UNIVERSITÉ DE LORRAINE | BIBLIOTHÈQUES UNIVERSITAIRES

AVERTISSEMENT

Ce document est le fruit d'un long travail approuvé par le jury de soutenance et mis à disposition de l'ensemble de la communauté universitaire élargie.

Il est soumis à la propriété intellectuelle de l'auteur. Ceci implique une obligation de citation et de référencement lors de l'utilisation de ce document.

D'autre part, toute contrefaçon, plagiat, reproduction illicite encourt une poursuite pénale.

Contact bibliothèque : ddoc-theses-contact@univ-lorraine.fr
(Cette adresse ne permet pas de contacter les auteurs)

LIENS

Code de la Propriété Intellectuelle. articles L 122. 4
Code de la Propriété Intellectuelle. articles L 335.2- L 335.10
http://www.cfcopies.com/V2/leg/leg_droi.php
http://www.culture.gouv.fr/culture/infos-pratiques/droits/protection.htm

UNIVERSITÉ DE LORRAINE

École doctorale IAEM Lorraine

Détection des anévrismes intracrâniens par apprentissage profond

THÈSE

présentée et soutenue publiquement le 22 mars 2024

pour l'obtention du

Doctorat de l'Université de Lorraine

(mention informatique)

par

Youssef Assis

Composition du jury

Présidente :	Carole Le Guyader	Professeure - INSA Rouen Normandie
Rapporteurs :	Christine Fernandez-Maloigne	Professeure - Université de Poitiers
	Antoine Vacavant	Professeur - Université Clermont Auvergne
Examinateurs :	Carole Lartizien	Directrice de recherche - CNRS
	Fabien Pierre	Maître de Conférences - Université de Lorraine
Encadrants :	René Anxionnat	Professeur - Université de Lorraine
	Erwan Kerrien	Chargé de recherche - Inria

Laboratoire Lorrain de Recherche en Informatique et ses Applications — UMR 7503

Mis en page avec la classe thesul.

Remerciements

Je souhaite avant tout exprimer ma plus sincère gratitude à mon directeur de thèse, Erwan Kerrien, mon directeur de thèse, pour son engagement exceptionnel dans ce projet et les conseils précieux qu'il m'a offerts tout au long de ces trois années, sur le plan tant professionnel que personnel. Un remerciement tout aussi chaleureux est destiné à mon co-directeur de thèse, René Anxionnat, pour le partage de son expertise, et m'avoir donné la chance de plonger dans son domaine, y compris l'opportunité d'assister à des procédures de traitement d'anévrismes. Mes remerciements s'adressent également à Liang Liao et à l'équipe du service de neuroradiologie diagnostique et interventionnelle du CHRU de Nancy pour leur précieuse collaboration.

Je tiens également à exprimer ma gratitude envers Christine Fernandez-Maloigne, Professeure à l'Université de Poitiers, et Antoine Vacavant, Professeur à l'Université Clermont Auvergne, pour avoir accepté d'être rapporteurs de cette thèse. Je souhaite aussi remercier Carole Lartizien, Directrice de recherche au CNRS, et Fabien Pierre, maître de conférences à l'Université de Lorraine, pour leur participation à mon jury en tant qu'examinateurs. Mes remerciements vont à Carole Le Guyader, Professeure à l'INSA Rouen Normandie, pour l'honneur qu'elle m'a fait en présidant le jury.

Je suis profondément reconnaissant envers toutes les personnes que j'ai eu le plaisir de rencontrer au LORIA durant ces années enrichissantes. Mes remerciements spéciaux vont aux membres de l'équipe Tangram : Marie-Odile, Pierre-Frédéric, Frédéric, Gilles, Brigitte, Véronique, Isabelle, ainsi qu'à mes collègues de bureau, qui ont contribué à rendre notre environnement de travail exceptionnellement chaleureux et accueillant. J'apprécie énormément les instants partagés avec Karim, Nariman, Romain, et Matthieu, qui ont été des moments de détente et d'évasion. Ma reconnaissance s'étend également à Nicolas, Radhouane, et Hugo. Un merci tout particulier à Laure Buhry pour son soutien constant et son écoute attentive en tant que membre du comité de suivi.

Enfin, je souhaite exprimer ma gratitude profonde à ma famille, particulièrement à mes parents, pour leur soutien inconditionnel. Je remercie sincèrement mes sœurs, Lamiae et Khadija, ainsi que mon frère Ayoub, pour leur encouragement constant. Une mention spéciale à Lamiae, pour son appui quotidien et son aide précieuse dans la recherche d'un équilibre entre mes obligations professionnelles et mes périodes de relaxation. Ce travail leur est intégralement dédié.

Table des matières

Table des figures **7**

Liste des tableaux **11**

Glossaire **13**

1 Introduction **15**

1.1 Contexte et motivations 15

1.1.1 Détection automatique en imagerie médicale 15

1.1.2 Diagnostic des anévrismes intracrâniens 16

1.1.3 Contexte et objectifs de la thèse 17

1.2 Contributions 18

1.2.1 Organisation du manuscrit 18

1.2.2 Publications et logiciels associés 20

2 Contexte clinique des anévrismes intracrâniens **23**

2.1 Anatomie et physiopathologie des anévrismes intracrâniens 24

2.1.1 Structure et fonction du cerveau et des vaisseaux sanguins 24

2.1.2 Caractéristiques des anévrismes intracrâniens 27

2.1.3 Facteurs de risque de rupture liés aux anévrismes intracrâniens 28

2.1.4 Options de traitement des anévrismes intracrâniens 29

2.2 Imagerie médicale pour la visualisation des anévrismes intracrâniens 30

2.2.1 Angiographie par tomodensitométrie (CTA) 31

2.2.2 Angiographie rotationnelle tridimensionnelle (3DRA) 31

2.2.3 Angiographie par résonance magnétique (MRA) 32

2.2.4 Comparaison des imageries 3D pour la visualisation des anévrismes 33

2.3 Diagnostic des anévrismes intracrâniens 35

2.3.1 Processus de prise en charge 35

2.3.2 Détection et localisation des anévrismes 35

2.3.3 Positionnement du plan de coupe pour une visualisation améliorée 36

2.3.4 Segmentation des anévrismes pour le suivi du traitement 37

3 La vision par ordinateur pour le diagnostic des anévrismes intracrâniens 41
3.1 Vision par ordinateur en imagerie médicale . . . 42
3.1.1 Diversité de tâches d'analyse d'images . . . 42
3.1.2 Évolution des approches d'analyse d'images . . . 44
3.2 Apprentissage profond pour l'analyse d'images médicales . . . 46
3.2.1 Réseaux de neurones convolutionnels et mécanismes d'attention . . . 46
3.2.2 Métriques d'évaluation . . . 51
3.2.3 Méthodologies d'évaluation . . . 55
3.2.4 Défis de l'apprentissage profond pour les anévrismes intracrâniens . . . 57
3.3 Détection et segmentation des anévrismes intracrâniens . . . 59
3.3.1 Approches classiques de traitement d'images . . . 59
3.3.2 Approches par apprentissage automatique . . . 60
3.3.3 Approches par apprentissage profond . . . 62
3.4 Bases de données et annotation des anévrismes intracrâniens . . . 69
3.4.1 Bases de données disponibles . . . 70
3.4.2 Techniques d'annotation . . . 72

4 Stratégie de données efficace pour la détection d'anévrismes intracrâniens 75
4.1 Approches de détection d'anévrismes par segmentation . . . 76
4.2 Stratégie d'échantillonnage et de génération de données . . . 79
4.2.1 Collection et annotation de données . . . 79
4.2.2 Approche par petits patchs . . . 80
4.2.3 Sélection et synthèse de données . . . 81
4.3 Réseau de neurones pour la détection des anévrismes . . . 82
4.3.1 Architecture du réseau de segmentation UNet . . . 82
4.3.2 Détails d'implémentation . . . 83
4.3.3 Métriques et méthodologies d'évaluation . . . 84
4.4 Étude expérimentale et discussion . . . 85
4.4.1 Étude par ablation . . . 85
4.4.2 Évaluation par validation croisée . . . 87
4.4.3 Synthèse et analyse des résultats . . . 89
4.5 Données de référence et intégration de l'a-priori médical . . . 91
4.5.1 Consolidation des annotations et reproductibilité . . . 91
4.5.2 Accélération de l'entraînement et optimisation des hyper-paramètres . . . 92
4.5.3 Intégration de l'a-priori médical pour améliorer la détection . . . 94

5 Détection d'anévrismes intracrâniens : une perspective de détection d'objet 101
5.1 Limitations des travaux existants de détection d'anévrismes . . . 102
5.2 Collection et préparation de données . . . 104
5.2.1 Ensembles de données et annotation . . . 104
5.2.2 Échantillonnage et génération de données . . . 105
5.3 Détection des anévrismes avec des représentations sphériques . . . 106

5.3.1 Réseau de détection d'objets et fonction de coût . . . 106
5.3.2 Détails d'implementation . . . 108
5.3.3 Métriques d'évaluation et méthodes de référence . . . 109
5.4 Étude expérimentale et résultats . . . 110
5.4.1 Étude par ablation . . . 111
5.4.2 Évaluation par validation croisée . . . 112
5.4.3 Évaluation par experts et consolidation des annotations . . . 114
5.4.4 Influence des petits patchs sur des données limitées . . . 116
5.5 Discussion et analyse qualitative . . . 117
5.5.1 Architectures de réseau et déséquilibre de classes . . . 117
5.5.2 Représentations sphériques et analyse des résultats . . . 118

6 Estimation de la pose des anévrismes intracrâniens dans un espace 3D 121
6.1 Positionnement de plans de coupes en imagerie médicale . . . 122
6.1.1 Formulation du problème d'estimation de la pose d'objets . . . 123
6.1.2 Approches géométriques pour l'estimation de la pose . . . 124
6.1.3 Approches par apprentissage pour l'estimation de la pose . . . 125
6.2 Collection et préparation de données . . . 127
6.2.1 Ensembles de données et annotation . . . 127
6.2.2 Échantillonnage et génération de données . . . 129
6.3 Estimation de la pose des anévrismes par points de repère . . . 129
6.3.1 Architecture du réseau et fonction de coût . . . 129
6.3.2 Détails d'implémentation . . . 131
6.3.3 Métriques d'évaluation . . . 131
6.4 Étude expérimentale et résultats . . . 132
6.4.1 Étude par ablation . . . 133
6.4.2 Évaluation par validation croisée . . . 134
6.5 Discussion et analyse des résultats . . . 137
6.5.1 Méthode pour l'estimation de pose et la détection d'objets . . . 137
6.5.2 Outil de visualisation et d'analyse des anévrismes . . . 138
6.5.3 Analyse qualitative des résultats . . . 139

7 Conclusion et perspectives 141
7.1 Résumé des contributions . . . 141
7.2 Perspectives futures . . . 143

Annexe 147

A Outil de vérification des détections d'anévrismes 147

Bibliographie 153

Table des figures

1.1 Anévrismes intracrâniens : sacculaire et fusiforme. Une augmentation de la pression artérielle peut provoquer une rupture de l'anévrisme extrêmement mortelle. 17
1.2 Représentation tridimensionnelle d'une image TOF-MRA montrant les plans de coupes axial, sagittal et coronal. L'anévrisme intracrânien, mesurant 4,83 mm, est indiqué par la flèche. 18

2.1 Système vasculaire cérébral [2] : artères, veines et capillaires. 24
2.2 Principales artères cérébrales : le polygone de Willis est situé à la base du cerveau. 25
2.3 Principales affections cérébro-vasculaires [2]. 26
2.4 (a) Les principales formes d'anévrismes cérébraux : sacculaire (gauche), fusiforme (milieu), et mixte (droite). (b) Illustration de la structure d'un anévrisme de forme sacculaire : Le dôme, situé à l'opposé de l'ostium, sépare la poche anévrismale du vaisseau parent [212]. . 28
2.5 Options de traitement des anévrismes intracrâniens. 30
2.6 Exemple d'image d'angiographie par tomodensitométrie (CTA) présentant la vue en 3D (a) ainsi que les plans de coupes 2D (b, c, d). L'anévrisme intracrânien, mesurant 4,18 mm, est identifié par la flèche sur les plans de coupe. Les vaisseaux sanguins sont difficile à visualiser dans la vue 3D. 31
2.7 Exemple d'image d'angiographie rotationnelle par rayons X (3DRA) illustrant la vue en 3D (a) et les plans de coupes 2D (b, c, d). L'anévrisme intracrânien de 6,83 mm est indiqué par la flèche. 32
2.8 Exemple d'image TOF-MRA montrant (a) la vue 3D et (b, c, d) les plans de coupes 2D. L'anévrisme intracrânien de 4,83 mm est indiqué par la flèche. 33
2.9 Détection en deux étapes d'un anévrisme intracrânien de 6,61 mm. Première analyse globale en vue 3D (gauche), suivie d'une analyse locale plus détaillée en plans de coupes canoniques (droite). 36
2.10 Analyse en 2D d'un anévrisme de 6,61 mm : mesures du collet, du diamètre, de la largeur, et l'évaluation de la connexion avec le vaisseau parent. Utilisation d'un plan de coupe reformaté avec projection MIP pour une analyse optimale de l'anévrisme et de sa connexion avec le vaisseau parent. 37
2.11 Segmentation d'un anévrisme intracrânien de 3,11 mm à partir d'images 3DRA [216] : analyse de l'anévrisme (rouge) et sa connexion avec le vaisseau parent (vert). 38

3.1 Comparaison des tâches de vision par ordinateur en imagerie médicale [192] : (a) classification d'images, (b) détection d'objets, (c) segmentation sémantique, et (d) segmentation d'instances. 43
3.2 Positionnement des coupes transversales d'IRM par rapport à l'articulation du genou [328]. 44
3.3 Exemple de réseau CNN pour la classification d'images [243] : Couches de convolution avec activation ReLU et pooling pour l'extraction des caractéristiques, suivies de couches MLP et de la fonction Softmax pour générer les scores de classification. 47
3.4 Réseaux YOLO [238] pour la détection d'objets en une seule étape : Encodeur suivi d'un bloc de détection pour la localisation et la classification des objets. 49

3.5 Réseau 3D UNet [65] : traitement d'une image 3D en entrée pour générer la segmentation en sortie, avec l'intégration de connexions "skip" par des opérations de concaténation. . . 50
3.6 Analyse des performances à travers des métriques basées sur des représentations graphiques. 54
3.7 Importance de choisir des métriques appropriées à la problématique (vérité terrain en bleu). 56
3.8 nnUNet [135] : un framework auto-configurable pour la segmentation d'images médicales, présentant une configuration automatique adaptée aux besoins de chaque ensemble de données. 67
3.9 Framework nnDetection [27] pour la détection d'objets : choix automatiquement des données d'apprentissage et ajustement de la structure du réseau selon plusieurs règles. 68
3.10 Exemples d'images TOF-MRA de la base de données CHUV [80] : anévrismes de tailles variées sont indiqués. 71
3.11 Variabilité d'annotation voxel par voxel d'un anévrisme intracrânien de 7,75 mm : (a) vue 3D et (b, c, d) exemples de coupes axiales. (e) Annotation approximative d'un anévrisme de 5,76 mm par une sphère englobante. 73

4.1 Défis liés à l'utilisation de réseaux basés sur la segmentation pour les tâches de détection d'objets. (a) Métriques de la compétition ADAM : évaluation de la distance entre les centres des prédictions (en vert) et les annotations (en rouge). (b) Le seuil de classification affecte le nombre et les dimensions des composantes connexes extraites. 78
4.2 Caractéristiques de la base de données du CHRU de Nancy : (a) exemple d'image après la suppression du crâne, (b) distribution des tailles des anévrismes. 79
4.3 Plugin 3D Slicer développé pour l'annotation des anévrismes : Chaque anévrisme est approximé par une sphère définie par deux points (F1, F2). 80
4.4 Anévrismes intracrâniens de différentes tailles au sein de patchs de $48 \times 48 \times 48$ voxels. . 81
4.5 (a) Exemple d'anévrismes synthétisés par distorsion aléatoire à partir d'un anévrisme de taille 2,92 mm. (b) Localisations des centres de patchs positifs (en rouge), des patchs négatifs (en bleu pour les vaisseaux sanguins et en vert pour le parenchyme). 82
4.6 Architecture du réseau 3D UNet [65] employé pour la segmentation d'images de dimensions $48 \times 48 \times 48$ voxels. Les dimensions des images et des cartes de caractéristiques sont illustrées. 83
4.7 Comparaison des métriques d'évaluation entre la vérité terrain (en vert) et les prédictions (en rouge) définies par des sphères : utilisation (b) des métriques d'ADAM basées sur la distance entre les centres des sphères et (c) des métriques basées sur l'intersection IoU entre les sphères. 84
4.8 (a) Évolution des coefficients Dice et Kappa pour les 5 modèles lors de la validation croisée. (b, c) Courbes FROC et Sensibilité-Précision de notre méthode utilisant les métriques ADAM. 88
4.9 Visualisation des résultats de notre méthode : les CC prédites en rouge, anévrismes détectés sont indiqués en flèche verte. (a, b) La ramification de petites artères peut être confondue avec un anévrisme (flèches rouges). (c) Petit anévrisme non détecté mesurant 2,74 mm (flèche jaune). 90
4.10 Variabilité des annotations de la base de données CHRU de Nancy : (a, b) Exemples d'anévrismes non annotés, mesurant respectivement 7,09 mm et 6,63 mm. (c, d) Deux annotations incorrectes, en réalité ne sont pas des anévrismes. 91
4.11 Exemples de faux positifs dans différentes localisations : (a) irrégularité de l'artère carotide, et (b, c, d) origine de petites artères. 94
4.12 Réseau UNet avec une supervision profonde : Combinaison des coûts pondérés, calculés par la formule $\sum_{i=1}^{4} \frac{1}{2^{i-1}} \times L_i$ à chaque résolution i. La vérité terrain est dérivée par interpolation. 96
4.13 (a) Comparaison des annotations sphériques (rouge) et des annotations raffinées (vert). (b) Segmentation manuelle des vaisseaux sanguins sur une image présentant un anévrisme (flèche). 98
4.14 Réseau UNet avec deux résolutions en entrée : Fusion des cartes de caractéristiques à chaque niveau de l'encodeur via une concaténation. 98

4.15 Réseau Double UNet traitant deux résolutions : Fusion des caractéristiques des vaisseaux sanguins (en bas) et des anévrismes (en haut) dans le décodeur avec un mécanisme d'attention. 99

5.1 Comparaison des données de la base CHUV [80] : (a) Rendu 3D et plans axiaux montrant notre annotation (rouge) par rapport à l'annotation initiale (bleu) incluant les vaisseaux sains. (b) Exemple d'anévrismes non annotés (3,06 mm et 3,22 mm) en haut, et annotations incorrectes (dysplasie simple ou irrégularité sur athérome à gauche, infundibulum à droite) en bas. 105
5.2 Architecture proposée pour la détection des anévrismes utilisant des sphères (C_x, C_y, C_z, r) et des scores de confiance (c). Les cartes de caractéristiques 2D sont représentées à des fins de visualisation ; leurs tailles réelles en 3D sont affichées. 107
5.3 Illustration de l'approche de prédiction par patchs sur le volume global d'une image TOF-MRA : division du volume en patchs de $96 \times 96 \times 96$ voxels avec un chevauchement de 32 voxels. Les prédictions sont extraites à partir de la partie centrale ($64 \times 64 \times 64$ voxels) de chaque patch. 109
5.4 Performance de notre méthode sur la base de données CHRU de Nancy : (a) Sensibilité de détection par taille d'anévrismes. (b) Scores de confiance des détections. 113
5.5 Exemples d'erreurs d'annotations dans la base de données CHUV [80] : (a, b) anévrismes non annotés, (c, d) annotations incorrectes (ici, départ de petites artères). 115
5.6 Évaluation comparative des performances de quatre techniques sur la base de données CHUV de Lausanne, avec annotations révisées. Notre méthode est DeepAneDet. 115
5.7 Performance de notre méthode sur la base CHUV de Lausanne en utilisant les annotations révisées : (a) sensibilité de détection par taille d'anévrismes, et (b) score de confiance par détections. 116
5.8 Comparaison des détections : notre méthode (sphères en rouge) vs *Modèle5* (boîtes en bleu). Les boîtes montrent une instabilité de taille lors des rotations, entraînant un faux positif (à droite). 118
5.9 Sphères de détection (rouge) et scores de confiance (bleu) : (a, b) exemples de petits anévrismes détectés par notre méthode (taille : 1,90 et 1,65 mm), (c) exemple de faux positif, et (d) anévrisme non détecté de taille 1,91 mm. 119

6.1 Visualisation d'un anévrisme de 2,69 mm de taille à travers les plans de coupes canoniques (a, b, c) et un plan de coupe reformaté (d) adapté à la direction de l'anévrisme, offrant une meilleure analyse de l'anévrisme et des structures environnantes. 122
6.2 Positionnement du plan de coupe au niveau du collet de l'anévrisme : les lignes médianes des vaisseaux et le centre de l'anévrisme sont utilisés pour définir le plan qui sépare l'anévrisme du vaisseau parent. 125
6.3 Estimation de la pose du genou [333] : Une première étape localise les RoIs avec un réseau Faster R-CNN sur une image à basse résolution, puis un second réseau repère précisément le centre du genou à partir d'une image à haute résolution. Le réseau V-Net segmente le plan d'orientation à partir de chaque RoI, suivi de PCA pour dériver la matrice de rotation. 126
6.4 Comparaison des tailles des anévrismes dans les ensembles de données CHRU et CHUV. . 127
6.5 Annotation de la pose des anévrismes : (a) les points d'annotation (Collet et Dôme, en jaune) et le vecteur de l'axe principal (en rouge). (b) La pose est déduite par le centre $C = (\mathrm{P1} + \mathrm{P2})/2$ et du vecteur d'axe $\vec{v} = \mathrm{P2} - C$. L'anévrisme peut également être représenté sous forme d'une sphère ayant un centre C et un rayon $r = |\vec{v}|$. A noter que les plans de coupes reformatés, contenant cet axe, permet une analyse morphologique et métrologique de l'anévrisme. 128
6.6 Notre architecture d'estimation de la pose de l'anévrisme. Les cartes de caractéristiques en 2D ne sont utilisées qu'à des fins de visualisation ; leurs tailles réelles en 3D sont affichées 130
6.7 Étude par ablation utilisant diverses configurations pour représenter la pose de l'anévrisme. 133

6.8 Performances des méthodes dans la détection des anévrismes sur l'ensemble de données du CHUV de Lausanne, en utilisant les annotations révisées. Ici, notre méthode est DeepAnePose. 136
6.9 Les performances de notre méthode d'estimation de la pose des anévrismes sur l'ensemble de données CHUV de Lausanne ont été évaluées. Ces performances ont été calculées sur les anévrismes correctement détectés et reposent sur des annotations révisées. 137
6.10 Extension du logiciel 3D Slicer développé pour la visualisation de la pose des anévrismes. 138
6.11 Résultats qualitatifs sur l'ensemble de données CHUV de Lausanne : les points prédites (bleu) et les vérités terrain associées (vert). Chaque plan de coupe reformaté a été déterminé en tournant autour de l'axe de l'anévrisme passant par les directions prédites. 139

A.1 Configuration de l'outil dans le logiciel 3D Slicer. 148
A.2 Deux méthodes pour accéder à l'outil depuis l'interface principale du logiciel 3D Slicer. . . 148
A.3 Structure du dossier des détections à vérifier. 149
A.4 Composants et fonctionnalités de l'outil de vérification des détections. 149

Liste des tableaux

2.1 Comparaison des techniques d'imagerie médicale 3D largement utilisées pour la visualisation des anévrismes intracrâniens. Ces mesures peuvent varier en fonction de l'équipement utilisé. 34

3.1 Exemples de métriques d'évaluation pour des problèmes de régression (valeurs prédites x, réelles $\hat{x}$, échantillons n) et de classification (composantes de matrice de confusion, détections K). 53

3.2 Analyse comparative des compétitions ADAM (en haut) et CADA (en bas) pour la tâche de détection d'anévrismes : Prédominance des approches en 3D utilisant des réseaux de segmentation d'images et le coefficient de Dice comme fonction de coût. L'astérisque (*) indique le rapport entre le nombre de patchs positifs et le nombre total de patchs dans l'ensemble d'entraînement. 66

4.1 Performances des meilleures méthodes de la compétition ADAM [285] : Prédominance des approches de segmentation et d'apprentissage ensembliste. Diverses opérations de post-traitement sont employées : filtrage des détections par nombre (N), par taille/volume (T) et par localisation (L). 76

4.2 Comparaison des variantes du modèle : *Modèle0* représente notre méthode proposée. Les métriques Dice et κ ont été évaluées sur l'ensemble de validation, tandis que les métriques d'ADAM, d'IoU, ainsi que par patchs ont été calculées sur l'ensemble de test. La sensibilité, le nombre de faux positifs (FP/cas), et la spécificité sont calculées à un seuil de classification par défaut de 50%. 86

4.3 Comparaison des méthodes de détection basées sur la segmentation : (En haut) Résultats officiels des 7 meilleures méthodes d'ADAM, évaluées sur les données privées d'ADAM. (En bas) Notre méthode est comparée à nnUNet en utilisant l'ensemble de données du CHRU. La variabilité des performances, mesurée par l'écart-type, est calculée sur les différents plis de test. 89

4.4 Optimisation des hyperparamètres sur un ensemble de test (40 images, 68 anévrismes). Les métriques sont basées sur un seuil IoU de 10%. * indique les éléments intégrés dans le modèle final. 94

4.5 Stratégies d'amélioration de la détection des petits anévrismes (en haut) et de réduction des faux positifs (en bas). Métriques d'évaluation basées sur un seuil d'IoU de 10%, appliquées à un ensemble de test composée de 68 anévrismes (13 anévrismes $<$ 2mm, 42 anévrismes entre 2 et 5mm). 96

5.1 Comparaison des anévrismes entre les ensembles de données CHRU et CHUV [80]. 105

5.2 Étude d'ablation de notre méthode : Expérimentations menées sur la base de données du CHRU de Nancy. Les métriques d'évaluation sont calculées en utilisant un seuil IoU de 10%. 112

5.3 Validation croisée à 5 plis sur la base du CHRU de Nancy : Comparaison avec des méthodes de référence. DeepAneDet est notre méthode proposée. Les scores de variabilité sont obtenus à partir des différents sous-ensembles de test. 113

5.4 Validation croisée à 5 plis (moyenne ± écart-type) sur l'ensemble de données CHUV. Chaque méthode a été entraînée avec les annotations initiales et testée avec ces mêmes annotations (en haut), ainsi qu'avec des annotations révisées (au milieu et en bas). * indique les méthodes entraînées uniquement sur des images présentant des anévrismes et testés sur l'ensemble des images, y compris celles sans anévrismes. 114

6.1 Étude par ablation sur la tâche d'estimation de la pose des anévrismes en utilisant la base de données du CHRU de Nancy. Les performances pour la tâche de détection ont été ajoutées. 134
6.2 Résultats (moyenne ± écart-type) de la détection des anévrismes sur les ensembles de données CHRU et CHUV. * indique l'utilisation des annotations révisées de la base CHUV pour l'évaluation. La méthode DeepAneDet fait référence à notre méthode de détection décrite dans le chapitre précédent. 135
6.3 Évaluation de la précision de l'estimation de la pose de notre méthode sur les ensembles de données CHRU et CHUV. Les modèles évalués en utilisant les annotations révisées sont indiqués par un "*". La méthode DeepAneDet désigne notre méthode de détection décrite dans le chapitre précédent. 136

Glossaire

2D : 2 dimensions
3D : 3 dimensions
3DRA : Three-dimensional rotational angiography
ACA : Anterior Cerebral Artery
ADAM : Aneurysm detection and segmentation
ADASYN : Adaptive synthetic
ANR : Agence nationale de la recherche
AP : Average precision
ACom : Anterior Communicating Artery
AUC : Area under curve
AVC : Accident vasculaire cérébral
BCE : Binary cross entropy
BRATS : Multimodal brain tumor segmentation challenge
CAD : Computer aided detection
CADA : Cerebral aneurysm detection and analysis
CBAM : Convolutional block attention module
CE-MRA : Contrast enhanced magnetic resonance angiography
CHRU : Centre hospitalier régional universitaire
CHUV : Centre hospitalier universitaire vaudois
COCO : Microsoft Common Objects in Context
CNN : Convolutional neural network
CT : Computed tomography
CTA : Computed tomography angiography
DAN : Dual attention network
DALI : Deep generative models, and data augmentation, labelling, and imperfections
DeiT : Data-efficient image Transformer
DSA : Digital subtraction angiography
ECA : External Carotid Artery
ECA-Nets : Efficient channel attention networks
FCN : Fully connected network
FLAIR : Fluid-attenuated inversion recovery
FN : False negative
FP : False positive
FPN : Feature pyramid network
FROC : Free-response receiver operating characteristic
GIoU : Generalized intersection over union
GPU : Graphics processing unit
HSA : Hémorragie sub-arachnoïdienne
IA : Intelligence artificielle
IABM : Intelligence artificielle en imagerie biomédicale
ICA : Internal Carotid Artery
IJCARS : International journal of computer assisted radiology and surgery
IoU : Intersection over union

IRM : Imagerie par résonance magnétique
IXI : Information extraction from images
KNN : k nearest neighbors
LORIA : Laboratoire Lorrain de Recherche en Informatique et ses Applications
MAV : Malformation artério-veineuse
MedT : Medical transformer
MICCAI : Medical image computing and computer-assisted intervention
MIP : Maximum intensity projection
MRI : Magnetic resonance imaging
MRA : Magnetic resonance angiography
ORASIS : Journées francophones des jeunes chercheurs en vision par ordinateur
PCom : Posterior Communicating Artery
PCA : Posterior Cerebral Artery
PVT : Pyramid Vision Transformer
RCAN : Residual channel attention network
ReLU : Rectified linear unit
ResNet : Residual neural network
R-CNN : Region-based convolutional neural network
RPN : Region proposal network
ROC : Receiver operating characteristic
RoI : Region of interest
SENet : Squeeze and excitation network
SFNR : Société française de neuroradiologie
SGD : Stochastic gradient descent
SMOTE : Synthetic minority over-sampling technique
SSD : Single-shot detector
SVM : Support vector machine
T : Tesla
TN : True negative
TP : True positive
TOF-MRA : Time-of-flight magnetic resonance angiography
TTA : Test time augmentation
UIATS : The unruptured intracranial aneurysm treatment score
VA : Vertebral Artery
VGG : Visual geometry group
ViT : Vision transformer
VR : Volume rendering
WEB : Woven endobridge
YOLO : You only look once

Chapitre 1

Introduction

1.1 Contexte et motivations

1.1.1 Détection automatique en imagerie médicale

La découverte des rayons X en 1895 a marqué un tournant déterminant dans le domaine médical, donnant naissance à l'imagerie médicale en tant que discipline à part entière. Cette discipline permet la visualisation des structures internes du corps humain sans avoir recours à des interventions chirurgicales [17]. Cette avancée majeure a joué un rôle décisif en fournissant une méthode non invasive pour observer et analyser les mécanismes internes du corps sous forme d'images en deux dimensions (2D). Dans les années 70, une étape significative a été franchie avec l'introduction de la tomodensitométrie (appelé aussi CT scan) à rayons X en pratique clinique, marquant ainsi la première des modalités d'imagerie modernes basées sur des coupes transversales [152]. Cette modalité offre une représentation tridimensionnelle (3D) plus réaliste des structures anatomiques, englobant des informations détaillées relatives aux dimensions, à la morphologie et à l'orientation spatiale des structures internes. Par la suite, dans les années 1980, l'émergence de l'imagerie par résonance magnétique (IRM) a inauguré une nouvelle ère dans le domaine de l'imagerie médicale. En recourant à des champs magnétiques et des ondes radio au lieu de radiations ionisantes, l'IRM a ouvert de nouvelles perspectives pour la détection et le diagnostic d'anomalies diverses, telles que les tumeurs et les lésions, y compris les anévrismes intracrâniens [5].

Au fil du temps, les progrès technologiques et la diversification des techniques d'imagerie médicale ont grandement renforcé les capacités diagnostiques, améliorant ainsi les soins et le traitement des patients [36]. Ces avancées ont été rendues possibles grâce à l'amélioration des équipements d'imagerie, à l'optimisation des protocoles de numérisation et aux développements de logiciels de reconstruction d'images en 3D [286]. Cependant, cette expansion sans précédent de l'imagerie médicale a également posé un défi considérable pour les professionnels de la santé, qui doivent faire face à un volume croissant de données d'imagerie à évaluer et à analyser manuellement dans des délais de plus en plus restreints [283]. Par exemple, dans certains hôpitaux, les radiologues doivent évaluer une image toutes les 3 à 4 secondes en moyenne pour répondre à leur charge de travail hebdomadaire [198].

L'analyse manuelle des images médicales représente une tâche exigeante pour les radiologues. Cette pratique nécessite un investissement important en temps et en compétences humaines, étant donné sa nature laborieuse et chronophage. En outre, la possibilité d'erreurs d'interprétation est non négligeable, souvent due à la fatigue et à la baisse de concentration, qui sont des conséquences fréquentes de l'analyse prolongée d'images complexes [41]. Les avancées technologiques ont contribué à une augmentation significative de la complexité des images médicales. Ces images se caractérisent aujourd'hui par des dimensions plus étendues avec davantage d'informations visuelles, nécessitant une vigilance et une attention particulières lors de leur analyse. Cette complexité est également influencée par les caractéristiques spécifiques des éléments à analyser dans ces images, tels que leur taille, leur nombre, leur emplacement, et surtout l'efficacité des techniques d'imagerie utilisées pour capturer et représenter leurs signaux. À titre d'exemple, la visualisation de certaines anomalies, comme les anévrismes intracrâniens, est rendue ardue en raison de leur petite taille et de leur subtilité sur les images médicales [217].

L'augmentation continue du nombre d'images médicales, ainsi que les défis associés à leur évaluation manuelle, ont engendré la nécessité de développer des outils d'aide à la décision. Ces outils visent à faciliter l'analyse et l'interprétation de ces images et ont pour objectif d'accompagner les professionnels de la santé, notamment les radiologues, dans leur pratique clinique. Ils visent à réduire le temps nécessaire pour l'analyse des images, voire améliorer l'efficacité et la précision du diagnostic. Récemment, des avancées significatives dans le domaine de l'intelligence artificielle (IA), en particulier dans le domaine de l'apprentissage profond, ont émergé comme des solutions prometteuses pour répondre à ces besoins [283]. L'intégration de ces outils au sein de la pratique clinique permet d'automatiser des tâches cruciales, telles que la détection, la segmentation, ainsi que l'évaluation quantitative et qualitative de plusieurs anomalies qui ont un impact sur la santé humaine. Parmi les anomalies particulièrement graves nécessitant une détection et une prise en charge précoces, les anévrismes intracrâniens occupent une position prépondérante [210]. L'utilisation d'outils automatiques basés sur l'apprentissage profond renforce la capacité à dépister de manière précoce et précise ces anomalies potentiellement mortelles, facilitant une intervention appropriée, et contribuant ainsi à améliorer les résultats pour les patients.

1.1.2 Diagnostic des anévrismes intracrâniens

Anévrismes intracrâniens

Les anévrismes intracrâniens, également appelés anévrismes cérébraux, sont des anomalies vasculaires complexes d'une grande importance clinique. Ils résultent de dilatations localisées des vaisseaux sanguins cérébraux, créant ainsi des poches de sang potentiellement dangereuses (voir Figure 1.1). Ces anomalies découlent de la fragilité des tissus vasculaires qui recouvrent la paroi artérielle, entraînant une dégradation progressive de leur structure au fil du temps [257]. Généralement de très petites structures mesurant entre 1 et 30 mm, les anévrismes se distinguent en deux configurations principales ; les anévrismes sacculaires et les anévrismes fusiformes, qui présentent des propriétés distinctes [137]. Toutefois, il est important de noter que 85% des anévrismes intracrâniens sont de type sacculaire.

L'incidence mondiale des anévrismes intracrâniens est estimée à être comprise entre 3,6% et 6% [297]. Malgré leur tendance à la stabilité sur de longues périodes et la possibilité pour de nombreux patients de vivre avec un anévrisme intracrânien pendant des durées prolongées, il convient de noter que chaque anévrisme comporte un risque potentiel de rupture extrêmement dangereuse [43]. Une telle rupture entraîne une hémorragie dans l'espace sous-arachnoïdien adjacent au cerveau, également connue sous le nom d'hémorragie sous-arachnoïdienne (HSA). Cette complication met en péril la vie des individus touchés, avec des taux de morbidité et de mortalité dépassant les 40% [95]. Dans les cas de survie, des perturbations cognitives peuvent affecter les patients à long terme, voire de manière permanente. Il est essentiel de souligner que l'aspect critique des anévrismes intracrâniens réside dans leur nature souvent asymptomatique [158]. En effet, la majorité de ces anévrismes demeurent silencieux et ne présentent que rarement des symptômes avant de se rompre. Par conséquent, la plupart des diagnostics surviennent de manière fortuite lors de l'évaluation d'imageries cérébrales effectuées pour des raisons médicales distinctes ou dans le cadre de programmes de dépistage, en grande partie grâce à la généralisation et à la disponibilité des techniques d'imagerie dans les établissements de santé.

Prévenir la rupture des anévrismes intracrâniens repose sur un diagnostic précoce et une surveillance régulière des anévrismes non rompus. Cette vigilance est essentielle pour détecter d'éventuels changements significatifs indiquant un risque potentiel de rupture, comme l'augmentation de la taille de l'anévrisme ou toute instabilité [132]. En présence de signes révélateurs d'un risque élevé de rupture, plusieurs options thérapeutiques sont envisageables pour réduire ce risque, notamment le traitement endovasculaire et la pose d'un clip chirurgical [204].

Détection des anévrismes intracrâniens

Le diagnostic précoce et la surveillance régulière des anévrismes intracrâniens non rompus reposent sur l'utilisation de techniques d'imagerie d'angiographie. Ces techniques permettent la visualisation des tissus cérébraux, mettant particulièrement l'accent sur les vaisseaux sanguins, incluant les anévrismes. L'angiographie par soustraction numérique (DSA) a longtemps été incontournable, mais son rôle diagnostic a aujourd'hui essentiellement disparu au profit de l'angiographie par tomodensitométrie (CTA) et

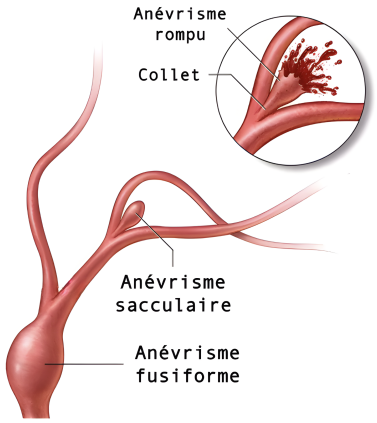


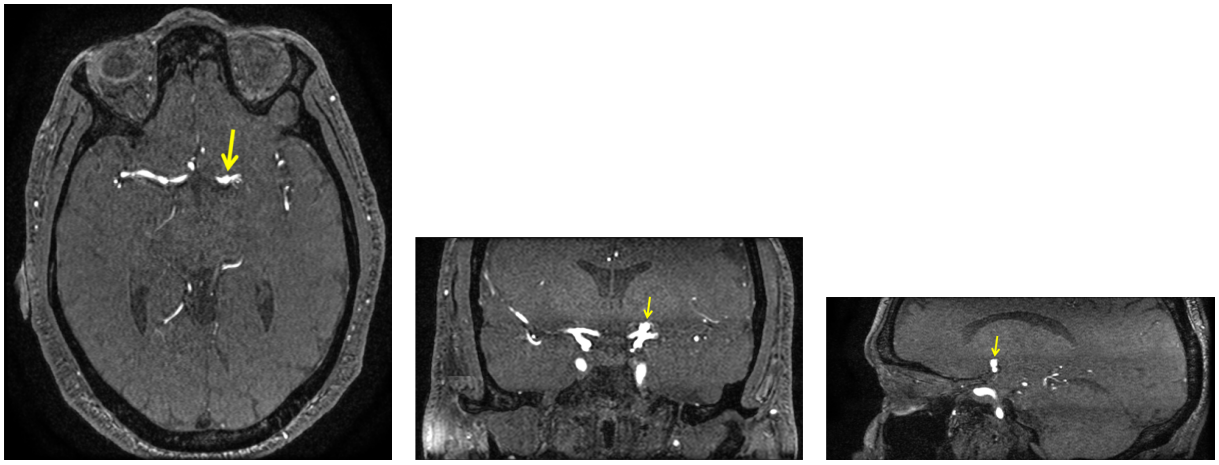
FIGURE 1.1 – Anévrismes intracrâniens : sacculaire et fusiforme. Une augmentation de la pression artérielle peut provoquer une rupture de l'anévrisme extrêmement mortelle.

l'angiographie par résonance magnétique en temps de vol (TOF-MRA). Cette dernière gagne en popularité en raison de son caractère non invasif et de l'absence de rayonnements ionisants [285]. Ces techniques d'imagerie génèrent, de manière non invasive, des images 3D du cerveau, en mettant en évidence les vaisseaux sanguins dans divers plans de coupe (voir un exemple d'image TOF-MRA en Figure 1.2). Par la suite, des radiologues analysent minutieusement ces images afin de détecter des anévrismes intracrâniens potentiels et évaluer la gravité ou le risque de rupture associé à chaque cas. Cette évaluation du risque repose sur divers facteurs morphologiques, tels que la taille, la forme et la localisation de l'anévrisme par rapport aux structures angio-architecturales environnantes [34, 88].

Cependant, la détection des anévrismes intracrâniens à partir d'images médicales est une tâche difficile, pour plusieurs raisons. La caractère volumineux et détaillé de ces images requiert une attention particulière pour repérer de telles structures de petite taille (entre 1 et 30 mm), rares (généralement entre 0 à 2 par patient) et dont la localisation varie d'un individu à l'autre. L'évaluation manuelle de ces images est une tâche chronophage demandant une concentration prolongée, et peut entraîner des variations d'interprétation entre les experts. Ces derniers ont souvent tendance à se concentrer sur les régions courantes, notamment le polygone de Willis, négligeant ainsi les sites moins fréquents [208]. De plus, la précision de la détection des anévrismes est influencée par les techniques d'imagerie employées. Certaines de ces modalités ne sont pas idéalement adaptées à la représentation des microstructures telles que les anévrismes, en particulier ceux de moins de 5 mm [217, 35]. Par conséquent, la sensibilité de détection des anévrismes varie, oscillant entre 60% et 88,1%, en fonction de l'expérience des radiologues, de la technique d'imagerie utilisée, de la qualité des images, de la taille et de l'emplacement des anévrismes [81]. Cette sensibilité est particulièrement réduite pour les anévrismes de petite taille (< 3 mm), pour lesquels elle peut atteindre 35%, ainsi que pour ceux situés à proximité des artères carotides internes et des artères cérébrales antérieures [308].

1.1.3 Contexte et objectifs de la thèse

Le diagnostic des anévrismes intracrâniens non rompus à partir d'images médicales représente un défi complexe au sein de la pratique clinique des professionnels de la santé, notamment des radiologues. Cette thèse a pour objectif de surmonter ces difficultés en développant des outils de détection et de visualisation automatisés pour soutenir les radiologues dans leur pratique quotidienne, ce qui permettra d'améliorer l'efficacité et la précision du processus diagnostique. À cette fin, la thèse se concentre sur l'exploration et la mise en œuvre de méthodes automatisées, en mettant particulièrement l'accent sur les avancées récentes en matière d'apprentissage profond, notamment l'utilisation de réseaux de neurones convolutionnels (CNN). Ces techniques seront spécifiquement adaptées pour répondre aux défis particuliers posés par les anévrismes intracrâniens, en particulier dans l'analyse d'images issues de l'angiographie par résonance magnétique en temps de vol (TOF-MRA).



(a) Axial (b) Sagittal (c) Coronal

FIGURE 1.2 – Représentation tridimensionnelle d'une image TOF-MRA montrant les plans de coupes axial, sagittal et coronal. L'anévrisme intracrânien, mesurant 4,83 mm, est indiqué par la flèche.

Ce projet de thèse s'inscrit dans le cadre du plan doctoral en intelligence artificielle (IA) de l'agence nationale de la recherche (ANR) et bénéficie d'un financement conjoint du centre hospitalier régional universitaire (CHRU) de Nancy et de la région Grand-Est. Le projet repose sur une collaboration entre le service de neuroradiologie diagnostique et thérapeutique du CHRU et le laboratoire lorrain de recherche en informatique et ses applications (LORIA). Les objectifs spécifiques sont les suivants :

1. **Collection de données :** La collection de données revêt une importance cruciale dans le domaine de l'apprentissage profond, étant donné que l'entraînement de ces modèles repose largement sur des ensembles de données annotés. Afin de surmonter les défis associés à l'annotation laborieuse des images médicales et de favoriser la création rapide d'un ensemble de données étendu, notre premier objectif est de concevoir une approche d'annotation adaptée aux anévrismes intracrâniens.
2. **Détection automatisée :** Concevoir des architectures de réseaux de neurones spécifiquement dédiées à la détection des anévrismes intracrâniens à partir d'images 3D, en utilisant les ensembles de données collectés. Ces architectures seront conçues et entraînées en prenant en considération les caractéristiques et les défis associés à ces lésions vasculaires peu fréquentes et de petite taille, y compris les limitations des données disponibles et le problème de déséquilibre entre les classes.
3. **Visualisation appropriée :** Proposer une approche automatique complémentaire pour faciliter la visualisation et l'analyse des anévrismes détectés. Cette approche vise à offrir une visualisation immédiate et déterminer un plan de coupe reformatté optimal des anévrismes, éliminant ainsi le besoin de recherches distinctes sur différents plans de coupe. Cette approche améliorera tant l'analyse qualitative que quantitative, simplifiant ainsi l'évaluation des caractéristiques morphologiques associées à chaque anévrisme.
4. **Évaluation rigoureuse :** Réaliser une évaluation et une validation approfondies des méthodes d'apprentissage profond développés dans ce contexte clinique revêt une importance capitale. Cette étape a pour objectif d'évaluer la pertinence, l'efficacité et la valeur ajoutée de ces modèles pour le flux de travail des radiologues, malgré la taille réduite des ensembles de données disponibles.

1.2 Contributions

1.2.1 Organisation du manuscrit

Le chapitre 2 se consacre à une analyse du contexte médical associé aux anévrismes intracrâniens. Il débute par une présentation de l'anatomie et de la physiologie cérébrales, ainsi que des vaisseaux sanguins.

Cette présentation est suivie d'une exploration des caractéristiques propres aux anévrismes, englobant leur structure, leurs types, leur développement et leur risque de rupture. De plus, dans ce chapitre, une description et une comparaison des différentes techniques d'imagerie couramment employées pour visualiser les anévrismes sont fournies, en mettant un accent particulier sur les avantages de l'imagerie TOF-MRA dans le processus de diagnostic. Enfin, une focalisation est portée sur le protocole suivi par les experts pour diagnostiquer et évaluer les anévrismes, incluant leur détection, leur visualisation appropriée et leur segmentation.

Le chapitre 3 est divisé en quatre sections, chacune consacrée à un aspect spécifique de l'automatisation du diagnostic des anévrismes intracrâniens. La première section traite des applications de la vision par ordinateur dans l'analyse d'images médicales, en mettant en avant les méthodes utilisées pour automatiser ces tâches. Elle introduit également les concepts de base de l'apprentissage automatique couramment employées dans l'analyse d'images médicales. La deuxième section se penche sur les réseaux de neurones convolutionnels (CNN) et les mécanismes d'attention, tout en discutant des métriques et des méthodologies d'évaluation adaptées aux spécificités l'imagerie médicale. Elle résume également les défis rencontrés dans l'automatisation du diagnostic des anévrismes. La troisième section examine les différents travaux de détection des anévrismes, couvrant à la fois les méthodes classique de traitement d'images et les approches plus récentes basées sur les CNNs en 3D. Elle met en lumière les avancées permises par ces réseaux et leur impact potentiel sur la routine clinique des radiologues. Enfin, étant donné que les algorithmes d'apprentissage dépendent fortement des ensembles de données annotées, la dernière section se concentre sur les bases de données disponibles pour les anévrismes et les techniques d'annotation employées.

Dans le domaine de la détection d'objets en imagerie médicale, deux approches couramment adoptées sont les approches basées sur la segmentation et celles basées sur la localisation directe des objets. Dans le chapitre 4, nous présentons notre première contribution : une méthode de détection automatique d'anévrismes basée sur la segmentation. Pour relever les défis liés à l'exploitation des ensembles de données médicales, nous proposons une stratégie de gestion des données efficace. Contrairement à l'annotation voxel par voxel généralement utilisée en imagerie médicale, nous introduisons une nouvelle méthode d'annotation approximative et rapide des anévrismes, élargissant ainsi les bases de données. Notre approche repose sur l'extraction de petits patches 3D, et nous détaillons les techniques d'échantillonnage et de synthèse de données utilisées pour compenser le déséquilibre et la rareté des signaux d'anévrismes dans les images. Ces techniques sont employées pour entraîner un réseau de segmentation de type 3D UNet [65]. Enfin, nous évaluons l'efficacité de notre stratégie de gestion des données pour la détection d'anévrismes, en utilisant une étude par ablation et des métriques d'évaluation appropriées.

Les approches de détection basées sur la segmentation sont largement utilisées dans la littérature. Cependant, ces approches nécessitent des opérations de post-traitement, ce qui peut avoir un impact sur leur évaluation et leurs performances [27]. Dans le chapitre 5, nous présentons une approche de détection 3D entièrement basée sur la régression. Nous commençons par mettre en évidence les limitations des approches existantes en termes de performances, d'évaluation limitée et d'architectures de réseaux moins adaptées aux images 3D. Ensuite, nous exposons notre méthode qui utilise une représentation sphérique plus adaptée pour les anévrismes par rapport à la représentation classique basée sur les boîtes englobantes. Nous détaillons également l'architecture du réseau de neurones et la fonction de perte utilisée pour surmonter le problème de déséquilibre de classes. Pour évaluer les performances de notre méthode, nous utilisons des métriques spécifiques pour la tâche de détection, en particulier la précision moyenne. Nous réalisons des expérimentations incluant une étude par ablation, une validation croisée et une évaluation avancée par des experts.

Le chapitre 6 présente un nouvel outil d'assistance aux experts dans leur pratique clinique. Cette approche étend la détection des anévrismes à l'estimation du plan de coupe optimal en fonction de l'orientation de chaque anévrisme, facilitant ainsi la visualisation et l'analyse qualitative de ces lésions vasculaires. Pour introduire notre méthode, nous mettons en évidence l'importance de l'estimation de la pose des anévrismes et discutons des avantages d'une approche simultanée de détection et d'estimation de leur orientation. Ensuite, nous détaillons notre méthode, en mettant en avant les données et les annotations utilisées pour obtenir des informations sur la localisation et l'orientation des anévrismes. Nous exposons également l'architecture du réseau neuronal, la fonction de perte et les détails de mise en œuvre. Pour évaluer les performances de notre approche, nous définissons des métriques d'évaluation spécifiques,

mettant l'accent sur l'estimation de la pose des anévrismes et leur détection. Nous présentons les résultats obtenus à partir de deux vastes bases de données, et nous accordons une attention particulière à l'analyse qualitative de l'estimation de la pose des anévrismes et de leur détection. Une discussion approfondie de ces résultats nous permet de mettre en évidence les forces et les limites de notre approche.

Enfin, dans le chapitre 7, nous synthétisons nos résultats et explorons des perspectives futures. Nos travaux se concentrent sur la détection et l'estimation de la pose des anévrismes intracrâniens dans des images cérébrales en 3D. Les performances sont prometteuses malgré la variabilité des données de référence des experts. Ces résultats soulignent l'efficacité de nos méthodes et leur potentiel pour améliorer la routine clinique des radiologues. Cependant, une évaluation clinique approfondie est nécessaire pour confirmer leur efficacité et définir les domaines d'amélioration. De plus, ces méthodes peuvent être étendues à d'autres problèmes liés à des anomalies de petite taille et peu fréquentes, ce qui pourrait enrichir les ressources diagnostiques des professionnels de la santé et améliorer la prise en charge des patients.

1.2.2 Publications et logiciels associés

Au cours de cette thèse, deux publications scientifiques ont été acceptées et présentées lors de conférences internationales majeures dans le domaine. De plus, une publication est en cours de soumission dans une revue internationale. Ces publications ont valorisé les résultats clés obtenus tout au long de nos recherche, démontrant les avancées significatives réalisées dans le domaine de la détection automatisée des anévrismes intracrâniens. En complément, plusieurs communications ont été réalisées à l'échelle nationale, contribuant ainsi à la diffusion des connaissances acquises au cours de cette thèse. Les publications et communications associées à ce travail de recherche sont les suivantes :

- **Chapitre 4 :**
  - Youssef Assis, Liang Liao, Fabien Pierre, René Anxionnat, and Erwan Kerrien. "An efficient data strategy for the detection of brain aneurysms from MRA with deep learning". In Deep Generative Models, and Data Augmentation, Labelling, and Imperfections (DALI), pages 226–234. Springer, 2021.
  - Youssef Assis, Liang Liao, Fabien Pierre, René Anxionnat, and Erwan Kerrien. "Une stratégie de données efficace pour la détection des anévrismes cérébraux avec l'apprentissage profond". 18ème journées francophones des jeunes chercheurs en vision par ordinateur (ORASIS), 2021.
  - Liang Liao, Youssef Assis, Fabien Pierre, René Anxionnat, and Erwan Kerrien. "Une stratégie efficace de préparation des données pour la détection des anévrismes cérébraux en IRM 3D-TOF par deep learning". Société Française de Neuroradiologie (SFNR), 2022.
- **Chapitre 5 :**
  - Youssef Assis, Liang Liao, Fabien Pierre, René Anxionnat, and Erwan Kerrien. "Intracranial Aneurysm Detection using Spherical Representation". Colloque Français d'Intelligence Artificielle en Imagerie Biomédicale (IABM), 2023.
  - Youssef Assis, Liang Liao, Fabien Pierre, René Anxionnat, and Erwan Kerrien. "Intracranial Aneurysm Detection : An object detection perspective". Soumis a International Journal of Computer Assisted Radiology and Surgery (IJCARS), 2023.
- **Chapitre 6 :**
  - Youssef Assis, Liang Liao, Fabien Pierre, René Anxionnat, and Erwan Kerrien. "Aneurysm Pose Estimation with Deep Learning". In Medical Image Computing and Computer Assisted Intervention (MICCAI), vol 14221. Springer, Cham. 2023. Ce travail a reçu le prix "MICCAI STAR Award".

Dans le but de favoriser de nouvelles avancées dans le domaine de la détection automatisée des anévrismes intracrâniens, nous avons pris l'initiative de rendre publiquement accessibles en ligne les codes que nous avons développés pour implémenter nos méthodes basées sur l'apprentissage profond. De plus, nous mettons à disposition des outils spécialement dédiés à l'annotation et à la visualisation des données d'imagerie cérébrale. Ces ressources sont disponibles à l'adresse suivante : `https://gitlab.inria.fr/yassis`. Elles comprennent les extensions suivantes du logiciel Slicer [93] :

- `Slicer-Sphere-Annotation` : un outil conçu pour simplifier l'annotation des anévrismes intracrâniens à partir d'images médicales. Il repose sur une annotation approximative en utilisant des sphères, définies par deux points, comme marqueurs d'annotation.
- `SlicerCheckDetections` : un outil développé et destiné aux radiologues pour permettre la visualisation et la validation des détections d'anévrismes générées par des modèles d'apprentissage profond. De plus, les détections peuvent être classifiées en fonction de leur localisation et leur morphologie.
- `Slicer-Reformat-Aneurysm` : un module utilisé pour visualiser les détections d'anévrismes et à faciliter la manipulation de plans de coupe reformatés, en fonction de l'orientation des anévrismes intracrâniens.

Chapitre 2

Contexte clinique des anévrismes intracrâniens

Sommaire

2.1 Anatomie et physiopathologie des anévrismes intracrâniens 24
2.1.1 Structure et fonction du cerveau et des vaisseaux sanguins 24
2.1.2 Caractéristiques des anévrismes intracrâniens 27
2.1.3 Facteurs de risque de rupture liés aux anévrismes intracrâniens 28
2.1.4 Options de traitement des anévrismes intracrâniens 29
2.2 Imagerie médicale pour la visualisation des anévrismes intracrâniens . . . 30
2.2.1 Angiographie par tomodensitométrie (CTA) 31
2.2.2 Angiographie rotationnelle tridimensionnelle (3DRA) 31
2.2.3 Angiographie par résonance magnétique (MRA) 32
2.2.4 Comparaison des imageries 3D pour la visualisation des anévrismes 33
2.3 Diagnostic des anévrismes intracrâniens 35
2.3.1 Processus de prise en charge 35
2.3.2 Détection et localisation des anévrismes 35
2.3.3 Positionnement du plan de coupe pour une visualisation améliorée 36
2.3.4 Segmentation des anévrismes pour le suivi du traitement 37

L'exploration du contexte clinique associé aux anévrismes intracrâniens offre un aperçu pour comprendre ces anomalies vasculaires et déterminer les exigences auxquelles doit répondre un outil d'assistance destiné aux radiologues. La section 2.1 se consacre à l'examen de la structure anatomique et physiologique des composants cérébraux et vasculaires, fournissant une base pour la compréhension de la formation des anévrismes et des facteurs de risque associés à leur rupture. Cela met en évidence l'importance de la détection précoce pour prévenir les graves conséquences d'une rupture éventuelle. La section 2.2 passe en revue les techniques d'imagerie médicale employées pour la visualisation et l'analyse des structures vasculaires en 3D. Parmi ces modalités, on compte l'angiographie par tomodensitométrie (CTA), l'angiographie rotationnelle par rayons X (3DRA), et l'angiographie par résonance magnétique (MRA). Au cœur de cette section, les avantages intrinsèques de l'angiographie par résonance magnétique en temps de vol (TOF-MRA) en tant qu'approche non invasive et non ionisante sont présentés. En outre, nous discutons des limites de ces techniques en ce qui concerne la visualisation des petites structures vasculaires, tels que les anévrismes intracrâniens. Enfin, la section 2.3 se consacre à une description du processus de diagnostic et de prise en charge des anévrismes intracrâniens, conformément aux protocoles cliniques en vigueur. Plus particulièrement, nous exposons les étapes suivies par les radiologues, à savoir la détection, la localisation, la visualisation optimale et la segmentation des anévrismes intracrâniens, chaque étape étant adaptée à des besoins spécifiques.

2.1 Anatomie et physiopathologie des anévrismes intracrâniens

2.1.1 Structure et fonction du cerveau et des vaisseaux sanguins

Cerveau humain

Le cerveau, une structure d'une complexité remarquable, se positionne au cœur de la physiologie du corps humain. Il assume un rôle dans la coordination et la régulation d'une vaste gamme de fonctions intrinsèques à l'existence humaine. Plus précisément, il se distingue comme le pivot de la cognition, des émotions et du contrôle moteur, conférant ainsi la capacité d'accomplir des processus cognitifs, de manifester des réponses émotionnelles et de gouverner les mouvements volontaires [298].

La structure du cerveau humain se divise en quatre régions principales, appelées lobes cérébraux : le lobe frontal, le lobe temporal, le lobe pariétal et le lobe occipital. Chacun de ces lobes présente des spécialisations fonctionnelles distinctes [255]. Le lobe frontal, situé à l'avant du cerveau, est associé à des fonctions telles que la parole, le langage, le raisonnement, la planification et la gestion des mouvements des membres. Il est souvent considéré comme le centre de la cognition supérieure et de la prise de décision. Le lobe temporal, situé près des tempes, est impliqué dans des processus tels que le traitement du langage, la mémoire et la régulation émotionnelle. Le lobe pariétal, situé en haut et à l'arrière du cerveau, est responsable du traitement des informations sensorielles, de la coordination des mouvements et de la perception spatiale. Enfin, le lobe occipital, situé à l'arrière du cerveau, est principalement dédié au traitement visuel, permettant l'interprétation des signaux visuels reçus par les yeux pour la perception et la compréhension de l'environnement. La coordination de ces lobes permet la perception, la pensée, la mémoire, la prise de décision et l'interaction avec l'environnement.

Le cerveau est composé d'un réseau de neurones interconnectés, comptant environ 100 milliards de neurones [121]. Ces connexions permettent une transmission rapide et sélective des informations entre les différentes régions cérébrales. Chaque neurone est hautement spécialisé et joue un rôle dans la formation de circuits électrochimiques nécessaires aux fonctions cérébrales. Outre les neurones, le cerveau héberge des cellules gliales qui apportent un soutien à l'activité neuronale en maintenant l'équilibre chimique du micro-environnement cérébral. De plus, elles contribuent à la nutrition des neurones en fournissant des nutriments et de l'oxygène tout en éliminant les déchets métaboliques. Cet approvisionnement en nutriments essentiels et en oxygène, ainsi que l'évacuation des déchets métaboliques, dépendent d'un réseau de vaisseaux sanguins cérébraux [124]. Cette collaboration étroite entre les neurones, les cellules gliales et les vaisseaux sanguins maintient un environnement cérébral optimal, garantissant que le cerveau puisse traiter l'information et remplir ses fonctions.

Vaisseaux sanguins cérébraux et polygone de Willis

Le système vasculaire du cerveau humain est un réseau complexe composé d'artères, de veines et de capillaires, chacun remplissant des rôles spécifiques au sein de la circulation sanguine (voir Figure 2.1). Les artères, les vaisseaux les plus larges, atteignent jusqu'à 6 mm de diamètre pour les plus importantes.

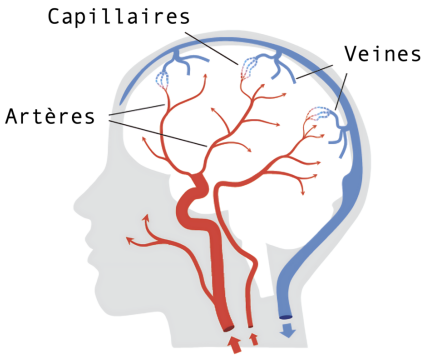


FIGURE 2.1 – Système vasculaire cérébral [2] : artères, veines et capillaires.

Ils acheminent le sang oxygéné et nutritif vers le cerveau, assurant ainsi les éléments essentiels au fonctionnement cellulaire cérébral. Les veines, quant à elles, ont des diamètres compris entre 0,4 et 3,5 mm. Elles facilitent le retour du sang appauvri en oxygène et chargé de produits métaboliques vers le cœur en vue de sa réoxygénation. Enfin, les capillaires, les vaisseaux sanguins les plus minuscules du cerveau, présentent un diamètre variant de 5 à 10 µm. Ils établissent la lien entre les artères et les veines, formant ainsi un réseau dense et étroitement entrelacé à travers les tissus cérébraux.

Les diverses composantes du système vasculaire cérébral, de tailles et densités variables, interagissent de manière coordonnée pour répondre aux besoins métaboliques des cellules cérébrales. À chaque battement cardiaque, le cerveau mobilise environ 15 à 20% du débit sanguin total du corps [315]. L'approvisionnement artériel du cerveau provient des artères carotides et vertébrales (voir Figure 2.2). Les artères carotides, situées de chaque côté du cou, se subdivisent en artères carotides externes et internes. Les artères carotides externes (ECA) forment un réseau ramifié destiné à vasculariser les structures faciales et cervicales supérieures. À l'intérieur de la glande parotide, les ECA se divisent en artère temporale superficielle et artère maxillaire. Les artères carotides internes (ICA) assurent l'irrigation de la majeure partie du cerveau, de l'oreille interne et de l'œil. Ces artères se subdivisent en plusieurs branches, dont l'artère ophtalmique, dédiée à l'irrigation des yeux. Les artères cérébrales antérieures (ACA), interconnectées par l'artère communicante antérieure (ACom), fournissent le flux sanguin au lobe frontal, au lobe pariétal et au lobe temporal. Les artères cérébrales moyennes (MCA) assurent l'irrigation des côtés des lobes temporaux et pariétaux. Enfin, les artères communicantes postérieures (PCom) rejoignent la circulation vertébro-basilaire au niveau des artères cérébrales postérieures (PCA).

D'autre part, les artères vertébrales (VA) sont situées à droite et à gauche dans la région cervicale de la colonne vertébrale. Les VA se divisent en deux branches principales : les artères cérébelleuses inféro-postérieures et les artères spinales. Elles se rejoignent ensuite au niveau de l'artère (ou tronc) basilaire. Les artères communicantes antérieures (ACom) et postérieures (PCom), établissant des connexions entre les circulations carotidiennes et vertébro-basilaire, forment un réseau appelé polygone de Willis.

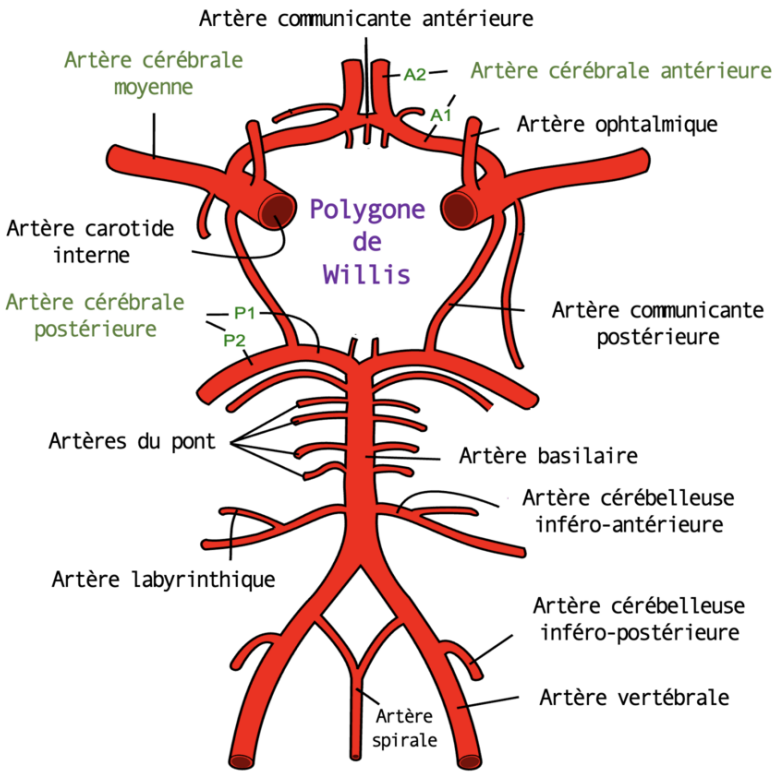


FIGURE 2.2 – Principales artères cérébrales : le polygone de Willis est situé à la base du cerveau.

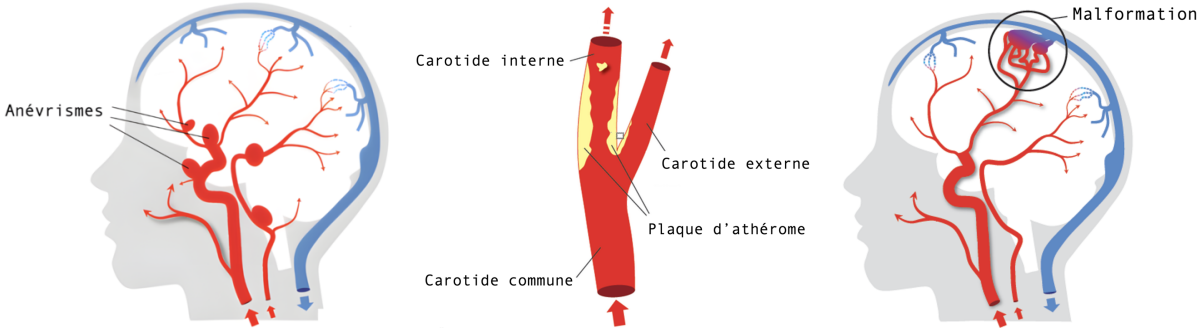
Le polygone de Willis constitue un disposition circulaire d'artères situé à la base du cerveau. Sa fonction principale consiste à assurer la continuité et la résilience de la circulation sanguine au sein du réseau vasculaire cérébral, facilitant ainsi la distribution d'oxygène et de nutriments dans différentes régions cérébrales. Les configurations du polygone de Willis varient d'un individu à l'autre, en termes de nombre, de forme et de taille des artères [149, 338]. Cette structure anatomique agit comme un mécanisme protecteur en cas de perturbations du flux sanguin, telles qu'une obstruction artérielle ou une réduction du débit. En cas d'obstruction d'une artère, par exemple, le polygone de Willis offre une voie alternative au sang pour atteindre les zones cérébrales souffrant d'un approvisionnement sanguin déficient.

Affections vasculaires

Le fonctionnement optimal du cerveau est étroitement lié à la santé de son système vasculaire, lequel assure un apport sanguin adéquat aux différentes régions cérébrales. Cependant, bien que ce système démontre une grande résilience et une capacité d'adaptation, il demeure vulnérable aux perturbations provoquées par diverses affections vasculaires. Parmi ces affections, les accidents vasculaires cérébraux (AVC, ou *Stroke* en anglais), suscitent des inquiétudes particulières en raison de leurs graves répercussions sur la santé des individus touchés [128].

Les AVC se caractérisent par une perturbation soudaine de l'apport sanguin au cerveau, qui survient généralement suite à l'obstruction d'un vaisseau sanguin (AVC ischémique, environ 80% des AVC) ou à sa rupture (AVC hémorragique, environ 20% des cas d'AVC) [38]. Cette perturbation entraîne une privation d'oxygène et de nutriments pour les cellules cérébrales de la région affectée. Les AVC se positionnent en tant que première cause de handicap physique acquis chez les adultes, la deuxième cause de démence, et la deuxième cause de mortalité, avec un taux de mortalité de 20% dans l'année suivant un AVC [299]. Notamment, il convient de souligner que les AVC hémorragiques présentent généralement une gravité supérieure par rapport aux AVC ischémiques [7].

Diverses anomalies vasculaires sont liées à des perturbations potentielles de la circulation sanguine cérébrale, accroissant le risque d'AVC et de leurs conséquences. Parmi ces anomalies, les anévrismes intracrâniens se démarquent par une dilatation anormale de la paroi d'une artère cérébrale, créant une saillie similaire à un petit ballon et liée à un risque de rupture (voir Figure 2.3a). Touchant jusqu'à 6% de la population mondiale, les anévrismes intracrâniens représentent environ 5% de tous les cas d'AVC [89]. Les sténoses cérébrales, quant à elles, se caractérisent par le rétrécissement des artères, généralement attribué à l'athérosclérose, formant une plaque athéromateuse qui s'épaissit progressivement (voir Figure 2.3b). Par conséquent, elles réduisent la circulation sanguine au niveau de l'artère touchée et constituent l'une des principales causes d'AVC ischémique à l'échelle mondiale [107]. Enfin, les malformations artério-veineuses (MAV) sont définies par une connexion anormale dans le réseau capillaire cérébral, se développant généralement après la naissance (voir Figure 2.3c). Leur principale conséquence est une augmentation anormale du débit sanguin et de la pression dans les veines, perturbant la circulation sanguine et augmentant le risque d'AVC hémorragique. Les malformations artério-veineuses sont relativement rares par rapport aux autres anomalies, touchant environ 0,1% de la population mondiale et représentant jusqu'à 2% de tous les cas d'AVC [6].



(a) Anévrismes intracrâniens (b) Sténose (c) Malformations artério-veineuses

FIGURE 2.3 – Principales affections cérébro-vasculaires [2].

Chacune de ces anomalies vasculaires expose les individus affectés à des risques spécifiques qui peuvent compromettre leur santé cérébrale. Il est impératif de procéder à une détection précoce et à des traitements appropriés pour ces affections. Toutefois, parmi ces anomalies, les anévrismes intracrâniens se distinguent par leur prévalence élevée au sein de la population mondiale. Ils sont associés à un risque de rupture, provoquant une hémorragie avec des taux de mortalité et de morbidité considérablement élevés [211]. Leur caractère souvent asymptomatique souligne la nécessité de mettre en place des stratégies spécifiques pour le diagnostic de cette condition particulière avant qu'elle n'atteigne des stades de développement avancés. Ainsi, la compréhension de la pathologie et des facteurs de risque liés aux anévrismes intracrâniens contribue à améliorer la gestion clinique et à prévenir les conséquences néfastes pour la santé.

2.1.2 Caractéristiques des anévrismes intracrâniens

Anévrismes et facteurs de formation

Les anévrismes intracrâniens, également appelés anévrismes cérébraux, se caractérisent par une dilatation localisée d'une artère du cerveau. Cette dilatation résulte principalement de la fragilité des parois vasculaires, qui est souvent causée par la pression exercée par le flux sanguin. Ces anévrismes sont particulièrement dangereux, car ils peuvent provoquer la rupture de l'artère dilatée, conduisant à une hémorragie appelée hémorragie sous-arachnoïdienne (HSA). Cette condition est associée à des taux de morbidité très élevés, ainsi qu'à un taux de mortalité d'environ 40% [158].

La formation et le développement des anévrismes intracrâniens sont des processus complexes et ne sont pas encore complètement élucidés. Ils sont influencés par des facteurs hémodynamiques, génétiques et anatomiques [227, 259]. Les forces hémodynamiques exercées par le flux sanguin sur la paroi des vaisseaux jouent un rôle crucial dans leur développement. En particulier, dans les zones de bifurcation, les variations de pression et de débit sanguin peuvent induire une turbulence et des contraintes excessives sur la paroi artérielle, favorisant ainsi sa détérioration progressive. De plus, des mutations génétiques contribuent à la susceptibilité individuelle à développer des anévrismes. Environ 20% des personnes atteintes d'un anévrisme intracrânien ont des antécédents familiaux [44].

Les caractéristiques anatomiques des vaisseaux cérébraux peuvent également prédisposer à la formation d'anévrismes. Certains individus peuvent avoir des vaisseaux sanguins avec des architectures particulières, telles qu'une paroi plus fine ou une disposition anormale des couches vasculaires, favorisant ainsi la formation d'anévrismes [248]. D'autres facteurs, comme l'hypertension, l'âge, le sexe, la consommation d'alcool et le tabagisme, contribuent également à la formation de ces affections vasculaires.

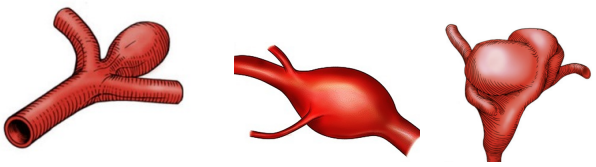
Tailles et formes

Les anévrismes intracrâniens se caractérisent par une variété de tailles millimétriques, mesurées en fonction de leur diamètre, qui peut varier de moins de 1 mm à jusqu'à 50 mm [137]. Au CHRU de Nancy, la majorité des cas diagnostiqués présente généralement des tailles inférieures à 15 mm, ce qui permet de les regrouper en trois catégories principales : les très petits anévrismes (moins de 3 mm), les anévrismes de taille moyenne (entre 3 et 5 mm) et les grands anévrismes (plus de 5 mm).

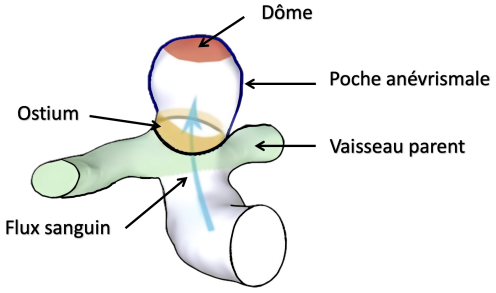
En ce qui concerne leur forme, les anévrismes se divisent principalement en deux catégories en fonction de leur emplacement dans l'arbre vasculaire : les anévrismes sacculaires et les anévrismes fusiformes [231]. Les anévrismes sacculaires, également connus sous le nom d'anévrismes en baies ou en sacs, prédominent parmi les cas d'anévrismes intracrâniens, représentant environ 80% à 90% des cas diagnostiqués [33, 90]. Ils se caractérisent par la formation de petites excroissances liées à l'artère, émanant d'une zone de fragilité dans la paroi artérielle (voir Figure 2.4a, gauche). En revanche, les anévrismes fusiformes se distinguent par une dilatation uniforme et progressive de segments étendus des artères concernées, sans la formation d'une poche discrète (voir Figure 2.4a, milieu). Il convient également de noter que certains anévrismes peuvent présenter des caractéristiques mixtes, combinant à la fois les aspects des anévrismes sacculaires et fusiformes (voir Figure 2.4a, droite).

La morphologie d'un anévrisme intracrânien est principalement définie par deux éléments : l'ostium, situé au collet de l'anévrisme, et le dôme (voir Figure 2.4b). L'ostium représente le point d'origine de l'anévrisme, marquant la connexion entre celui-ci et le vaisseau parent à partir duquel il se développe. D'autre part, le dôme constitue la partie affaiblie et dilatée de l'anévrisme, s'étendant au-delà de la paroi

de l'artère. Sous l'influence du flux sanguin entrant, le dôme émerge comme la zone la plus vulnérable de l'anévrisme, exposant ainsi l'anévrisme à un risque de rupture et à un potentiel saignement dans le tissu cérébral environnant. Cette caractérisation de la morphologie des anévrismes intracrâniens est cruciale pour l'évaluation du risque de rupture et la planification des interventions thérapeutiques.



(a) Formes les plus courantes d'anévrismes.



(b) Structure de l'anévrisme.

FIGURE 2.4 – (a) Les principales formes d'anévrismes cérébraux : sacculaire (gauche), fusiforme (milieu), et mixte (droite). (b) Illustration de la structure d'un anévrisme de forme sacculaire : Le dôme, situé à l'opposé de l'ostium, sépare la poche anévrismale du vaisseau parent [212].

Localisation

Les anévrismes intracrâniens peuvent émerger à divers endroits le long des artères cérébrales, chaque site engendrant une forme spécifique d'anévrisme. Les bifurcations des grandes artères cérébrales, où les vaisseaux se ramifient en branches plus petites, sont fréquemment liées à la formation d'anévrismes, étant soumises à d'importantes contraintes hémodynamiques [259]. En raison de sa localisation centrale dans le cerveau, le polygone de Willis est soumis de manière significative aux forces hémodynamiques générées par le flux sanguin. Cette exposition favorise la dilatation des artères, ce qui peut contribuer à la formation d'anévrismes. Environ 85% de tous les anévrismes intracrâniens se développent au niveau du polygone de Willis [208].

La fréquence de formation des anévrismes varie selon les régions vasculaires. En général, 90% des anévrismes sont situés sur la circulation antérieure : entre 30 et 35% sur l'artère communicante antérieure (ACom), entre 30 et 35% sur le siphon carotidien, et environ 20% sur la bifurcation de l'artère cérébrale moyenne (ACA). D'autre part, 10% se localisent sur la circulation vertébro-basilaire (termino-basilaire et artère cérébelleuse inféro-postérieure). La localisation des anévrismes et l'impact des forces hémodynamiques peuvent moduler leur expansion, affectant ainsi le risque de rupture qui leur est associé.

2.1.3 Facteurs de risque de rupture liés aux anévrismes intracrâniens

Le processus évolutif de la formation des anévrismes intracrâniens jusqu'à leur rupture s'étend sur plusieurs années, débutant par une prédisposition congénitale ou acquise de la paroi artérielle, rendant les vaisseaux plus sensibles aux forces du flux sanguin. Au fil du temps, ces contraintes entraînent des altérations pathologiques telles que l'amincissement de la paroi, la détérioration de son architecture, et la diminution de sa résistance aux contraintes du flux sanguin. Ces changements favorisent la croissance et le développement des anévrismes, augmentant ainsi leur vulnérabilité à la rupture. Il est estimé que 50 à 80% des anévrismes intracrâniens demeurent intacts tout au long de la vie d'un individu [43]. En conséquence, entre 20 et 50% se rompent, entraînant une hémorragie sous-arachnoïdienne (HSA). Environ 25% des personnes décèdent dans les 24 heures suivant la rupture, un chiffre qui atteint 50% dans les trois mois qui suivent [143].

Le diagnostic précoce et la surveillance des anévrismes avant leur rupture revêtent une importance capitale. Cette vigilance s'accompagne de l'évaluation spécifique du risque de rupture pour chaque cas. Ce risque est influencé par divers facteurs, incluant des aspects inhérents à l'anévrisme lui-même et des éléments liés au patient, incluant :

- **Taille et forme :** La taille des anévrismes constitue un paramètre essentiel dans l'évaluation de leur risque de rupture. En règle générale, les anévrismes de petite taille, mesurant moins de 5 mm, présentent un risque de rupture faible et peuvent être surveillés sans nécessiter un traitement immédiat, sauf en cas de signes de croissance ou de développement. Toutefois, à mesure que la taille de l'anévrisme augmente, le risque de rupture s'accroît considérablement. Les anévrismes de plus grande taille deviennent ainsi plus vulnérables à la rupture, car la pression exercée par le flux sanguin sur leur paroi augmente proportionnellement, affaiblissant leur structure vasculaire. D'autre part, la forme de l'anévrisme joue également un rôle déterminant dans l'évaluation du risque de rupture. Les anévrismes sacculaires, en raison de leur paroi plus mince, présentent un risque plus élevé de rupture par rapport aux anévrismes fusiformes [247].
- **Rapport de taille :** La prise en compte du rapport de taille est également pertinente dans l'évaluation d'un anévrisme. Ce rapport correspond à la mesure de son diamètre divisé par celui de son collet, c'est-à-dire sa connexion au vaisseau parent. Un rapport de taille élevé constitue un indicateur de la susceptibilité de l'anévrisme à la rupture [252].
- **Rapport d'aspect :** Un autre paramètre à considérer est le rapport d'aspect d'un anévrisme, calculé en divisant sa largeur par son diamètre. Un rapport d'aspect élevé indique que l'anévrisme présente une forme plus allongée ou irrégulière, tandis qu'un rapport d'aspect faible suggère une forme plus sphérique ou arrondie. Les anévrismes avec un rapport d'aspect élevé sont généralement considérés comme moins stables et présentent un risque accru de rupture [252].
- **Localisation et angle d'écoulement :** L'emplacement des anévrismes peut avoir une influence sur leur croissance et, par conséquent, sur leur risque de rupture. Les anévrismes localisés dans le polygone de Willis sont associés à un risque plus élevé de rupture en raison de la circulation sanguine intense dans cette région vasculaire [208]. De plus, la position relative de l'anévrisme par rapport aux autres structures cérébrales et l'angle d'écoulement de l'anévrisme peuvent également avoir un impact sur le risque de rupture [20].
- **Autres facteurs :** D'autres facteurs peuvent également influencer le risque de rupture des anévrismes, en plus de leurs caractéristiques morphologiques. Parmi ces facteurs, on peut citer la présence d'anévrismes multiples, la vitesse de croissance de l'anévrisme, et des éléments liés au patient, tels que son origine ethnique, son âge, ses antécédents de rupture d'anévrisme, les antécédents familiaux, ainsi que la présence d'hypertension artérielle.

En se basant sur ces divers facteurs de risque de rupture, les cliniciens évaluent la gravité de chaque anévrisme dans le cadre de la prise en charge clinique. Pour faciliter cette évaluation, ils recourent fréquemment des modèles tels que le score PHASES [34] et le score UIATS [88]. Le score PHASES évalue des paramètres liés au patient, comme l'origine ethnique, l'hypertension, et l'âge, ainsi que des paramètres liés à l'anévrisme, tels que la taille, la localisation et les antécédents de rupture, pour fournir une évaluation du risque de rupture sur les 5 prochaines années. Cependant, il présente des limites, notamment la sous-représentation de certains groupes de patients (jeunes et fumeurs) et l'absence de prise en compte des facteurs de risque liés au traitement. D'autre part, le score UIATS, utilisé au CHRU de Nancy, offre une analyse plus approfondie. Il intègre des facteurs de risque liés au traitement et des éléments supplémentaires concernant à la fois l'anévrisme (vitesse de croissance, forme, rapport d'aspect, etc.) et le patient (espérance de vie, symptômes cliniques, comorbidité, etc.). Si l'anévrisme est considéré comme présentant un faible risque de rupture, le patient fera l'objet d'un suivi régulier, généralement entre 6 mois et 5 ans. En revanche, si l'anévrisme est jugé à risque élevé de rupture, une intervention préventive sera planifiée pour prévenir toute éventuelle rupture.

2.1.4 Options de traitement des anévrismes intracrâniens

Deux options de traitement fréquemment employées pour traiter les anévrismes intracrâniens sont la pose d'un clip neurochirurgical et l'embolisation endovasculaire [204]. Ces procédures visent à isoler l'anévrisme de la circulation sanguine afin de minimiser les risques d'expansion et, par conséquent, de rupture. La pose d'un clip, également appelée "clipping", requiert une craniotomie pour accéder à l'anévrisme, suivie par la mise en place d'un clip métallique autour de son collet (voir Figure 2.5a). L'objectif est d'interrompre le flux sanguin vers l'anévrisme, l'excluant ainsi de la circulation normale. Une fois le

clip en place, la portion du crâne retirée est soigneusement remise en position et fixée à l'aide de plaques et de vis spéciales, permettant ainsi une récupération optimale et une protection adéquate du cerveau. Cette procédure exige une grande précision et une expertise en neurochirurgie.

En comparaison, l'embolisation endovasculaire, également appelée "coiling", est une procédure moins invasive. Elle implique l'insertion d'un cathéter dans une artère périphérique, généralement l'artère fémorale, pour être ensuite guidé jusqu'à l'anévrisme (voir Figure 2.5b). Des coils (petites bobines en platine) de différentes tailles et formes sont déployés à l'intérieur de la poche anévrismale pour favoriser la formation d'un caillot, excluant ainsi l'anévrisme de la circulation sanguine. Une fois l'embolisation terminée, le cathéter est retiré et le vaisseau sanguin est surveillé attentivement. Cette technique offre un temps de récupération plus court et réduit les risques associés à l'intervention chirurgicale.



(a) Pose d'un clip métallique au collet de l'anévrisme (b) Utilisation de coils pour isoler l'anévrisme

FIGURE 2.5 – Options de traitement des anévrismes intracrâniens.

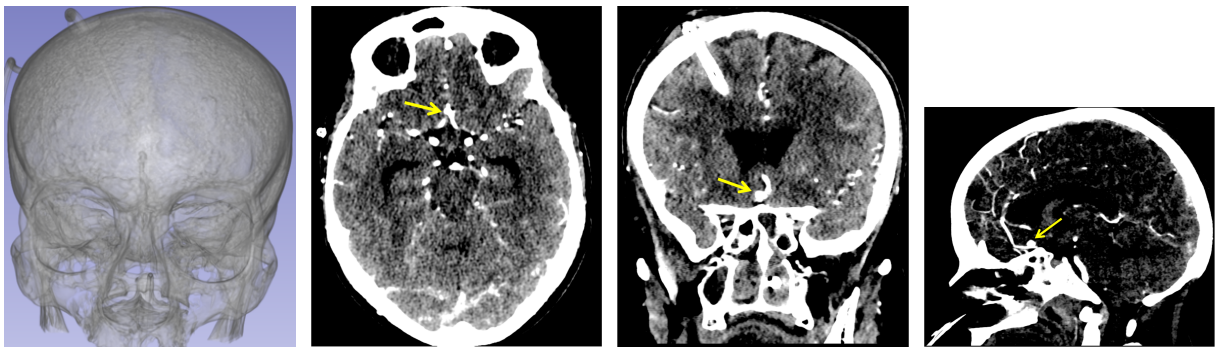
Le choix entre la pose d'un clip neurochirurgical et l'embolisation endovasculaire pour le traitement des anévrismes intracrâniens nécessite une collaboration étroite entre le neurochirurgien et le neuroradiologue interventionnel. Cette décision doit tenir compte de l'expertise des équipes médicales, des caractéristiques spécifiques de l'anévrisme et du patient, ainsi que des risques associés à chaque option de traitement. L'embolisation endovasculaire, en raison de sa moindre invasivité et d'une récupération plus rapide, est privilégiée pour les anévrismes de petite taille, les formes irrégulières, et les localisations difficiles d'accès par la pose de clip. Les anévrismes présentant un large collet dans des zones de bifurcations vasculaires ou de parois latérales peuvent poser des défis pour les coils et sont traités avec de nouveaux dispositifs d'occlusion tels que le "Woven EndoBridge" (WEB) [14]. Toutefois, l'embolisation endovasculaire peut présenter des limites d'efficacité dans certains cas, accroissant le risque de récidive ou de rupture ultérieure. La pose d'un clip neurochirurgical peut être préférée pour les anévrismes de grande taille ou de forme particulière en raison de son potentiel à isoler plus complètement l'anévrisme. Après la procédure de traitement, une surveillance régulière s'impose pour évaluer l'efficacité du traitement, la stabilité de l'anévrisme au fil du temps, et détecter précocement d'éventuelles complications postopératoires.

2.2 Imagerie médicale pour la visualisation des anévrismes intracrâniens

La détection, l'évaluation, le traitement et la surveillance des anévrismes intracrâniens reposent sur des techniques avancées d'imagerie angiographique, permettant la visualisation précise des vaisseaux sanguins, incluant les anévrismes intracrâniens. L'angiographie par tomodensitométrie (CTA) et l'angiographie par résonance magnétique en temps de vol (TOF-MRA) ont progressivement remplacé l'angiographie numérique soustraite (DSA) dans le domaine du diagnostic. Cette dernière est principalement réservée à la thérapie, complétée par la fluoroscopie pour un suivi en temps réel et l'angiographie rotationnelle par rayons X (3DRA) pour une analyse tridimensionnelle. Cette section se focalise spécifiquement sur l'utilisation des modalités d'imagerie tridimensionnelles (CTA, 3DRA, et TOF-MRA) dans le contexte de la visualisation des anévrismes intracrâniens. L'objectif est de mettre en évidence leurs avantages, leurs limites et leurs applications cliniques.

2.2.1 Angiographie par tomodensitométrie (CTA)

L'angiographie par tomodensitométrie, également appelée angioscanner, combine les principes de la tomodensitométrie (CT) et de l'angiographie pour la visualisation des vaisseaux sanguins et des tissus spécifiques dans une région anatomique, telle que le cerveau. Cette technique repose sur l'utilisation d'un scanner émettant des rayons X, conjointement avec l'administration intraveineuse d'un agent de contraste iodé. Étant donné que les rayons X ne permettent pas une visualisation directe des vaisseaux sanguins, un agent de contraste iodé, doté d'un numéro atomique élevé, est requis. Ce produit améliore la visibilité des vaisseaux lors de l'imagerie par rayons X. Après l'injection, une série d'images en coupes transversales est acquise à intervalles réguliers. Un laps de temps est prévu pour permettre la diffusion de l'agent de contraste dans la circulation sanguine cérébrale, facilitant ainsi la visualisation détaillée de l'arbre vasculaire, y compris des anévrismes intracrâniens (voir un exemple d'image en Figure 2.6).



(a) Rendu du volume (b) Axial (c) Sagittal (d) Coronal

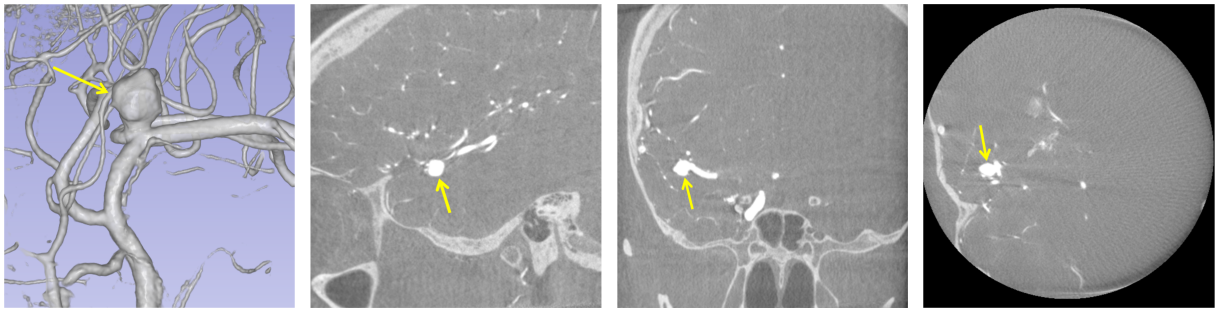
FIGURE 2.6 – Exemple d'image d'angiographie par tomodensitométrie (CTA) présentant la vue en 3D (a) ainsi que les plans de coupes 2D (b, c, d). L'anévrisme intracrânien, mesurant 4,18 mm, est identifié par la flèche sur les plans de coupe. Les vaisseaux sanguins sont difficile à visualiser dans la vue 3D.

La CTA présente un avantage majeur grâce à son acquisition rapide des images, généralement en quelques secondes, ce qui en fait une option précieuse pour les cas aigus ou d'urgence [264]. De plus, elle est largement accessible dans de nombreux établissements médicaux, en faisant une modalité d'imagerie couramment disponible. Cependant, elle comporte des limites à prendre en considération. Tout d'abord, la CTA expose les patients aux rayonnements ionisants des rayons X, avec des implications potentielles à long terme pour la santé. De plus, l'injection d'agent de contraste iodé comporte des inconvénients tels que le risque de réactions allergiques et d'altération de la fonction rénale [8]. La qualité et la résolution spatiale des images en CTA dépendent de l'agent de contraste utilisé et de la vascularisation du patient, ce qui peut restreindre la précision de la visualisation des petites artères, surtout lorsque le contraste atteint certaines zones tardivement ou lorsque l'opacification des vaisseaux est insuffisante. De plus, les rayons X peuvent générer des artefacts en présence de structures corporelles denses comme les os ou les implants métalliques, entraînant du bruit dans les images. Enfin, la CTA peut avoir des limites dans la détection de structures de petite taille, pouvant ne pas être visualisées avec une grande précision [19].

2.2.2 Angiographie rotationnelle tridimensionnelle (3DRA)

La technique d'angiographie rotationnelle tridimensionnelle se distingue par sa résolution spatiale élevée parmi les modalités d'angiographie [295, 9]. Elle associe les rayons X à un agent de contraste iodé pour visualiser en détail les structures vasculaires dans une région spécifique. Cette approche s'appuie sur les principes de l'angiographie par rayons X 2D classique, tout en surmontant certaines de ses limitations, telles que les problèmes de superposition vasculaire et les morphologies complexes des vaisseaux sanguins. La particularité de la 3DRA réside dans sa capacité à capturer une séquence d'images tout en réalisant une rotation continue autour du crâne. Pour la réalisation d'une imagerie 3DRA, un agent de contraste iodé est administré via un cathéter inséré dans l'artère fémorale, guidé jusqu'aux artères cérébrales.

Pendant l'injection du produit de contraste, la source de rayons X effectue une rotation autour de la tête du patient, acquérant ainsi un ensemble d'images 2D prises sous divers angles et à intervalles réguliers. Ces images sont ensuite reconstituées par un logiciel de reconstruction pour former une représentation 3D des vaisseaux sanguins dans la région ciblée (voir un exemple d'image en Figure 2.7).



(a) Rendu du volume (b) Axial (c) Sagittal (d) Coronal

FIGURE 2.7 – Exemple d'image d'angiographie rotationnelle par rayons X (3DRA) illustrant la vue en 3D (a) et les plans de coupes 2D (b, c, d). L'anévrisme intracrânien de 6,83 mm est indiqué par la flèche.

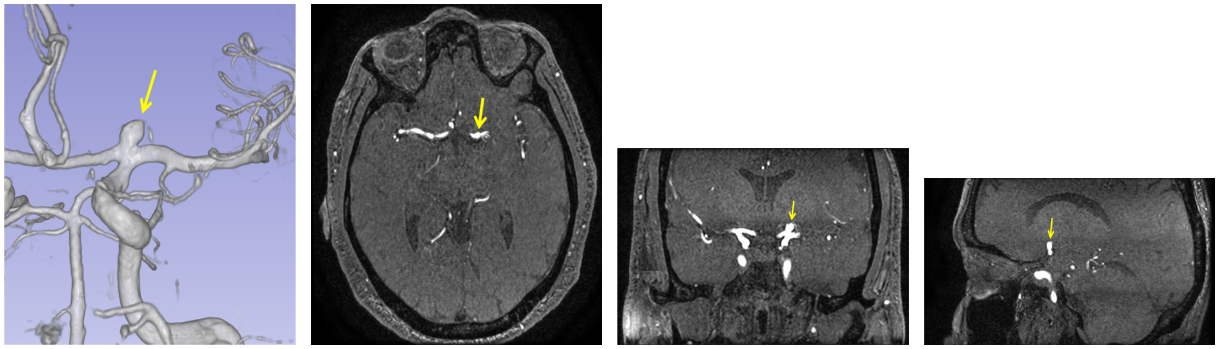
La 3DRA présente des avantages indéniables, offrant des représentations détaillées des structures vasculaires, y compris les vaisseaux de petite taille et les anomalies vasculaires [295]. Les images produites sont de haute qualité, présentant un minimum de bruit et d'artefacts. Sa rapidité d'acquisition, généralement entre 5 et 10 secondes, en fait un choix privilégié pour l'imagerie vasculaire en salle d'intervention. Cependant, ces avantages sont nuancés par diverses contraintes. Tout d'abord, la 3DRA est sélective, ce qui signifie qu'elle se concentre sur une zone spécifique des vaisseaux sanguins. Cette caractéristique peut poser des limites dans des applications diagnostiques qui exigent une vision globale de l'arbre vasculaire. De plus, bien que rares, il existe des risques théoriques de complications iatrogènes neurologiques (thromboemboliques, dissections) ainsi que des complications au point de ponction. En outre, l'accès à la 3DRA peut être limité en raison des coûts élevés associés à l'équipement, généralement disponible dans les milieux hospitaliers spécialisés. Par ailleurs, l'exposition aux radiations et l'administration de produits de contraste iodé représentent des contraintes majeures. Enfin, une autre limitation de la 3DRA réside dans sa capacité relativement limitée à différencier les tissus mous par rapport à certaines autres modalités d'imagerie, ce qui peut entraîner des représentations anatomiques moins précises.

2.2.3 Angiographie par résonance magnétique (MRA)

L'imagerie par résonance magnétique (IRM) représente une technique avancée en imagerie médicale, exploitant les propriétés magnétiques des atomes d'hydrogène dans les tissus biologiques. Cette technique propose une variété de séquences, comprenant les séquences pondérées en T1, en T2, FLAIR, ainsi que l'angiographie par résonance magnétique (MRA). Chacune caractérisée par des paramètres d'accentuation spécifiques, permettent une exploration des structures anatomiques. La MRA, spécifiquement dédiée à la visualisation des vaisseaux sanguins, est largement préférée pour l'évaluation vasculaire en raison de son principal avantage : l'utilisation d'un champ magnétique plutôt que de rayonnements ionisants, permettant des examens sans agent de contraste. Son principe repose sur le comportement des protons dans les atomes d'hydrogène présents dans le sang circulant, créant ainsi un contraste dans les images [306, 165]. Plusieurs modalités de MRA existent, chacune se démarquant par des approches spécifiques et des applications cliniques variées [114]. Parmi elles, les deux modalités les plus fréquemment employées sont la MRA à temps de vol (TOF-MRA) et la MRA avec renforcement du contraste (CE-MRA). Toutefois, la TOF-MRA est privilégiée pour le dépistage en raison de son caractère non invasif, éliminant ainsi la nécessité d'utiliser un agent de contraste, contrairement à la CE-MRA.

Pour réaliser une imagerie TOF-MRA du cerveau, des antennes spécifiques sont positionnées autour de la tête afin de capter les signaux radiofréquences émis par les tissus cérébraux. L'appareil d'IRM génère un champ magnétique statique mesuré en unités de Tesla (T), alignant les protons dans les tissus

corporels, y compris le sang circulant dans les vaisseaux sanguins, dans la direction du champ magnétique. Ces protons agissent comme de minuscules aimants. Des bobines spéciales, situées autour de la région d'intérêt, émettent des impulsions radiofréquence spécifiques, provoquant un basculement temporaire des spins des protons du sang. Lorsque l'impulsion radiofréquence est interrompue, les protons reviennent progressivement à leur état d'équilibre avec le champ magnétique statique. Cependant, en raison du mouvement du sang, les protons subissent un déphasage partiel lors des impulsions répétées, maintenant ainsi une contribution significative du signal de résonance magnétique détecté par le sang en mouvement. Ce phénomène est à la base du principe de l'angiographie par temps de vol. Les antennes de l'IRM captent ensuite les signaux de résonance magnétique émanant des tissus à différents moments après l'application des impulsions radiofréquence. Ces données sont recueillies sous forme de signaux, analysés en fonction de leur intensité, position spatiale, et temps de retour au signal d'équilibre pour la reconstruction d'images en coupes transversales (voir un exemple d'image TOF-MRA en Figure 2.8).



(a) Rendu du volume (b) Axial (c) Sagittal (d) Coronal

FIGURE 2.8 – Exemple d'image TOF-MRA montrant (a) la vue 3D et (b, c, d) les plans de coupes 2D. L'anévrisme intracrânien de 4,83 mm est indiqué par la flèche.

La technique d'imagerie TOF-MRA présente plusieurs avantages dans l'imagerie des vaisseaux sanguins. Son caractère non invasif constitue un atout majeur, éliminant ainsi l'exposition aux rayonnements ionisants et la nécessité d'utiliser des agents de contraste. Cette non-invasivité en fait une option privilégiée pour le dépistage des affections vasculaires cérébrales, réduisant les risques liés aux procédures invasives susceptibles de provoquer des réactions allergiques. Par ailleurs, la TOF-MRA est compatible avec tout matériel endovasculaire tel que les coils et les stents, présentant peu d'artefacts métalliques en pratique. Cependant, il est essentiel de noter certaines limites de cette technique. Son temps d'acquisition, généralement de quelques minutes, peut sembler relativement long par rapport à d'autres modalités d'imagerie. Par conséquent, des artefacts de mouvement du patient et du bruit peuvent parfois compromettre la qualité et la précision des images obtenues [203]. Bien que la TOF-MRA soit efficace pour visualiser la plupart des structures vasculaires, elle peut présenter des limites pour les structures très petites (< 3 mm) [140]. Dans de tels cas, l'utilisation de techniques complémentaires peut être nécessaire pour une visualisation plus précise. Toutefois, une configuration appropriée de l'équipement utilisé, notamment en ajustant le champ magnétique (3T, 5T, 7T), peut améliorer la visualisation de ces structures de dimensions réduites.

2.2.4 Comparaison des imageries 3D pour la visualisation des anévrismes

Chacune des techniques d'imagerie précédentes présente des avantages et des limites spécifiques. Le choix de la technique appropriée dépend de plusieurs facteurs, notamment la disponibilité des équipements au sein de l'établissement de santé, l'urgence de la situation, et des considérations liées à la sécurité du patient telles que l'exposition aux rayons X, l'utilisation d'agents de contraste, et le champ magnétique. De plus, l'objectif clinique, qu'il s'agisse d'un diagnostic ou d'un traitement, peut également influencer la sélection de la technique d'imagerie la plus adaptée. Une synthèse des caractéristiques propres à chaque technique est présentée dans la table 2.1.

	CTA	**3DRA**	**TOF-MRA**
Cas d'utilisation	Diagnostic	Traitement	Diagnostic
Invasive	+	++	−
Accessibilité	+++	+	++
Temps d'acquisition	$5-10$ sec	$5-10$ sec	$5-10$ min
Résolution	++	+++	+
Taille voxel (mm^3)	$0,2-0,5$	$0,1-0,4$	$0,3-0,5$
Dimension (voxels)	$512\times512\times200$	$512\times512\times502$	$512\times512\times254$
Qualité	++	+++	+
Artefacts	+++	−	+
Sensibilité (%)	$84-97,2$	$92,3-100$	$76,6-92$
Spécificité (%)	$85-100$	$93-100$	$87-92$

TABLE 2.1 – Comparaison des techniques d'imagerie médicale 3D largement utilisées pour la visualisation des anévrismes intracrâniens. Ces mesures peuvent varier en fonction de l'équipement utilisé.

Parmi les diverses techniques d'imagerie disponibles, la 3DRA se positionne comme la méthode de référence pour la visualisation des anévrismes intracrâniens, privilégiée dans les situations requérant une grande précision, surtout lors de mesures morphologiques pendant le traitement des anévrismes intracrâniens [9]. En contraste, la CTA et la TOF-MRA sont principalement utilisées à des fins diagnostiques et sont largement disponibles dans de nombreux établissements de santé. La CTA est souvent privilégiée en situations d'urgence en raison de sa brève durée d'acquisition, généralement entre 5 et 10 secondes. Mais la TOF-MRA se distingue par sa nature non invasive, ce qui en fait une option privilégiée pour les activités de dépistage. Cette technique élimine l'exposition aux rayonnements et l'utilisation de produits de contraste, réduisant ainsi les risques potentiels pour le patient.

La qualité et la résolution spatiale des images médicales jouent un rôle pour représenter avec précision les signaux des anévrismes intracrâniens. La technique 3DRA se distingue par sa haute résolution spatiale, généralement entre 0,1 et 0,4 mm^3/voxel, permettant une visualisation précise des structures vasculaires, y compris les petits anévrismes de moins de 3 mm. Sa capacité à se focaliser sur une zone spécifique du système vasculaire réduit les artefacts et le bruit, améliorant ainsi la fiabilité des mesures morphologiques des anévrismes, telles que la taille, le collet, le rapport d'aspect, et le volume [9]. Ces mesures sont utilisées dans le traitement des anévrismes, notamment pour évaluer la taille des coils nécessaires pour remplir et isoler la poche anévrismale. En comparaison, la CTA et la TOF-MRA présentent également une bonne résolution spatiale, bien que généralement inférieure à celle de la 3DRA. La CTA affiche une résolution entre 0,2 et 0,5 mm^3/voxel, tandis que la TOF-MRA se situe autour de 0,3 à 0,5 mm^3/voxel. Cependant, la CTA peut être plus sujette aux artefacts et au bruit, ce qui peut influencer la qualité des images obtenues. Pour la TOF-MRA, elle peut être sujette aux artefacts de flux, notamment dans le cas des anévrismes de grande taille. Il est toutefois important de noter que ces différentes caractéristiques peuvent varier en fonction de la configuration matérielle et logicielle de l'équipement utilisé.

Les anévrismes sont de petites structures vasculaires situées au sein d'images 3D de grandes dimensions, fournissant des informations visuelles détaillées. Dans un ensemble composé de plusieurs millions de voxels représentant les vaisseaux sanguins et les autres tissus cérébraux, les anévrismes se distinguent par quelques centaines de voxels. Pour visualiser et analyser ces structures, il est impératif de disposer d'images présentant une résolution spatiale et une qualité appropriées. Le taux de sensibilité, évaluant la capacité des techniques à représenter l'intégralité des anévrismes dans les images, varie en fonction de divers paramètres tels que l'équipement utilisé, la taille et la localisation des anévrismes [217, 196]. Parmi les techniques disponibles, la 3DRA et la CTA démontrent généralement une sensibilité supérieure à celle de la TOF-MRA. Plus précisément, la 3DRA peut détecter la grande majorité des anévrismes, avec une sensibilité oscillant entre 92,3 et 100%. En revanche, la TOF-MRA présente des taux de sensibilité compris entre 76,6 et 92% [217, 196]. En particulier, la TOF-MRA peut avoir des difficultés à détecter les anévrismes de moins de 3 mm [91]. Ces techniques d'imagerie ne sont pas exemptes de risque de faux positifs, où des structures telles que des vaisseaux sains, d'autres tissus cérébraux, des artefacts ou du bruit peuvent être incorrectement interprétés comme des anévrismes dans les images. Ces erreurs

réduisent la spécificité de ces techniques, qui varie de 87 à 92% pour la TOF-MRA [58, 196].

Chaque technique d'imagerie présente des avantages et des limites qui se manifestent particulièrement lors de l'exploration de structures millimétriques telles que les anévrismes. Souvent, une approche combinée de plusieurs modalités est adoptée pour capitaliser sur les forces spécifiques de chacune, garantissant une visualisation plus exhaustive et précise. Dans le contexte du traitement des anévrismes, la 3DRA est préférée pour des mesures précises, complétée par l'angiographie numérique soustraite (DSA). Malgré les défis potentiels de la TOF-MRA dans la visualisation des petits anévrismes (< 3 mm), généralement associés à un faible risque de rupture, son caractère non invasif en fait une option de choix pour le diagnostic et l'orientation thérapeutique. Les avancées continues dans les équipements d'imagerie, combinées à une configuration adaptée de ces techniques, ouvrent des perspectives d'amélioration continue de leurs performances.

2.3 Diagnostic des anévrismes intracrâniens

2.3.1 Processus de prise en charge

La prise en charge des patients atteints d'anévrismes intracrâniens peut être envisagée selon deux approches distinctes. La première approche intervient lorsque le patient présente des symptômes pouvant être liés à un anévrisme et se présente à l'hôpital, tels que des céphalées sévères, des altérations visuelles, des crises d'épilepsie, des étourdissements ou d'autres signes neurologiques. Les médecins entreprennent alors une évaluation clinique approfondie, collectant les antécédents médicaux et réalisant un examen physique. En se basant sur ces informations et les manifestations symptomatiques, les médecins peuvent suspecter la présence d'anévrismes intracrâniens. Pour confirmer cette hypothèse, un examen d'imagerie médicale est effectué. La deuxième approche, plus courante, repose sur des programmes de dépistage ou sur la découverte fortuite lors d'examens d'imagerie cérébrale réalisés pour d'autres raisons. Cette approche est facilitée par la démocratisation des équipements IRM, par l'amélioration de la qualité de l'imagerie et par la formation de radiologues compétents. L'objectif du dépistage est de détecter les anévrismes chez les individus avant l'apparition de symptômes évocateurs. Grâce à cette approche proactive, les professionnels de la santé ont considérablement augmenté le nombre d'anévrismes identifiés chaque année à un stade précoce.

La visualisation et l'évaluation des anévrismes intracrâniens reposent sur l'utilisation de modalités d'imagerie d'angiographie. Les radiologues s'engagent dans l'analyse et l'interprétation des images acquises, suivant des méthodologies variables selon les protocoles spécifiques du centre hospitalier et de leur formation. Ce processus englobe la détection, la localisation, l'optimisation de la vue d'analyse, et/ou la segmentation des anévrismes. Il permet une évaluation qualitative et quantitative des anévrismes, fournissant des informations sur leur nombre, leur emplacement, leurs caractéristiques morphologiques et leur gravité. Ce processus peut également être utilisé pour le suivi et l'évaluation des anévrismes traités.

2.3.2 Détection et localisation des anévrismes

L'approche de détection des anévrismes intracrâniens à partir des images médicales 3D varie selon les protocoles cliniques adoptés, généralement suivant une démarche en deux étapes (voir Figure 2.9). La première étape consiste à visualiser et analyser les vaisseaux sanguins dans une vue 3D. Les radiologues appliquent un seuillage basé sur l'intensité des images pour discriminer les vaisseaux sanguins des autres tissus cérébraux. Durant cette phase, les radiologues analysent la surface des vaisseaux à la recherche d'anomalies correspondant aux anévrismes, en portant une attention particulière aux zones où ces anomalies ont tendance à se former, telles que le polygone de Willis [208]. Bien que cette approche s'avère efficace pour détecter les anévrismes de grande taille, elle peut parfois négliger les cas d'anévrismes situés dans des régions moins typiques ou présentant des difficultés de visualisation, en particulier les anévrismes de petite taille (< 5 mm), entraînant ainsi des résultats de faux négatifs (FN). D'autre part, en raison de la complexité de l'anatomie cérébrale et de la présence d'artefacts de flux, notamment au niveau de la carotide interne, des faux positifs (FP) peuvent également se produire.

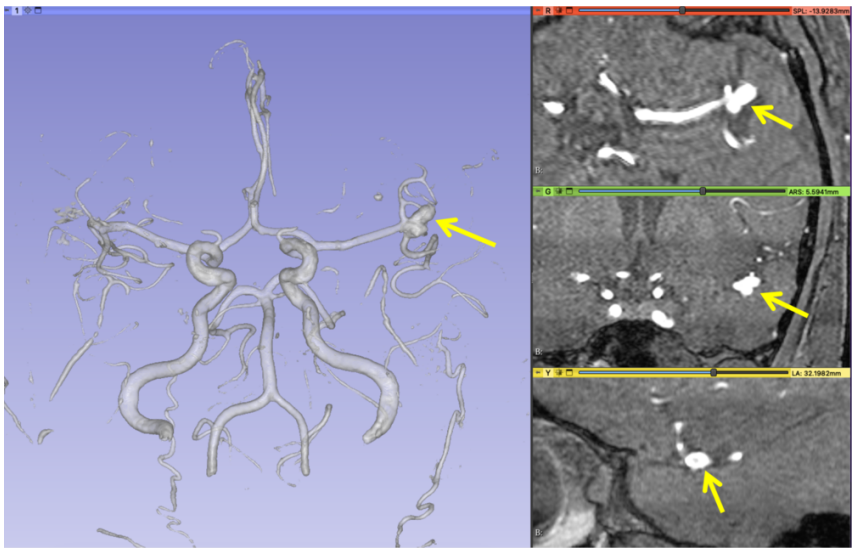


FIGURE 2.9 – Détection en deux étapes d'un anévrisme intracrânien de 6,61 mm. Première analyse globale en vue 3D (gauche), suivie d'une analyse locale plus détaillée en plans de coupes canoniques (droite).

Pour surmonter les limitations de la première étape, la seconde repose sur une analyse plus approfondie à l'aide de plans de coupes 2D canoniques (axial, sagittal et coronal). Ces plans de coupes 2D permettent aux radiologues d'explorer l'image selon différentes perspectives, améliorant ainsi la visualisation et la compréhension de l'arbre cérébro-vasculaire, y compris des structures moins visibles ou non visibles en vue 3D, telles que les petits vaisseaux. Pendant cette étape, les radiologues sélectionnent des images de coupe dans les différents plans, identifiant celles qui offrent une meilleure visualisation des anévrismes [80]. Par exemple, lors de l'examen du plan axial, les radiologues peuvent repérer les sites fréquents où les anévrismes ont tendance à se former. La vue sagittale permet de mieux visualiser des zones spécifiques, comme le tronc basilaire. La vue coronale se révèle utile pour inspecter des zones telles que les artères cérébrales antérieures ou les segments de l'artère cérébral moyenne.

Cette méthode d'analyse, évoluant de la visualisation 3D à une exploration plus ciblée en 2D, vise à perfectionner la détection des anévrismes intracrâniens par les radiologues. L'objectif global est d'améliorer simultanément la sensibilité et la spécificité de détection. La fusion d'informations provenant des vues en 3D et des coupes en 2D renforce la visualisation et l'analyse des structures cérébrales sous divers angles. De plus, cette approche permet aux radiologues d'évaluer de manière plus précise la taille, la localisation et l'impact des anévrismes sur les structures environnantes du cerveau.

Cependant, il est essentiel de noter que cette méthode n'est pas sans limites. Les images sont riches et détaillées, avec pour objectif de localiser des structures très petites, sans informations sur leur nombre, leur taille ou leur emplacement. Cette complexité peut rendre la tâche ardue pour les radiologues. De plus, l'approche manuelle, bien que riche en informations, exige un investissement significatif en termes de temps et de concentration. Cette exigence peut prolonger la durée nécessaire pour effectuer une évaluation exhaustive des images dans le contexte de la pratique clinique quotidienne des radiologues. En outre, en raison de la complexité de la tâche et du risque d'erreurs humaines, certains anévrismes peuvent échapper à la détection, entraînant des faux négatifs. Ces erreurs peuvent avoir des conséquences sur la santé des patients.

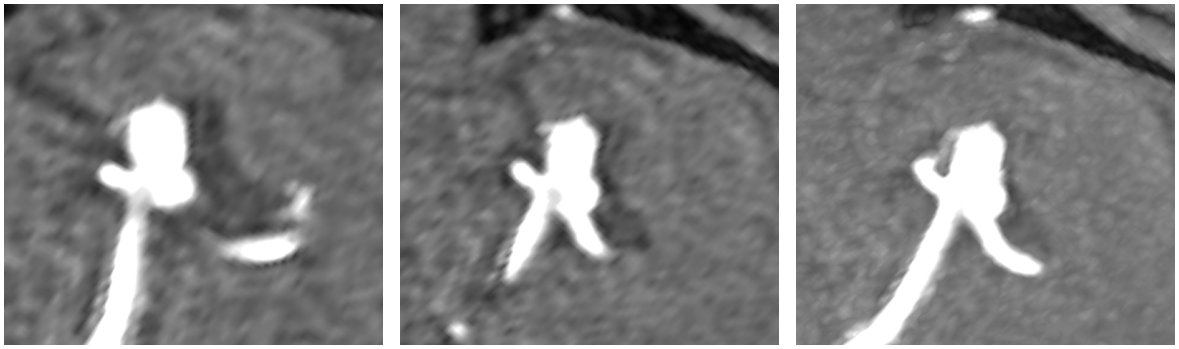
2.3.3 Positionnement du plan de coupe pour une visualisation améliorée

L'analyse des anévrismes intracrâniens englobe l'évaluation de diverses caractéristiques morphologiques telles que le diamètre, la largeur, la taille du collet, la forme, l'orientation, et la localisation par rapport aux structures angio-architecturales environnantes (voir section 2.1.3). L'estimation directe de ces caractéristiques à partir de la vue 3D peut manquer de précision en raison de la variabilité significative

par rapport aux images 2D [9]. Les défis émergent des problèmes de qualité d'image et des variations liées au seuillage, susceptibles d'introduire des erreurs potentielles dans ces mesures qualitatives. La limite de résolution entre les vaisseaux sanguins et les tissus cérébraux pose une contrainte majeure, car, après le seuillage, les tissus cérébraux et les petits vaisseaux sanguins deviennent invisibles dans la vue 3D. Cette limitation restreint la possibilité d'une évaluation précise de la localisation de l'anévrisme par rapport à ces structures environnantes.

Afin d'améliorer la fiabilité des mesures, les radiologues adoptent une approche centrée sur l'analyse en vue 2D. Deux méthodes sont couramment utilisées : les plans de coupes et les images de projection d'intensité maximale (MIP). Dans la première approche, les radiologues naviguent et sélectionnent le plan de coupe le plus approprié en fonction de la localisation et de l'orientation de l'anévrisme (voir Figure 2.10a). À partir de ce plan de coupe, ils calculent les mesures associées à l'anévrisme. Cependant, cette méthode reconnaît les limites inhérentes à la sélection manuelle du plan de coupe optimal. Le choix du plan de coupe canonique (axial, sagittal et coronal) via le défilement ne garantit pas un alignement parfait avec l'orientation de l'anévrisme, ce qui limite la précision du calcul des mesures, notamment de la taille (du collet au dôme). D'où l'importance d'utiliser des plans de coupes reformatés qui s'alignent mieux sur l'orientation de l'anévrisme (voir Figure 2.10b). De plus, la résolution spatiale limitée des images peut influencer la précision de la délimitation des contours des anévrismes, introduisant des incertitudes dans leur caractérisation. Néanmoins, la complexité des formes d'anévrisme peut également compliquer la détermination précise de certaines caractéristiques, comme l'orientation. La prise en compte de ces facteurs lors de l'analyse manuelle des images garantit une évaluation plus précise des anévrismes.

La deuxième approche consiste à générer des images MIP à partir d'un angle de projection offrant la meilleure visualisation de l'anévrisme en fonction de sa localisation et de son orientation, tout en évitant les superpositions indésirables avec d'autres structures vasculaires (voir Figure 2.10c). Cette méthode est souvent appliquée lors des traitements d'anévrismes, en particulier lorsque des scanners à rayons X sont disponibles pour déterminer les angles de projection les plus adaptés.



(a) Plan de coupe canonique (axial). (b) Plan de coupe reformaté. (c) Plan de coupe reformaté avec projection MIP (5 mm d'épaisseur).

FIGURE 2.10 – Analyse en 2D d'un anévrisme de 6,61 mm : mesures du collet, du diamètre, de la largeur, et l'évaluation de la connexion avec le vaisseau parent. Utilisation d'un plan de coupe reformaté avec projection MIP pour une analyse optimale de l'anévrisme et de sa connexion avec le vaisseau parent.

2.3.4 Segmentation des anévrismes pour le suivi du traitement

En complément des caractéristiques morphologiques telles que la taille, la forme et la localisation des anévrismes intracrâniens, le volume émerge comme un paramètre essentiel dans le traitement et le suivi de ces anomalies vasculaires [180, 50]. Il se positionne en tant qu'indicateur pour évaluer l'efficacité des traitements antérieurs, facilitant ainsi la gestion à long terme des patients. La surveillance des variations de volume au fil du temps offre aux cliniciens l'opportunité de déterminer si le traitement a réussi à réduire la poche de l'anévrisme ou à prévenir sa croissance. De plus, la segmentation de la paroi vasculaire joue un rôle clé dans les simulations liées à l'hémodynamique et au risque de rupture, offrant une analyse approfondie des forces et des pressions exercées sur les parois de l'anévrisme [137]. Enfin, la mesure

du volume s'avère être un paramètre précieux pour déterminer la taille des coils nécessaires lors de l'embolisation endovasculaire visant à traiter et isoler un anévrisme. Cependant, au CHRU de Nancy, une bonne approximation de ce volume est généralement réalisée en prenant en compte le diamètre, la largeur et la taille du collet de l'anévrisme.

La segmentation vise à délimiter précisément la zone occupée par l'anévrisme dans les images médicales (voir Figure 2.11). En pratique clinique, cette tâche est souvent effectuée manuellement par les radiologues. Ils identifient et annotent les régions correspondant à l'anévrisme en parcourant les différents plans de coupes 2D contenant l'anévrisme. Cette approche, nécessitant de l'expertise, peut être chronophage, surtout pour les anévrismes de forme complexe, entraînant une variabilité de segmentation, notamment au niveau des contours de l'anévrisme. Les contours obtenus peuvent également être imprécis en raison des limitations de résolution et des artefacts liés à la technique d'imagerie utilisée. Certains centres médicaux proposent des outils de segmentation semi-automatique ou automatique pour faciliter cette tâche laborieuse.

Cependant, l'efficacité de ces outils dépend souvent de la qualité de l'image et de la complexité de la forme de l'anévrisme, pouvant entraîner des erreurs nécessitant une intervention manuelle pour correction et validation des résultats. Une fois l'anévrisme segmenté, les cliniciens mesurent son volume en se basant sur la résolution spatiale de l'image et le nombre de voxels qui le composent. En intégrant les informations de la phase de détection (voir section 2.3.2) et la visualisation dans le plan de coupe optimal (voir section 2.3.3) avec les données quantitatives issues de la segmentation, les cliniciens parviennent à une vision et une analyse plus complètes de l'anévrisme intracrânien. Il est toutefois crucial de noter que l'utilisation de diverses techniques d'imagerie exerce une influence significative sur la fiabilité des mesures morphologiques. Une illustration concrète de cette influence réside dans le fait qu'une incertitude de seulement un demi-voxel (0,2 mm en IRM) lors de la segmentation se traduit par une variation de 24% dans le volume estimé d'un anévrisme de 5 mm de diamètre. En revanche, la 3DRA se distingue par une précision supérieure par rapport aux autres modalités d'imagerie [9].

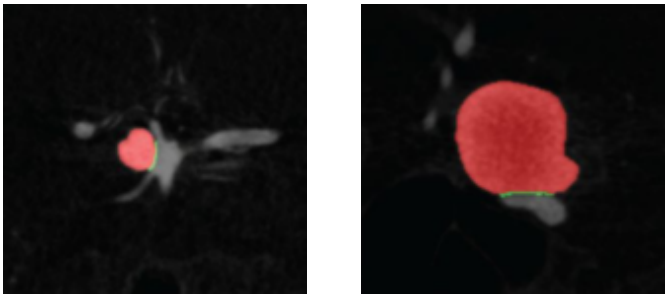


FIGURE 2.11 – Segmentation d'un anévrisme intracrânien de 3,11 mm à partir d'images 3DRA [216] : analyse de l'anévrisme (rouge) et sa connexion avec le vaisseau parent (vert).

Conclusion

Ce chapitre a exploré le contexte clinique des anévrismes intracrâniens, mettant en avant leur physiopathologie caractérisée par des structures de petite taille millimétrique, leur rareté, et une prédominance de forme sacculaire. La prévalence élevée de morbidité et de mortalité associée à la rupture de chaque anévrisme souligne l'impératif d'un diagnostic précoce des anévrismes non rompus. Diverses techniques d'imagerie 3D ont été examinées, mettant en relief l'utilité diagnostique de l'angiographie par résonance magnétique en temps de vol (TOF-MRA) en raison de sa nature non-invasive et de sa disponibilité au sein des établissements de santé. Enfin, les différentes étapes cliniques, telles que la détection, le positionnement optimal des coupes et la segmentation, ont également été exposées. Nous avons montré que ces tâches demandent une très grande expertise et restent malgré tout fastidieuses. L'approche manuelle soulève des défis, étant associée à des performances qui, parfois, pourraient être améliorées (sensibilité, spécificité pour la détection et précision pour la segmentation). Le travail exposé dans la suite de ce manuscrit aborde les problématiques de la détection des anévrismes intracrâniens dans les images TOF-

MRA, ainsi que l'estimation du plan de coupe d'analyse par des approches automatiques. Sans être exclusif, les cibles prioritaires pour ces tâches sont les anévrismes de petite taille (< 5 mm), dont la détection demeure particulièrement difficile en TOF-MRA. Bien que la segmentation ne soit pas l'objet dans ce travail de thèse, nous avons néanmoins exploité des méthodes de segmentation d'anévrismes. Notamment, il apparaît que, pour les petits anévrismes, la 3DRA semble être plus adaptée en termes de précision pour cette tâche. Le prochain chapitre examinera la bibliographie afin de passer en revue les méthodes de la vision par ordinateur proposées pour automatiser et assister les radiologues dans leur pratique clinique du diagnostic des anévrismes intracrâniens.

Chapitre 3

La vision par ordinateur pour le diagnostic des anévrismes intracrâniens

Sommaire

3.1 Vision par ordinateur en imagerie médicale 42
3.1.1 Diversité de tâches d'analyse d'images 42
3.1.2 Évolution des approches d'analyse d'images 44
3.2 Apprentissage profond pour l'analyse d'images médicales 46
3.2.1 Réseaux de neurones convolutionnels et mécanismes d'attention 46
3.2.2 Métriques d'évaluation 51
3.2.3 Méthodologies d'évaluation 55
3.2.4 Défis de l'apprentissage profond pour les anévrismes intracrâniens 57
3.3 Détection et segmentation des anévrismes intracrâniens 59
3.3.1 Approches classiques de traitement d'images 59
3.3.2 Approches par apprentissage automatique 60
3.3.3 Approches par apprentissage profond 62
3.4 Bases de données et annotation des anévrismes intracrâniens 69
3.4.1 Bases de données disponibles 70
3.4.2 Techniques d'annotation 72

L'augmentation continue du volume d'images médicales générées au sein des établissements de santé a engendré une demande croissante pour l'analyse et l'interprétation de ces images, en particulier dans un contexte où les délais sont de plus en plus restreints. En réponse à cette exigence, différentes méthodes d'analyse automatisée ont été mise en place pour soutenir les radiologues dans leur routine clinique. Ce chapitre se concentre sur les méthodes appliquées au diagnostic des anévrismes intracrâniens, en mettant particulièrement l'accent sur celles basées sur les algorithmes d'apprentissage profond. Dans la section 3.1, nous explorons les principes fondamentaux de la vision par ordinateur et leur intégration dans l'analyse des images médicales. Cette section couvre diverses tâches d'analyse d'images et trace l'évolution des techniques utilisées dans ce domaine. La section 3.2 est dédiée aux concepts de base des algorithmes de réseaux de neurones adaptés à l'analyse d'images, y compris les réseaux de neurones convolutionnels et les mécanismes d'attention. Elle présente également les mesures et les méthodologies d'évaluation adaptées à l'imagerie médicale tout en mettant en lumière les défis liés à l'application de ces techniques aux anévrismes intracrâniens. Dans la section 3.3, nous explorons les méthodes de détection et de segmentation des anévrismes intracrâniens, en mettant l'accent sur les approches tridimensionnelles basées sur les réseaux de neurones. Enfin, la section 3.4 met en évidence le rôle des bases de données dans le développement des modèles d'apprentissage. Nous présentons les bases de données disponibles pour les anévrismes intracrâniens, ainsi que les techniques d'annotation associées.

3.1 Vision par ordinateur en imagerie médicale

La vision par ordinateur constitue un sous-domaine de l'intelligence artificielle (IA), situé à l'intersection des mathématiques et de l'informatique. Son objectif principal consiste à permettre aux machines, notamment les ordinateurs, d'analyser et d'interpréter des données visuelles telles que des images et des vidéos, en vue d'extraire des informations pertinentes et de formuler des recommandations [296]. Ce domaine englobe diverses applications, allant de la surveillance vidéo à la robotique, en passant par l'imagerie médicale. L'intégration de la vision par ordinateur en radiologie, par exemple, offre des perspectives prometteuses pour optimiser les procédures, réduisant ainsi la charge de travail des radiologues et améliorant leur efficacité [87]. Cette section présente les tâches fondamentales de la vision par ordinateur appliquées à l'imagerie médicale, ainsi que les méthodes employées pour leur automatisation.

3.1.1 Diversité de tâches d'analyse d'images

L'imagerie médicale joue un rôle central dans le diagnostic, le traitement et le suivi de diverses pathologies affectant la santé humaine. Pour faciliter, voire automatiser, l'analyse et l'interprétation des images médicales, la vision par ordinateur définit une diversité de tâches incluant la classification, la détection, la segmentation et l'estimation de la pose d'objets [87]. Adaptées à des besoins spécifiques, ces tâches permettent d'extraire des informations pertinentes des images, en se concentrant sur des zones d'intérêt telles que les organes ou les anomalies, comme les anévrismes intracrâniens. Les informations ainsi extraites peuvent être de nature qualitative, comme l'identification d'objets spécifiques, ou quantitative, en mesurant des paramètres tels que les dimensions et le volume d'objets.

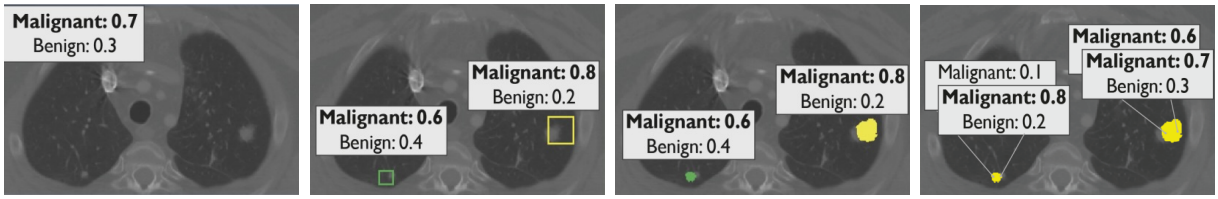
Classification d'images

La classification d'images attribue des catégories à une image en fonction de son contenu, avec des scores de confiance pour chaque catégorie (voir Figure 3.1a). Cette tâche se divise en deux types principaux : multi-classe, où une seule catégorie est attribuée par image, et multi-étiquettes, permettant plusieurs catégories par image. Cette approche est couramment utilisée pour trier des ensembles d'images, facilitant ainsi une analyse ultérieure basée sur des priorités. En imagerie naturelle, la classification est souvent utilisée pour catégoriser des éléments tels que des animaux, des personnes ou des objets du quotidien [163, 78], ainsi que des éléments artificiels [79, 314]. Dans le domaine médical, elle sert dans les programmes de dépistage à répartir les images en catégories telles que "saines" ou "anormales" pour la classification binaire, ou en plusieurs classes comme "normales", "bénignes" ou "malignes". Cette classification contribue à l'optimisation des ressources et à la diminution des erreurs diagnostiques [157]. Par exemple, elle est utilisée pour classifier des radiographies vertébrales [330], des tumeurs cérébrales [77] et des anomalies des artères du polygone de Willis [123].

Dans le diagnostic des anévrismes intracrâniens, la classification d'images a été appliquée dans des travaux ultérieures pour classifier des sous-volumes d'images TOF-MRA selon la présence ou l'absence d'anévrismes [119, 150]. Elle peut être également utilisée pour classifier les images selon des critères plus spécifiques, comme la présence d'anévrismes rompus ou des aspects liés à leur potentiel de danger [208, 215]. Toutefois, la classification d'images peut s'avérer limitée pour des applications nécessitant une localisation plus précise. Dans ces cas, d'autres tâches d'analyse d'images, telles que la localisation, la détection d'objets et la segmentation, sont requises pour acquérir des informations plus fines [60].

Localisation et détection d'objets

La localisation d'objets est une tâche de régression qui vise à déterminer la position des objets d'intérêt dans les images. Cette tâche nécessite la détermination des coordonnées des objets, telles que leurs centres (x, y, et z pour les images 3D), ainsi que l'évaluation des régions qu'ils occupent, en utilisant par exemple des représentations spécifiques comme des boîtes englobantes (avec largeur, hauteur, et profondeur en 3D) ou des cercles/sphères (avec rayon). Néanmoins, avant d'effectuer la localisation, il est essentiel tout d'abord de confirmer la présence des objets, ce qui peut être accompli par la tâche de classification. Lorsque la classification est combinée avec la localisation, on parle alors de détection d'objets (voir Figure 3.1b). En imagerie naturelle, la détection d'objets est appliquée dans divers contextes tels que la surveillance vidéo, le suivi d'objets et la conduite autonome [342]. En imagerie médicale, les systèmes



(a) (b) (c) (d)

FIGURE 3.1 – Comparaison des tâches de vision par ordinateur en imagerie médicale [192] : (a) classification d'images, (b) détection d'objets, (c) segmentation sémantique, et (d) segmentation d'instances.

de détection d'objets ont simplifié le processus de diagnostic de nombreuses pathologies, apportant un soutien précieux aux radiologues [60, 178]. Ces outils peuvent être utilisés en tant que premiers ou seconds lecteurs, offrant ainsi une double vérification des images. Cette automatisation réduit la charge de travail et a le potentiel d'améliorer la précision du diagnostic. Les tâches de détection d'objets ont trouvé des applications variées en radiologie, telles que la détection de nodules pulmonaires [258, 271], de fractures [282], de lésions sclérotiques [183] et d'anévrismes intracrâniens [27, 318]. De plus, la détection d'objets peut être utilisée pour estimer la taille des objets, par exemple en utilisant les dimensions des boîtes englobantes ou les diamètres des cercles/sphères. Toutefois, pour des besoins de caractérisation plus précise, tels que le suivi de la progression tumorale ou l'évaluation de l'efficacité des traitements impliquant le calcul du volume [71], il est alors préférable de recourir à des tâches de segmentation.

Segmentation d'images

La segmentation d'images implique la classification des voxels (ou pixels 2D) ou la délimitation des frontières afin d'identifier avec précision les régions correspondant aux objets d'intérêt dans les images. Cette tâche permet d'obtenir des informations sur les caractéristiques morphologiques des objets, comme leur taille, forme et volume. On distingue principalement deux types de segmentation : la segmentation sémantique et la segmentation d'instances [202]. La segmentation sémantique attribue à chaque voxel une étiquette indiquant la catégorie de l'objet (voir Figure 3.1c). Cependant, cette approche peut se révéler moins efficace lorsque plusieurs objets de la même catégorie sont présents dans une image, comme c'est souvent le cas avec les anévrismes, car elle a tendance à les considérer comme une seule entité. En revanche, la segmentation d'instances, tout en attribuant une étiquette de catégorie, assigne également à chaque voxel un identifiant unique correspondant à l'instance de l'objet dans l'image (voir Figure 3.1d).

Les techniques de segmentation d'images sont fréquemment employées en imagerie médicale pour analyser qualitativement et quantitativement les structures anatomiques, ainsi que pour la localisation et la détection d'anomalies [187]. Cependant, elles présentent des défis liés à la précision des mesures, comme la taille, la forme et le volume des objets, qui dépendent des modalités d'imagerie [32, 9]. De plus, l'utilisation d'images segmentées pour des tâches de localisation et de détection d'objets nécessite des étapes de post-traitement, notamment l'extraction et l'analyse des composantes connectées, le calcul de position et dimensions d'objets, ainsi que la détermination des scores confiance par objet. Cela peut diminuer l'efficacité des méthodes de segmentation par rapport aux méthodes directes de détection d'objets [139].

Estimation de la pose d'objets

L'estimation de la pose d'un objet est une tâche qui consiste à déterminer sa localisation et son orientation par rapport à un système de coordonnées fixe. La localisation se représente généralement par les coordonnées du centre (x, y, z) de l'objet, tandis que l'orientation peut être décrite de diverses manières, telles que les angles d'Euler, les quaternions ou la matrice de rotation [261]. L'importance de l'estimation de la pose se manifeste dans diverses applications nécessitant une compréhension et une interaction avec l'environnement. En imagerie naturelle, elle est couramment utilisée pour estimer la pose de caméras ou d'objets, notamment en robotique [288], réalité augmentée [194], et navigation autonome [311]. Dans le domaine médical, l'estimation de la pose est utilisée, en particulier dans des contextes tels que la chirurgie et les interventions assistées par ordinateur. Cette technique permet de mieux comprendre l'environnement chirurgical en évaluant la position et l'orientation des instruments

chirurgicaux par rapport aux structures anatomiques impliquées [115, 233, 307]. De plus, cette tâche peut être utilisée pour optimiser les processus d'acquisition et de reconstruction des images en coupes transversales [325, 328]. Par exemple, en IRM, la considération de la localisation et de l'orientation des structures anatomiques vise à assurer une représentation cohérente, tout en contribuant à la réduction des artefacts (voir Figure 3.2). En outre, le positionnement du tube à rayons X par rapport à un organe ou une articulation facilite le processus d'acquisition et améliore la qualité des images générées [99, 156].

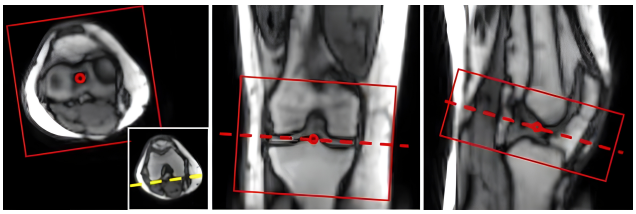


FIGURE 3.2 – Positionnement des coupes transversales d'IRM par rapport à l'articulation du genou [328].

3.1.2 Évolution des approches d'analyse d'images

Les tâches d'analyse et d'interprétation d'images s'adaptent aux particularités de chaque problème [60]. Il est courant de combiner plusieurs de ces tâches au sein d'un même système. Par exemple, un tel système peut débuter par la classification d'images, puis procéder à la localisation ou à la segmentation des objets présents dans les images classifiées [150, 239]. Auparavant, ces tâches s'appuyaient fortement sur l'expertise humaine et utilisaient des techniques classiques de traitement d'images. Cependant, avec l'intégration croissante de l'intelligence artificielle et des algorithmes d'apprentissage automatique, une tendance vers des méthodes plus automatisées a émergé.

Les techniques classiques de traitement d'images reposent sur l'extraction et l'analyse des caractéristiques des images [3]. Elles utilisent des filtres traditionnels tels que les filtres gaussiens pour réduire les irrégularités, les filtres de Sobel et de Prewitt pour détecter les contours, ainsi que les filtres de Laplace pour améliorer le contraste [178]. De plus, l'analyse des histogrammes des images est courante pour comprendre la distribution des intensités d'images, ce qui permet d'ajuster le contraste ou de réaliser une segmentation basée sur les niveaux de gris [171]. Ces techniques sont particulièrement utiles dans des domaines tels que la reconnaissance de formes. Cependant, elles présentent des limitations dans des contextes spécifiques comme l'imagerie médicale, où elles sont moins efficaces pour traiter le bruit, les artefacts et la variabilité due aux équipements d'imagerie et aux variations anatomiques [23]. De plus, l'extraction des caractéristiques des images nécessite une expertise à la fois en traitement d'images et en médecine. Ces techniques classiques sont souvent limitées pour des tâches plus avancées, comme l'évaluation qualitative ou l'interprétation de structures de formes complexes, et nécessitent une intervention manuelle pour ajuster les résultats. Face à ces limitations, des méthodes plus avancées de vision par ordinateur ont été adoptées, qui simulent l'expertise humaine et offrent une plus grande flexibilité pour des analyses de haut niveau.

Vision par ordinateur et apprentissage automatique

La vision par ordinateur, une branche de l'IA, se focalise sur l'analyse et l'interprétation des données visuelles. Dans un premier temps, cette discipline s'est appuyé sur le raisonnement symbolique, notamment par l'utilisation de systèmes experts [220]. Ces systèmes, conçus pour simuler les processus de décision des experts humains, étaient basés sur un ensemble de règles et de bases de connaissances. Ils incluaient divers critères tels que les variations du contraste, les caractéristiques morphologiques des objets et leurs relations spatiales dans l'image. Cependant, la création de ces règles par des experts se révélait extrêmement exigeante en termes de temps et de ressources. De plus, la complexité de ces systèmes augmentait de manière significative avec la difficulté des problèmes traités, limitant leur efficacité et leur applicabilité dans des contextes variés [334]. Face à ces défis, la vision par ordinateur a adopté des approches basées sur l'apprentissage automatique. Ces approches visent à fournir aux machines la capacité d'acquérir des connaissances à partir de données brutes sans programmation explicite, éliminant ainsi la nécessité de définir des règles par des experts [24].

L'apprentissage automatique, qui s'appuie sur des modèles statistiques, se divise principalement en trois catégories d'apprentissage : supervisé, non supervisé, semi-supervisé, et par renforcement [253]. Dans l'apprentissage supervisé, des modèles statistiques sont entraînés avec des données annotées, chaque échantillon étant associé à une étiquette, appelée également "vérité terrain". L'apprentissage non supervisé, en revanche, entraîne les modèles en utilisant des données non annotées pour identifier des motifs ou clusters à partir des caractéristiques intrinsèques des données. L'apprentissage semi-supervisé combine ces deux approches, utilisant à la fois des données annotées et non annotées. Il est pertinent quand l'annotation est difficile, comme en médecine [61]. L'apprentissage par renforcement, différent, repose sur des systèmes récompensés ou pénalisés selon leurs actions, favorisant l'acquisition de stratégies optimales dans un environnement spécifique.

L'apprentissage automatique englobe un large éventail d'algorithmes, tels que les SVM (Support Vector Machines), les k-NN (k-Nearest Neighbors), les K-Means, les arbres de décision et les réseaux neuronaux, chacun d'entre eux étant adapté à des problèmes distincts [253]. Les SVM, un algorithme d'apprentissage supervisé, se distinguent dans les problèmes de régression et de classification, surtout avec des données présentant une séparation non linéaire [256]. Ils fonctionnent en transformant les données dans un espace de dimension plus élevée pour une meilleure représentation, suivie d'une étape d'optimisation qui minimise l'erreur entre les prédictions et la vérité terrain. Cette méthode est particulièrement efficace dans les situations où les marges de séparation entre les classes sont floues ou complexes. Le k-NN est un autre algorithme supervisé, utilisé principalement pour la classification. Il assigne des étiquettes aux nouvelles données en fonction de leur similarité avec les données d'entraînement [70]. L'algorithme identifie les "K" points les plus proches et attribue la classe ou la moyenne des valeurs la plus commune parmi ces voisins, ce qui le rend utile dans des contextes où les relations de proximité sont un indicateur fiable de similarité. Quant à K-Means, c'est un algorithme de clustering non supervisé. Il partitionne les données en "K" groupes distincts, cherchant à minimiser la variance au sein de chaque cluster [190]. Cet algorithme est efficace pour regrouper des données non étiquetées en fonction de leurs caractéristiques intrinsèques, comme dans l'analyse de marché ou la segmentation de la clientèle. Les arbres de décision, utilisés pour la classification, segmentent les données en sous-groupes plus petits en utilisant une série de règles de décision [312]. Cette méthode offre une visualisation claire des chemins de décision, ce qui facilite la compréhension et l'explication des prédictions de l'algorithme.

La variété de ces algorithmes d'apprentissage offre des solutions adaptées pour divers problèmes, de la classification simple à des tâches plus complexes de clustering ou de régression [191]. Toutefois, dans le domaine de l'imagerie, en particulier médicale, l'apprentissage profond est devenu l'approche privilégiée, notamment en raison des améliorations significatives des performances qu'il a apportées [178].

Apprentissage profond et réseaux de neurones

L'apprentissage profond, qui s'appuie sur les réseaux de neurones, s'inspire de la complexité et de l'adaptabilité du cerveau humain. Ces réseaux sont composés de multiples couches de neurones interconnectés, organisés hiérarchiquement pour traiter l'information [106]. Capables de traiter une variété de données, telles que des images en 2D ou 3D, ils transmettent ces informations à travers ces couches pour produire des sorties spécifiques à la tâche en question. Chaque neurone contribue en recevant des signaux d'entrée, calculant une somme pondérée à l'aide de poids pour les connexions, puis appliquant une fonction d'activation non linéaire. Ces fonctions d'activation introduisent une non-linéarité essentielle, permettant au réseau de détecter et d'interpréter des relations et caractéristiques complexes dans les données [226]. Durant l'entraînement, les connexions entre les neurones des couches successives sont ajustées pour optimiser la performance globale du réseau. Une des forces majeures des réseaux de neurones réside dans leur capacité à extraire automatiquement des caractéristiques abstraites à partir des données brutes. Cette capacité contraste fortement avec les méthodes classiques de traitement d'image, où l'extraction des caractéristiques nécessite une expertise pour concevoir des filtres adaptés.

L'entraînement des réseaux de neurones est un processus central dans le développement de modèles statistiques performants. Ce processus consiste à minimiser une fonction de coût qui, dans l'apprentissage supervisé, mesure l'écart entre les prédictions du réseau et la vérité terrain. Pendant l'entraînement, le réseau est soumis à des ensembles de données d'entraînement, divisées en lots, et traités à travers plusieurs itérations, ou époques. Les algorithmes de rétropropagation, tels que la descente de gradient stochastique (SGD) et Adam, ajustent progressivement les poids du réseau pour optimiser la performance [277, 63]. Ces

algorithmes s'appuient sur deux phases essentielles appliquées à chaque lot de données d'entraînement. La première phase, la propagation avant, consiste à introduire les données dans le réseau pour produire des prédictions. La deuxième phase, la propagation arrière, implique le calcul de la fonction de coût et l'ajustement des poids en fonction des gradients calculés, dans le but de réduire cette fonction. Deux problèmes principaux peuvent survenir lors de l'entraînement : la disparition du gradient, où les gradients deviennent progressivement trop faibles pour entraîner une convergence efficace, et l'explosion du gradient, où les gradients deviennent excessivement grands, menaçant la stabilité de l'entraînement. La gestion appropriée du taux d'apprentissage est cruciale pour prévenir ces problèmes [225]. Un taux d'apprentissage élevé peut accélérer la convergence mais risque de dépasser les minima locaux, tandis qu'un taux trop bas peut conduire à une convergence lente ou à un état de stagnation. Enfin, le processus d'entraînement dépend fortement des capacités des unités de traitement graphique (GPU), qui varient en fonction de facteurs tels que le nombre de paramètres du réseau, la dimension des données et la taille des lots d'entraînement. La gestion de la mémoire GPU revêt une importance particulière dans le traitement d'images médicales, caractérisées par leur grande dimensionnalité [266].

Le sur-apprentissage est un autre problème qui peut se manifester durant le processus d'entraînement des réseaux de neurones. Il survient lorsque le modèle s'ajuste de manière excessive aux données d'entraînement, au détriment de sa capacité à généraliser vers de nouvelles données. Ce phénomène résulte souvent d'un manque de diversité dans les données d'entraînement [244, 292]. Pour contrer ce phénomène, diverses techniques sont employées. L'augmentation des données est l'une de ces stratégies. Elle consiste à diversifier l'ensemble d'entraînement en y intégrant des variations des données existantes, telles que des rotations, des changements d'échelle ou des translations [66]. La régularisation est une autre approche efficace, qui vise à simplifier le modèle en imposant des pénalités sur les poids trop importants dans la fonction de coût [206]. En outre, des techniques comme le Dropout et la normalisation par lot sont largement utilisées. Le Dropout réduit le sur-apprentissage en désactivant aléatoirement certains neurones pendant l'entraînement, ce qui empêche le réseau de devenir trop dépendant de caractéristiques spécifiques des données d'entraînement [273]. La normalisation par lot, quant à elle, normalise les entrées de chaque couche, ce qui facilite et accélère l'entraînement tout en réduisant le sur-apprentissage [133]. Enfin, l'arrêt anticipé, qui consiste à arrêter l'entraînement lorsque la performance sur un ensemble de validation ne s'améliore plus, prévient l'ajustement excessif aux données d'entraînement [322].

3.2 Apprentissage profond pour l'analyse d'images médicales

Les réseaux de neurones, adaptés à divers types de données, présentent diverses architectures. Parmi elles, les réseaux de neurones convolutionnels (CNN) sont particulièrement efficaces dans le traitement d'images, grâce à leur capacité à extraire des caractéristiques visuelles complexes [170, 164]. L'intégration de mécanismes d'attention dans les CNNs renforce cette capacité, en focalisant le réseau sur des zones ou caractéristiques spécifiques des images [113, 84]. Cette section aborde les réseaux CNNs et les mécanismes d'attention, ainsi que les métriques et méthodologies d'évaluation adaptées à l'imagerie médicale. Nous nous limitons au cas particulier de l'apprentissage supervisé, qui est notre cadre applicatif.

3.2.1 Réseaux de neurones convolutionnels et mécanismes d'attention

Réseaux de neurones convolutionnels (CNN)

Les réseaux de neurones convolutionnels, s'inspirant du système visuel humain, utilisent des filtres de dimensions réduites pour extraire des caractéristiques visuelles à divers niveaux de complexité au sein de leur architecture. Les couches initiales de ces réseaux se concentrent sur l'identification d'éléments fondamentaux tels que les contours et les textures, similairement aux filtres dans le traitement d'images traditionnel. Aux niveaux plus profonds, les CNNs discernent des structures plus abstraites [292]. Ils offrent trois avantages majeurs [106] : premièrement, le partage de paramètres entre les filtres appliqués à différentes positions de l'image réduit le nombre total de paramètres, améliorant ainsi l'efficacité computationnelle ; deuxièmement, ils possèdent une invariance spatiale, ce qui facilite la reconnaissance d'objets indépendamment de leur emplacement dans l'image ; et troisièmement, ils sont capables de représenter les caractéristiques des images sous la forme de cartes de caractéristiques sémantiques.

La structure des CNNs est constituée de plusieurs éléments distincts, chacun jouant un rôle dans le processus de traitement et d'analyse (voir Figure 3.3) :

- **Couches de convolution :** Ces couches utilisent des filtres entraînables pour détecter automatiquement des structures spécifiques dans les images [110]. Chaque couche de convolution comprend plusieurs filtres de petite taille, telle que 3×3. Par convolution, les poids de chaque filtre sont appliqués uniformément à différentes positions de l'image, permettant ainsi une extraction de caractéristiques invariantes par translation. Les résultats des opérations de convolution sont ensuite agrégés dans des cartes de caractéristiques, représentant l'ensemble des informations extraites.
- **Fonctions d'activation non-linéaire :** Ces fonctions sont utilisées pour capturer et modéliser les dépendances non linéaires entre les données d'entrée. Elles sont généralement appliquées après les couches de convolution pour introduire de la non-linéarité au sein des cartes de caractéristiques. Elles activent sélectivement les neurones, déterminent les activations à transmettre aux couches suivantes, et évitent les problèmes d'annulation et d'explosion des gradients [104, 266]. Parmi les exemples courants de ces fonctions, on trouve ReLU, LeakyReLU, Softmax et sigmoïde [226].
- **Couches d'échantillonnage :** Les couches de sous-échantillonnage (ou pooling), telles que le max-pooling et l'average-pooling, réduisent la taille des cartes tout en préservant les informations essentielles. Elles élargissent également le champ de focalisation tout en réduisant le nombre de paramètres du réseau. En contraste, les couches de sur-échantillonnage utilisent des opérations d'interpolation, comme l'interpolation par le plus proche voisin, pour augmenter la taille des cartes de caractéristiques. Elles compensent la perte d'informations qui peut survenir lors du passage par des couches de sous-échantillonnage tout en améliorant la qualité visuelle du résultat final [292].
- **Couches entièrement connectées (MLP) :** Généralement situées vers la sortie du réseau, ces couches établissent des connexions complètes entre tous les neurones des couches précédentes et elles-mêmes. Leur principale fonction est d'agréger et d'analyser de manière globale les caractéristiques extraites tout au long du réseau neuronal. Cette agrégation permet de prendre des décisions finales ou de générer des prédictions en fonction des caractéristiques extraites de l'image.

Les CNNs se caractérisent par leur variabilité en termes de profondeur, qui correspond au nombre de couches, et de largeur, déterminée par le nombre de filtres ou de neurones dans chaque couche. Ces facteurs définissent la complexité du modèle, mesurée par le nombre de paramètres, et influencent sa capacité à extraire des caractéristiques abstraites. La configuration des couches des CNNs dépend de plusieurs facteurs, tels que la dimension des images, la taille des objets d'intérêt, les ressources de calcul disponibles (GPU), et les exigences spécifiques de la tâche d'analyse [106].

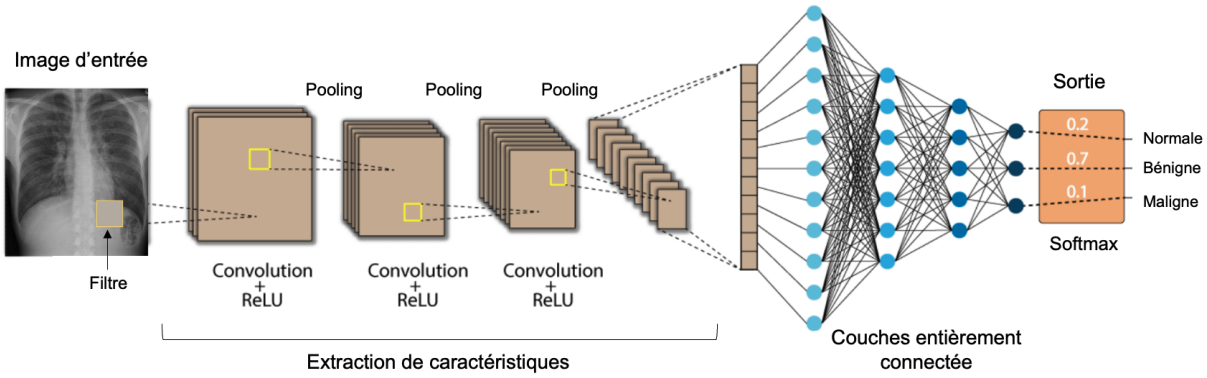


FIGURE 3.3 – Exemple de réseau CNN pour la classification d'images [243] : Couches de convolution avec activation ReLU et pooling pour l'extraction des caractéristiques, suivies de couches MLP et de la fonction Softmax pour générer les scores de classification.

Adaptation des architectures CNNs selon les tâches d'analyse

Les CNNs suivent généralement une structure de base adaptable à la tâche d'analyse spécifique. Un encodeur, constitué de couches de convolution, de fonctions d'activation non linéaires et de sous-

échantillonnage, génère des cartes de caractéristiques. En classification d'images, l'encodeur est suivi par des couches MLP et une fonction d'activation appropriée (sigmoïde pour la classification binaire, Softmax pour la classification multi-classes) pour produire les scores de classification. En détection d'objets, des modules supplémentaires sont ajoutés après l'encodeur, combinant des couches de convolution et/ou MLP avec des fonctions d'activation pour la localisation des objets et l'évaluation de leurs scores de classification. En segmentation d'images, un bloc décodeur similaire à l'encodeur, mais utilisant des couches de sur-échantillonnage, restaure les dimensions des cartes de caractéristiques, permettant ainsi la génération de scores de classification pour chaque voxel de l'image.

Classification d'images L'évolution des CNNs a débuté avec les recherches de Fukushima [97] dans les années 1980, qui ont établi les fondations pour leur développement futur. En 1998, le réseau LeNet [170] a été conçu pour classer des images de petite taille (32×32 pixels) en 10 classes, avec un total de 7 couches, incluant 2 couches de convolution utilisant des filtres de taille 5×5. La compétition ImageNet [78] a joué un rôle clé en offrant une large base de données publique pour la classification d'images de 224×224 pixels en 1000 classes, stimulant ainsi l'avancement des CNNs. L'introduction de réseaux plus profonds a permis d'améliorer considérablement les performances. AlexNet [164], par exemple, a introduit des techniques de régularisation telles que les couches Dropout et la fonction d'activation ReLU, ainsi qu'une implémentation GPU efficace pour les opérations de convolution. Le réseau VGG [269], composé de 16 couches, a ensuite démontré sa capacité à extraire des caractéristiques d'image complexes. Cependant, la profondeur accrue des CNNs a introduit le problème de l'annulation des gradients. ResNet [118] a abordé ce problème en intégrant des connexions résiduelles, facilitant l'entraînement de réseaux encore plus profonds. Inception [278], de son côté, a permis d'extraire des caractéristiques à diverses échelles spatiales en utilisant des couches de convolution avec des filtres de tailles différentes en parallèle. Ces avancées ont non seulement amélioré la classification d'images mais ont également été appliquées à d'autres domaines d'analyse d'images, élargissant ainsi le champ d'application des CNNs.

Détection d'objets Deux approches de réseaux de détection d'objets sont couramment utilisées : les détecteurs en deux étapes et les détecteurs en une seule étape. Les détecteurs en deux étapes suivent un processus de détection en deux phases. La première phase génère des régions d'intérêt (RoI) potentielles, en utilisant des méthodes comme la recherche sélective [291] ou les réseaux de proposition de régions (RPN) [239]. Les RoIs sont ensuite triées et filtrées en fonction de leur score de confiance. Dans la seconde phase, les RoIs sélectionnées sont normalisées et soumises à un second CNN pour la classification des objets et la régression des boîtes englobantes. Des modèles comme R-CNN [102], Fast R-CNN [101] et Faster R-CNN [239] illustrent cette approche. Ces modèles se distinguent par leur précision en matière de localisation d'objets, mais leur complexité computationnelle les rend moins adaptés à la détection en temps réel et avec des capacités de calcul restreintes [342].

En revanche, les détecteurs en une seule étape éliminent la nécessité d'une étape distincte de proposition de régions. Ces détecteurs agissent directement sur l'image complète, la subdivisant en une grille et génèrent des prédictions pour chaque cellule de grille contenant des objets (voir Figure 3.4). Cette méthode se distingue par sa vitesse, produisant des résultats en un seul passage à travers le réseau. L'architecture de ces réseaux se compose de deux composantes essentielles : un encodeur et un bloc de détection. L'encodeur extrait les caractéristiques de l'image, que le bloc de détection utilise pour prédire à la fois les boîtes englobantes des objets et leurs scores, comprenant la confiance de la détection et la catégorie de l'objet. Parmi les détecteurs en une seule étape fréquemment utilisés, on retrouve YOLO [237], SSD [179] et RetinaNet [176]. YOLO se démarque par son équilibre entre rapidité et précision, tandis que SSD repose sur l'extraction de multiples cartes de caractéristiques à différentes échelles pour améliorer la détection d'objets de tailles variées. En parallèle, RetinaNet introduit une fonction de coût Focale mieux adaptée aux problèmes de déséquilibre entre les classes. De plus, RetinaNet tire parti de l'architecture Feature Pyramid Network (FPN) [175], permettant une représentation multi-échelle des images, facilitant ainsi la détection d'objets de tailles variées.

La plupart des systèmes de détection d'objets utilisent des stratégies basées sur des ancres. Ces éléments, agissant comme des points de référence aux dimensions et emplacements variés, sont conçus pour améliorer l'alignement des prédictions avec les objets réels. Ils facilitent l'adaptation du modèle aux diverses tailles et formes des objets [239]. En pratique, une gamme large d'ancres de différentes tailles est employée pour couvrir diverses dimensions et proportions d'objets. Toutefois, le choix des dimensions

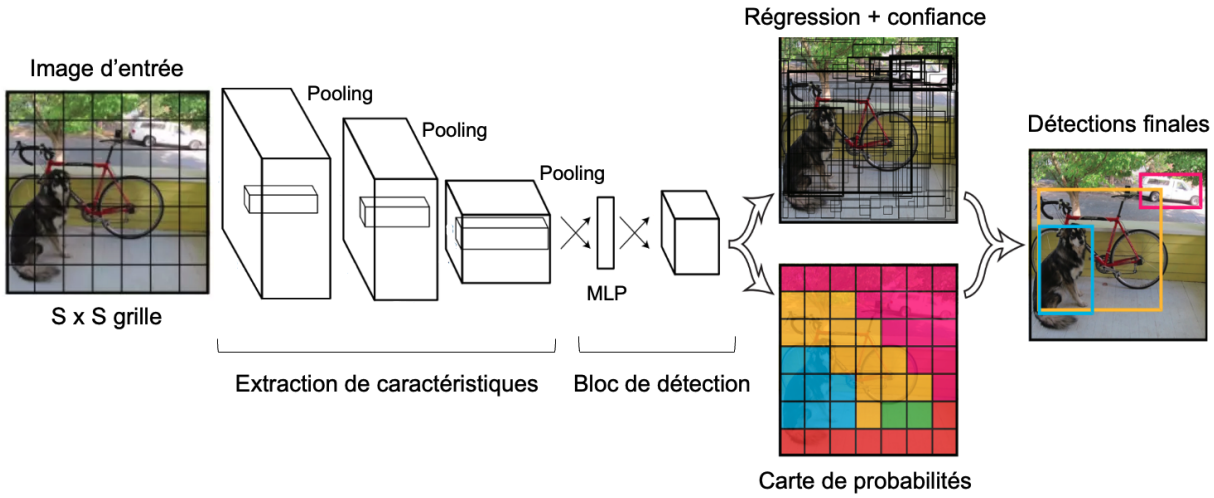


FIGURE 3.4 – Réseaux YOLO [238] pour la détection d'objets en une seule étape : Encodeur suivi d'un bloc de détection pour la localisation et la classification des objets.

des ancres est crucial, un mauvais choix pouvant diminuer la performance de la détection. De plus, la superposition de nombreux objets peut entraîner des ambiguïtés avec les ancres, provoquant des erreurs de classification ou de localisation [342]. Un autre défi est lié à l'augmentation des besoins en mémoire et en calcul due à l'utilisation intensive d'ancres, aggravant le déséquilibre entre les ancres positives et négatives [323]. Pour surmonter ces problèmes, des méthodes de détection sans ancre ont été élaborées, offrant des performances comparables aux approches traditionnelles fondées sur les ancres [329, 271].

Segmentation d'images Les CNNs pour la segmentation d'images ont évolué au fil du temps. Dans un premier temps, des approches telles que le réseau entièrement convolutif (FCN) [184] étaient basées sur des architectures conçues pour la classification d'images. Les couches MLP ont été remplacées par des convolutions de taille 1×1, produisant des prédictions de la même taille que l'image d'entrée. Cependant, cette approche avait une limitation majeure, la perte d'information spatiale lors du sur-échantillonnage, où les cartes de caractéristiques étaient agrandies 32 fois pour atteindre la taille de l'image d'entrée. Pour remédier à cette limitation, les FCNs ont intégré des connexions résiduelles basées sur une opération de sommation, fusionnant les cartes de caractéristiques des couches de convolution avec celles issues du sur-échantillonnage. L'architecture UNet [245], spécifiquement adaptée à l'analyse d'images biomédicales, a amélioré les FCNs en introduisant une phase d'expansion. Elle traite les images en passant par des couches de convolution et de sur-échantillonnage successives, en ajoutant des connexions résiduelles ("skip") basées sur une opération de concaténation entre les phases d'encodage et de décodage à différents niveaux de résolution (voir Figure 3.5). Des développements ultérieurs ont introduit des variantes pour la segmentation 3D, avec le 3D UNet [65] étendant l'UNet classique à trois dimensions et le V-Net [201] améliorant l'intégration du contexte à l'aide de filtres plus grands (5×5) et de connexions résiduelles entre les couches de convolution.

Mécanismes d'attention visuelle

Les CNNs se sont révélés performants dans diverses applications d'analyse d'images. L'adoption de mécanismes d'attention dans ces réseaux leur permet de se focaliser sélectivement sur certaines régions ou attributs spécifiques des images, émulant ainsi le processus d'attention humain dans la reconnaissance visuelle [113, 86]. Ces mécanismes modulent activement le degré d'importance accordé à différentes zones ou caractéristiques pendant l'analyse d'une image, contribuant à l'amélioration des performances des CNNs. On distingue principalement deux types de mécanismes : l'attention par canal et l'attention spatiale.

Attention par canal L'attention par canal utilise des modules de réseaux de neurones pour mettre en évidence les caractéristiques les plus pertinentes au sein des cartes de caractéristiques. Ces modules

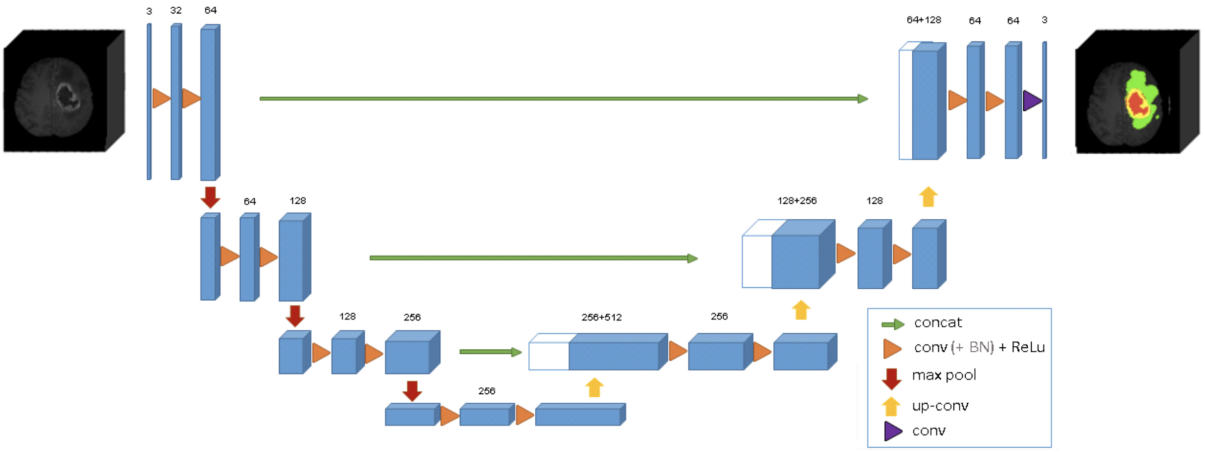


FIGURE 3.5 – Réseau 3D UNet [65] : traitement d'une image 3D en entrée pour générer la segmentation en sortie, avec l'intégration de connexions "skip" par des opérations de concaténation.

incluent des couches de convolution, de pooling, des couches MLP, ainsi que des fonctions d'activation pour générer des pondérations pour chaque canal. Un exemple de cette approche est le réseau SENet [127], qui intègre des opérations de compression et d'excitation pour ajuster l'importance attribuée à chaque carte de caractéristiques. La phase de compression, réalisée par un average pooling, réduit les informations spatiales de chaque carte. Ensuite, la phase d'excitation intervient, où des statistiques globales sont traitées par des couches MLP avec des fonctions d'activation telles que ReLU et sigmoïde, afin de générer des pondérations pour chaque canal. D'un autre côté, le modèle RCAN [331] améliore cette approche en introduisant des connexions résiduelles, intégrées pour optimiser et accélérer le processus d'apprentissage, ce qui permet un ajustement plus précis des poids des caractéristiques. Pour simplifier le mécanisme d'attention tout en maintenant des performances compétitives, le réseau ECA-Nets [301] propose un mécanisme simplifié basé sur une seule couche de convolution, réduisant ainsi la complexité computationnelle tout en préservant un niveau élevé de performance. De plus, les réseaux de segmentation Attention UNet [218] et UNet++ [337] sont spécialement conçus pour filtrer les caractéristiques moins pertinentes au niveau des connexions skip du réseau UNet [245], garantissant que seules les caractéristiques essentielles sont propagées, améliorant ainsi le processus de reconstruction de la segmentation.

Attention spatiale L'attention spatiale se concentre sur l'importance relative des différentes zones au sein des cartes de caractéristiques, permettant ainsi aux réseaux CNNs de diriger leur analyse vers les parties les plus pertinentes de l'image [113]. Plusieurs réseaux CNNs utilisent cette forme d'attention pour améliorer leurs performances. Par exemple, le réseau PSANet [332] exploite l'attention positionnelle autonome (PSA), permettant à chaque élément d'une carte de caractéristiques d'établir des connexions dynamiques avec d'autres éléments en fonction de leur position spatiale relative. Les poids de l'attention sont déterminés en estimant les similarités entre les paires d'éléments à l'aide de couches de convolution. D'autre part, GENet [126] introduit un mécanisme appelé "Gather-Excite", qui se divise en deux étapes : la première, "gather", agrège les cartes de caractéristiques sur une étendue spatiale donnée, tandis que la seconde, "excite", redistribue ces informations agrégées et l'entrée originale pour produire des pondérations par rapport aux différentes régions spatiales. Enfin, les Deformable Convolutional Networks (DCN) [72] intègrent des couches de convolution déformables dans leur structure. Contrairement aux convolutions classiques avec des grilles de filtrage fixes, les couches de convolution déformables permettent aux filtres de s'ajuster dynamiquement aux variations locales de l'image. Cela signifie que la forme et l'orientation des grilles de filtrage peuvent changer en fonction des particularités de chaque région de l'image, créant ainsi des masques d'attention de manière implicite.

Combinaison de l'attention par canal et spatiale L'intégration de l'attention par canal et de l'attention spatiale a été étudiée pour évaluer leur synergie dans l'extraction de caractéristiques sémantiques et spatiales. Ces deux formes d'attention peuvent être intégrées de manière parallèle ou séquentielle.

Par exemple, le module CBAM [310] les combine de manière séquentielle en utilisant une opération de multiplication et est composé de couches de convolution, de pooling, de couches MLP et d'une fonction sigmoïde. Cette approche a montré son efficacité dans la classification d'images et la détection d'objets avec différentes architectures de CNNs. D'un autre côté, le module DAN [96] fusionne les deux types d'attention en additionnant les cartes de caractéristiques générées par l'attention par canal et l'attention spatiale, améliorant ainsi les performances d'un réseau ResNet [118] pour la segmentation d'images. De même, le modèle SCA-CNN [56] intègre l'attention spatiale, via une convolution et une fonction Softmax, et l'attention par canal, via une couche de max-pooling suivie de couches MLP et une fonction Softmax. Cette combinaison permet de créer des cartes de caractéristiques enrichies, , exploitant les avantages de chaque type d'attention pour une analyse plus approfondie des images.

Auto-attention et Transformers L'évolution des mécanismes d'attention en vision par ordinateur a été marquée par l'introduction du mécanisme d'auto-attention, une forme d'attention spatiale [113]. Les réseaux CNNs se fondaient sur des couches de convolution agissant localement sur les images, ce qui présentait des limitations pour capturer les relations à longue distance entre les pixels de l'image. En 2018, le réseau Non-Local [303] a apporté une amélioration significative en intégrant le mécanisme d'auto-attention dans un réseau CNN, permettant ainsi l'encodage des dépendances globales. Ce mécanisme évalue la dépendance et attribue un poids à chaque paire de pixels de l'image, indépendamment de leur distance. Malgré sa complexité computationnelle, l'auto-attention a ouvert la voie à de nouvelles architectures entièrement basées sur ce mécanisme, supprimant même le besoin de couches de convolution dans certaines applications. En 2020, Vision Transformer (ViT) [84] est devenue la première architecture entièrement basée sur les mécanismes d'auto-attention. Pour atténuer la complexité computationnelle liée aux dépendances entre chaque paire de pixels, ViT décompose l'image en patchs (sous-images), évaluant les dépendances à l'aide de blocs d'auto-attention. Cependant, bien que ViT ait montré des performances concurrentielles par rapport aux CNNs, ses exigences en termes de données d'entraînement et de ressources de calcul ont posé plusieurs limitations. Pour surmonter ces limites, des variantes comme le Data-efficient Image Transformer (DeiT) [287] se sont concentrées sur l'apprentissage avec des ensembles de données plus restreints. Le Pyramid Vision Transformer (PVT) [302] utilise une structure pyramidale pour analyser les images à différentes échelles, et le Swin Transformer [182] privilégie une approche hiérarchique, adaptant l'analyse à chaque niveau de résolution pour des cartes de caractéristiques plus fines.

Les architectures Transformers ont principalement été appliquées aux images naturelles en 2D, tandis que leur application en imagerie médicale 3D demeure un défi en raison des exigences en données et puissance de calcul. Face à ces contraintes, des approches hybrides ont été développées, combinant les capacités d'encodage des mécanismes d'auto-attention avec les réseaux CNNs. Par exemple, TransUNet [54] intègre les Transformers en tant qu'encodeur au sein du réseau de segmentation UNet. Medical Transformer (MedT) [293] adopte une approche à deux branches d'attention basées sur les Transformers, l'une locale, opérant sur des patchs d'images, et l'autre globale et peu profonde, opérant sur l'ensemble de l'image 3D. Par ailleurs, SwinUNet [47] combine les architectures Swin Transformer [182] et 3D UNet pour améliorer la construction de cartes de caractéristiques plus pertinentes. En dépit de défis persistants, comme la disponibilité restreinte de données médicales, des approches de transfert de connaissances et de réduction de la complexité des Transformers ont été suggérées pour surmonter ces obstacles [246, 260]. Toutefois, leur adoption à grande échelle en imagerie médicale demeure limitée.

3.2.2 Métriques d'évaluation

L'évolution des architectures neuronales, allant des CNNs aux Transformers, a élargi leurs domaines d'application. Néanmoins, l'évaluation de leurs performances est essentielle, particulièrement dans le secteur médical où les enjeux de santé sont critiques. Cette évaluation implique la comparaison des prédictions avec des données de référence (vérité terrain). En imagerie médicale, elle rencontre des défis spécifiques tels que la limitation des données disponibles, la variabilité entre les images de patients et les équipements d'imagerie, la fiabilité des données de référence, et l'absence de consensus sur les métriques d'évaluation adaptées pour chaque problème [192]. Cette section aborde les métriques d'évaluation appropriées à ces enjeux, traitant tant les problèmes de régression (prédiction de valeurs continues telles que distance, coordonnées, angles) que de classification (prédiction de catégories).

Problèmes de régression

Pour évaluer les modèles de régression, plusieurs mesures peuvent être utilisées. La distance euclidienne, par exemple, quantifie l'écart linéaire entre les coordonnées de points dans un espace multidimensionnel (voir Équation 3.1). Lorsqu'il s'agit d'évaluer l'erreur d'angle entre deux vecteurs, la mesure de similarité cosinus est couramment employée (voir Équation 3.2). Cette dernière analyse l'alignement angulaire entre deux vecteurs, une métrique pertinente dans des applications telles que l'estimation de la pose d'objets [156]. Pour mesurer l'écart entre des échantillons d'observations, des métriques comme l'erreur absolue moyenne (MAE) et l'erreur quadratique moyenne (MSE) sont employées [39]. La MAE mesure la moyenne des erreurs absolues entre prédictions et valeurs réelles (voir Équation 3.3). Elle a tendance à minimiser l'impact des valeurs aberrantes, traitant les grandes erreurs de la même manière que les petites. À l'inverse, la MSE estime la moyenne des carrés des erreurs, ce qui la rend plus adaptée lorsque les erreurs importantes sont considérées comme plus critiques (voir Équation 3.4). En effet, du fait de son caractère quadratique, la MSE accentue les écarts significatifs et se révèle mieux adaptée à l'évaluation de valeurs de petite amplitude [185].

Problèmes de classification

La matrice de confusion est un outil fréquemment utilisé pour évaluer la performance des modèles de classification, qu'ils soient appliqués à des problèmes binaires ou multi-classes. Elle permet de comparer les prédictions à la vérité terrain, tant pour les éléments de la classe ciblée (considérée comme positive) que pour les autres classes (considérées comme négatives). Cette section se concentre sur la matrice de confusion dans le cadre d'un problème binaire. Elle se compose de quatre éléments clés : vrais positifs (TP), faux positifs (FP), vrais négatifs (TN) et faux négatifs (FN), dont les significations diffèrent en fonction de l'analyse réalisée. Dans les tâches de classification ou de segmentation, cet outil sert à mesurer les performances des images ou des voxels classés dans chaque catégorie. En ce qui concerne la détection d'objets, elle facilite l'examen de la localisation et de la classification des prédictions par rapport aux données réelles. Il convient de mentionner que dans le cas de la détection d'objets, la catégorie des TN reste indéfinie à cause du nombre illimité de zones sans objets dans les images.

Dans le contexte des algorithmes d'apprentissage profond, l'établissement de la matrice de confusion nécessite l'application d'un seuil de classification pour convertir les prédictions continues en étiquettes de classe. Par exemple, un score de classification supérieur à 50% pourrait être attribué à la classe "anévrisme", tandis qu'un score inférieur ou égal à 50% serait classé comme "non anévrisme". Ce seuil a un impact direct sur les composantes de la matrice de confusion et, par conséquent, sur les métriques d'évaluation associées. Ces métriques se divisent en deux catégories principales : les métriques fondamentales, calculées avec un seuil de classification fixe, et les métriques basées sur des représentations graphiques, qui examinent les variations de performance à différents seuils de classification [122, 192].

Métriques d'évaluation fondamentales

Exactitude L'exactitude mesure le pourcentage de prédictions correctes parmi l'ensemble des prédictions (voir Équation 3.5). Cependant, dans des cas de déséquilibre entre les classes, cette métrique peut être trompeuse. Elle risque de privilégier la classe majoritaire au détriment de la classe minoritaire. Par exemple, dans une situation où la classe positive représente seulement 10% du total, un modèle prédisant constamment la classe négative atteindrait un taux d'exactitude de 90%, malgré son inefficacité à identifier correctement la classe positive. Pour une évaluation plus représentative, d'autres métriques sont souvent privilégiées, telles que l'exactitude équilibrée (voir Équation 3.6), la précision et la sensibilité.

Spécificité La spécificité mesure la capacité d'un modèle à classifier correctement les éléments de la classe négative, qui ne présentent pas la condition ciblée (voir Équation 3.7). Une spécificité élevée est importante dans les contextes où les erreurs de faux positifs ont des conséquences graves ou coûteuses. Par exemple, en dépistage médical, une spécificité élevée aide à exclure les cas non pathologiques, évitant ainsi des traitements superflus, réduisant les risques de complications et améliorant l'utilisation des ressources médicales. Cependant, cette mesure se limite aux cas négatifs, ne couvrant pas l'identification des cas positifs, pour laquelle d'autres métriques comme la sensibilité ou la précision sont privilégiées.

Sensibilité Contrairement à la spécificité, la sensibilité évalue le pourcentage des éléments positifs correctement identifiés parmi tous les éléments positifs (voir Équation 3.8). Dans des domaines tels que le diagnostic médical, où la détection précise de conditions rares ou graves est impérative, la sensibilité joue un rôle important. Toutefois, la sensibilité ne prend pas en compte les faux positifs, dont un nombre élevé peut entraîner des coûts et une gestion inefficace des ressources pour traiter les cas mal classifiés [122]. Par conséquent, d'autres métriques complémentaires telles que la précision sont souvent utilisées pour évaluer la précision des prédictions de la classe positive.

Précision La précision est le pourcentage de prédictions positives qui sont correctement classifiées parmi toutes les prédictions positives (voir Équation 3.9). Cette métrique est utile pour mesurer la fiabilité des prédictions positives, notamment dans les systèmes de détection d'objets. Toutefois, pour une évaluation plus globale de la performance, il est recommandé d'utiliser des métriques qui prennent en compte à la fois la fiabilité des prédictions positives (précision) et la capacité à identifier correctement tous les cas positifs (sensibilité). Le score F et l'intersection sur l'union (IoU) sont des exemples de telles métriques.

Score F Le score F, ou F_β, est une métrique qui intègre la précision et la sensibilité (voir Équation 3.10). Le coefficient β ajuste le poids relatif de la sensibilité par rapport à la précision, permettant ainsi de personnaliser l'évaluation selon les exigences spécifiques de chaque application. Un cas particulier est le score F1 (où $\beta = 1$), connu également sous le nom de score Dice. Le score F1 est fréquemment utilisé pour mesurer l'équilibre entre la précision et la sensibilité, et est largement appliqué, notamment dans la segmentation d'images, surtout dans des situations où il y a un déséquilibre de classes [275].

Intersection sur union (IoU) L'IoU, également appelé l'indice Jaccard, est couramment utilisé pour évaluer les performances des modèles de détection d'objets [192, 280]. Il quantifie le degré de superposition entre les objets détectés par le modèle et ceux présents dans la vérité terrain (voir Équation 3.11). Cette métrique offre une évaluation précise de la capacité du modèle à effectuer une classification précise des objets (déterminer la classe d'objet détectée) et à les localiser avec précision dans l'image [221].

Kappa de Cohen (κ) Le coefficient de Kappa de Cohen est une mesure qui évalue l'accord entre deux ensembles, en prenant en considération la possibilité d'un accord aléatoire [199]. Cette caractéristique

Régression	**Distance euclidienne**	$\sqrt{(x_2-x_1)^2+(y_2-y_1)^2+(z_2-z_1)^2}$	(3.1)
	Similarité cosinus	$\cos^{-1}\left(\frac{\vec{v_1}\cdot\vec{v_2}}{\|\vec{v_1}\|\cdot\|\vec{v_2}\|}\right)$	(3.2)
	MAE	$\frac{1}{n}\sum_{i=1}^{n}\|x_i-\hat{x}_i\|$	(3.3)
	MSE	$\frac{1}{n}\sum_{i=1}^{n}(x_i-\hat{x}_i)^2$	(3.4)
Classification	**Exactitude**	$(TP+TN)/(TP+TN+FP+FN)$	(3.5)
	Exactitude équilibrée	$\frac{1}{2}\cdot[TP/(TP+FN)+TN/(TN+FP)]$	(3.6)
	Spécificité	$TN/(TN+FP)$	(3.7)
	Sensibilité	$TP/(TP+FN)$	(3.8)
	Précision	$TP/(TP+FP)$	(3.9)
	Score F_β	$\left[(1+\beta^2)\cdot TP\right]/\left[(1+\beta^2)\cdot TP+\beta^2\cdot FN+FP\right]$	(3.10)
	IoU	$TP/(TP+FP+FN)=\mathrm{F1}/(2-\mathrm{F1})$	(3.11)
	Kappa de Cohen	$\frac{P_o-P_e}{1-P_e}=2\cdot\frac{TP\cdot TN-FN\cdot FP}{(TP+FP)\cdot(FP+FN)+(TP+FN)\cdot(FN+TN)}$	(3.12)
	Précision moyenne	$\sum_K(\text{Sensibilité}_{K+1}-\text{Sensibilité}_K)\cdot\text{Précision}_{K+1}$	(3.13)

TABLE 3.1 – Exemples de métriques d'évaluation pour des problèmes de régression (valeurs prédites x, réelles $\hat{x}$, échantillons n) et de classification (composantes de matrice de confusion, détections K).

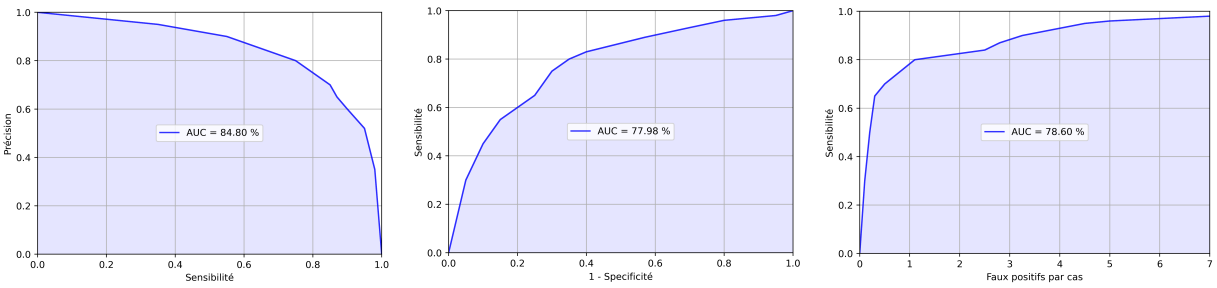
le distingue des autres métriques en offrant une évaluation plus sophistiquée que la simple mesure du pourcentage d'accord. Le coefficient κ compare l'accord observé (P_o) à l'accord qui serait attendu par hasard (P_e), permettant ainsi de corriger pour les accords aléatoires (voir Équation 3.12). Les valeurs de κ varient de -1 (absence totale d'accord) à 1 (accord parfait), avec 0 indiquant un accord au hasard. Ce coefficient se révèle particulièrement utile dans des contextes caractérisés par un déséquilibre des classes [67].

Métriques basées sur des représentations graphiques

Le calcul des métriques précédemment évoquées exigent l'application d'un seuil de classification aux prédictions. Souvent, le choix de ce seuil est effectué de manière arbitraire et peut varier entre différentes méthodes, ce qui rend la comparaison entre ces méthodes complexe [192]. Afin de garantir une évaluation plus objective, il est courant de recourir à des métriques basées sur des représentations graphiques telles que la courbe Précision-Sensibilité, la précision moyenne (AP), ainsi que les courbes ROC (*Receiver Operating Characteristic*) et FROC (*Free-Response* ROC). Ces métriques permettent d'évaluer les performances à différents seuils de classification, facilitant ainsi l'analyse des méthodes dans leurs configurations optimales. De plus, elles simplifient le choix du seuil qui présente les meilleurs performances.

Courbe Précision-Sensibilité L'analyse de la variation de la précision en fonction de la sensibilité est essentielle dans divers contextes, comme le diagnostic des pathologies, où l'objectif est d'atteindre une sensibilité élevée tout en minimisant les faux positifs [137]. La courbe Précision-Sensibilité est couramment employée pour cette analyse, mettant en évidence les fluctuations de précision et de sensibilité à différents seuils de classification (voir Figure 3.6a). Le seuil optimal, visant à maximiser simultanément précision et sensibilité, peut être déterminé de deux manières : en calculant le score F1, qui exprime l'équilibre entre précision et sensibilité à divers seuils, et en sélectionnant le seuil produisant le score maximal ; ou en identifiant directement le "point de fonctionnement optimal" en haut à droite de la courbe. De plus, l'aire sous la courbe (AUC) fournit une mesure globale des performances, simplifiant ainsi la comparaison entre différentes méthodes, une AUC élevée indiquant une capacité à obtenir une grande précision tout en classifiant correctement les éléments positifs.

Précision moyenne (AP) La précision moyenne propose une alternative pour quantifier l'aire sous la courbe Précision-Sensibilité (AUC), adaptée à l'évaluation des méthodes de détection d'objets [221]. Son calcul repose sur une somme pondérée des précisions obtenues à différents seuils de classification. Tout d'abord, les détections sont classées en fonction de leurs scores de confiance, plaçant ainsi les détections les plus fiables en premier. Chaque détection est ensuite évaluée en comparant la prédiction avec la vérité terrain, en utilisant un seuil IoU prédéfini. Les TP et les FP cumulés sont ensuite calculés par rapport à l'ensemble des détections au-dessus du seuil de confiance correspondant. Les scores de précision et de sensibilité sont calculés pour chaque détection en fonction des valeurs cumulées de TP et FP. L'AP est définie comme la somme pondérée des précisions en fonction des valeurs de sensibilité calculées pour l'ensemble des détections K (voir Équation 3.13). Une valeur élevée de l'AP indique que le modèle parvient à maintenir une grande précision tout en identifiant correctement les objets réels.



(a) Courbe Précision-Sensibilité (b) Courbe ROC (c) Courbe FROC

FIGURE 3.6 – Analyse des performances à travers des métriques basées sur des représentations graphiques.

Courbes ROC et FROC La courbe ROC analyse le compromis entre la sensibilité et la spécificité pour évaluer la capacité discriminatoire d'un modèle entre les classes positives et négatives [40]. Elle représente graphiquement la sensibilité en fonction du taux de faux positifs (1 - spécificité) à différents seuils de classification (voir Figure 3.6b). Cependant, dans le contexte des tâches de détection d'objets, les vrais négatifs et par conséquent, la spécificité, ne sont pas clairement définis. Pour de telles situations, la courbe FROC est plus appropriée, évaluant la sensibilité en fonction du nombre de faux positifs par cas (voir Figure 3.6c). Néanmoins, l'utilisation de l'aire sous la courbe FROC ne garantit pas toujours des comparaisons robustes entre différentes méthodes. Cette limitation découle de la variation potentielle du nombre maximal de faux positifs par cas pour chaque méthode, ce qui réduit la fiabilité des comparaisons entre divers systèmes de classification [22].

3.2.3 Méthodologies d'évaluation

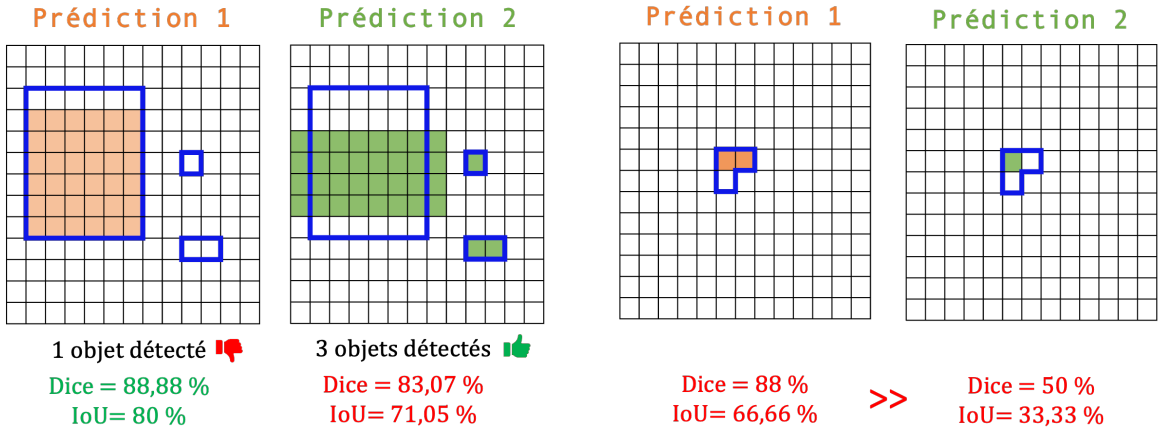
L'évaluation des performances en imagerie médicale présente des défis uniques par rapport à d'autres domaines [192]. Les ensembles de données sont souvent limités et l'annotation des données de référence se révèle complexe et sujette à une variabilité significative. De plus, le déséquilibre entre les classes peut introduire un biais en faveur de la classe majoritaire [45]. Ces caractéristiques particulières engendrent des défis spécifiques pour une évaluation fiable, ce qui requiert des méthodologies adaptées.

Qualité et fiabilité des annotations

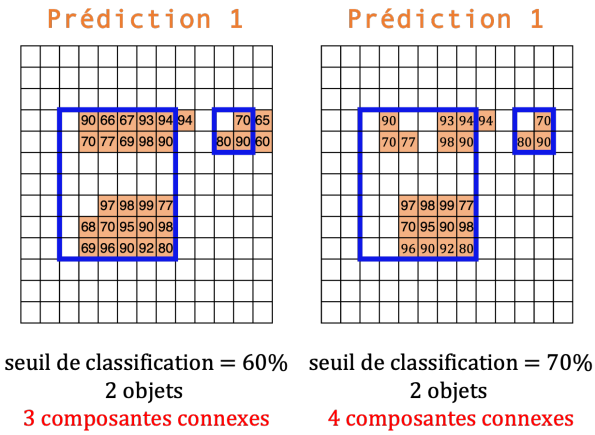
La disponibilité limitée d'images médicales constitue un défi majeur pour une évaluation significative, souvent restreinte à quelques dizaines à centaines d'images. Cette contrainte découle principalement de la nécessité de préserver la confidentialité des patients et de la complexité intrinsèque de l'annotation des images. L'annotation est une tâche exigeante et chronophage qui exige une grande précision, concentration et expertise médicale [305]. Pour accroître la taille des ensembles de données et surmonter ces obstacles, il est impératif de renforcer la collaboration institutionnelle et d'investir dans la collecte, le partage et l'annotation des données. En outre, les annotations peuvent présenter des variations, à la fois entre différents annotateurs (inter-observateurs) et chez un même annotateur (intra-observateurs) [155]. Cette variabilité est influencée par l'expertise des annotateurs, la complexité des images médicales, leur richesse en détails, leur qualité variable, ainsi que les caractéristiques des éléments à annoter (nombre, taille, localisation, etc.). Elle peut se traduire par diverses erreurs d'annotation, telles que des fausses annotations, des cas non annotés et des annotations imprécises. Pour réduire cette variabilité et garantir des données de référence robustes, l'adoption de procédures d'annotation impliquant plusieurs experts est recommandée [155]. Diverses méthodes peuvent être envisagées pour déterminer la vérité finale, comme le calcul de la moyenne des annotations, le recours au vote majoritaire avec consensus, ou d'autres approches [254].

Choix de métriques

Le choix des métriques pour évaluer les performances des modèles d'apprentissage dépend de plusieurs facteurs, tels que la nature de la tâche, les caractéristiques des objets, et les ensembles de données [192, 280]. En segmentation d'images, l'objectif est de classifier les voxels dans leurs catégories respectives. Cependant, dans la détection d'objets, l'évaluation se concentre à la fois sur la classification des objets et leur localisation dans l'image. La localisation est souvent évaluée en utilisant la métrique d'intersection IoU. La présence d'objets de petite taille peut introduire un biais en faveur des objets de grande taille, nécessitant une évaluation de chaque instance d'objet (voir Figure 3.7a). En outre, de petites variations liées à des facteurs tels que la variabilité entre différents observateurs ou la nature non déterministe des modèles d'apprentissage [276] peuvent significativement affecter les mesures d'évaluation, justifiant le choix de métriques adaptées à la problématique (voir Figure 3.7b). L'utilisation de méthodes de segmentation pour des problèmes de détection d'objets peut être moins appropriée. Par exemple, une augmentation du seuil de classification au niveau des voxels peut entraîner une augmentation du nombre de composantes connexes, ce qui réduit la précision de l'évaluation au niveau d'entités d'objets (voir Figure 3.7c). Par ailleurs, pour une évaluation complète des modèles, des métriques basées sur des représentations graphiques sont utiles pour évaluer les performances à différents seuils de classification.



(a) Confusion entre la détection d'objets et la segmentation. (b) Défis de l'évaluation des objets de petite taille.



(c) Augmentation du seuil de classification entraîne plus de composantes connexes.

FIGURE 3.7 – Importance de choisir des métriques appropriées à la problématique (vérité terrain en bleu).

Validation croisée

La validation croisée est une méthode d'évaluation fortement recommandée lorsque les données sont limitées [12]. Elle implique la division des données en sous-ensembles, appelés "plis", et l'exécution itérative d'une procédure combinant l'entraînement et l'évaluation du modèle. À chaque itération, l'un des plis est réservé exclusivement à l'évaluation, tandis que les autres sont utilisés pour l'entraînement. Cette approche itérative permet d'évaluer la performance du modèle sur différentes partitions de données, assurant ainsi une évaluation plus robuste avec des indicateurs de variabilité de performance. Plusieurs variantes de la validation croisée sont disponibles, dont la méthode "leave-one-out" se démarque en utilisant chaque donnée individuelle comme ensemble de test tour à tour, avec les autres données pour l'entraînement [235]. Toutefois, cette variante nécessite d'importantes ressources de calcul pour l'entraînement des modèles, ce qui limite son applicabilité à grande échelle.

Évaluation clinique

L'évaluation sur un ensemble de test permet de sélectionner des méthodes potentiellement applicables dans un contexte clinique. Les données d'entraînement et d'évaluation, souvent limitées dans leur capacité à refléter la diversité des cas réels [111], requièrent une validation clinique approfondie par des professionnels de santé pour évaluer correctement le potentiel de généralisation de ces méthodes. Les systèmes de diagnostic assisté par ordinateur (CAD), en particulier, sont fréquemment employés en tant que premiers ou seconds relecteurs, offrant ainsi une vérification supplémentaire des résultats diagnostiques [49]. Cette

double vérification vise à améliorer l'efficacité des professionnels de santé tout en assurant la qualité des décisions cliniques. L'intégration réussie de ces systèmes dans la routine clinique nécessite une collaboration étroite entre les experts médicaux et les développeurs. La mise en place de protocoles de validation rigoureux, comprenant des évaluations continues et des ajustements basés sur les retours d'expérience clinique, est indispensable pour optimiser la fiabilité de ces outils automatisés en milieu clinique.

3.2.4 Défis de l'apprentissage profond pour les anévrismes intracrâniens

Ensembles de données limités

L'entraînement et l'évaluation des modèles d'apprentissage en imagerie médicale sont entravés par la disponibilité limitée de bases de données annotées, particulièrement pour les données 3D. Contrairement à d'autres anomalies, telles que les tumeurs cérébrales, qui bénéficient de bases de données publiques plus étendues [1, 272], les bases de données pour les anévrismes intracrâniens sont plus restreintes et privées, comprenant quelques dizaines à centaines d'images [81]. Cette rareté est principalement due à la complexité et au coût élevé associés à l'annotation des données [146]. Plusieurs approches peuvent être envisagées pour surmonter ce défi. Une consiste à simplifier le processus d'annotation, par exemple, en utilisant des annotations approximatives [155]. Une autre stratégie repose sur l'apprentissage semi-supervisé, qui exploite à la fois les données annotées et non annotées pour entraîner des modèles plus robustes malgré la rareté des données annotées [61].

L'apprentissage contrastif, axé sur la détection des similitudes entre les échantillons, ainsi que l'apprentissage à partir de peu d'échantillons, sont également des méthodes potentielles pour travailler avec des quantités limitées de données [304, 162, 339]. En outre, l'apprentissage fédéré, permettant de former des modèles sur des données réparties sans nécessiter un transfert direct, offre une autre solution pour utiliser des données de sources diverses tout en préservant la confidentialité et la sécurité des données [262]. Bien que ces approches aient été appliquées dans divers domaines, leur utilisation spécifique dans le diagnostic des anévrismes demeure un champ en développement. Enfin, les techniques d'augmentation de données, comme la rotation, l'effet de miroir et la synthèse d'images, peuvent pallier le manque de données et améliorer la généralisation des modèles [62].

Spécificités des anévrismes intracrâniens

L'analyse d'images médicales se distingue de l'imagerie naturelle par des défis spécifiques, découlant de la complexité des modalités d'imagerie, des structures anatomiques détaillées, des contours d'objets incertains et complexes, ainsi que de la grande diversité des échelles des objets [129]. Cette complexité est encore accentuée dans le cas des anévrismes. En effet, ces anomalies vasculaires se distinguent par leur petite taille variable, de moins d'un millimètre à plus de 30 mm, leur nombre restreint et inconnu a priori chez les individus (généralement de 0 à 2), leur ressemblance avec les segments courbés des vaisseaux sains, ainsi que leur localisation variable au niveau de l'arbre vasculaire.

Le diagnostic des anévrismes par les approches d'apprentissage est donc plus complexe que l'analyse de pathologies plus grandes, en raison de leur petite taille et du manque d'informations visuelles détaillées. Les opérations de pooling dans les CNNs peuvent réduire la précision nécessaire pour identifier ces petites anomalies [181]. De plus, une annotation extrêmement précise est essentielle pour ces anévrismes, car des erreurs minimes peuvent introduire une grande variabilité dans les données d'entraînement. Des efforts pour améliorer la qualité d'image, standardiser les annotations et encourager la collaboration entre experts aident à augmenter la fiabilité des données [236, 134, 214]. En parallèle, les méthodes ensemblistes, bien qu'exigeantes en données et ressources de calcul, présentent une alternative pour pallier les erreurs éventuelles d'annotation [249]. D'autres pathologies de petite taille, telles que les polypes et les nodules pulmonaires, ont souscrit des recherches en détection automatique. Bien que ces recherches puissent être source d'inspiration, elles ne répondent pas spécifiquement au cas des anévrismes. Les polypes, similaires en morphologie aux anévrismes, sont principalement analysés dans des images naturelles (2D) et sont donc compatibles avec les algorithmes standards d'analyse d'images [29, 144]. En revanche, l'analyse des anévrismes nécessite des images médicales en 3D. Leur taille est similaire à celle des nodules pulmonaires, qui se distinguent par des contrastes plus marqués dans les images CT et disposent également d'importantes bases de données publiques [258, 185]. Cependant, la complexité des anévrismes réside dans leur

similarité avec les courbures des vaisseaux sanguins, exigeant des algorithmes capables de se focaliser sur leur signal rare au sein d'images riches en détails.

Problème de déséquilibre de classes

Le déséquilibre des classes est un problème courant en apprentissage automatique, notamment lorsqu'il s'agit d'objets rares ou de petite taille, comme les anévrismes. Cette situation se caractérise par un nombre limité de voxels positifs (classe minoritaire, quelques dizaines de voxels) par rapport aux autres tissus cérébraux (classe majoritaire, des millions de voxels). L'entraînement de réseaux CNNs sur des données présentant un tel déséquilibre a tendance à privilégier les classes majoritaires, ce qui a un impact négatif sur les performances du modèle [45]. Pour atténuer ce problème, on peut utiliser des techniques d'échantillonnage de données et des fonctions de perte adaptées [148].

L'échantillonnage de données est une approche visant à résoudre le déséquilibre des classes en ajustant le nombre d'échantillons dans les classes minoritaires et majoritaires de l'ensemble d'entraînement jusqu'à atteindre l'équilibre souhaité. Deux approches principales sont couramment utilisées à cette fin : le sous-échantillonnage et le sur-échantillonnage. Le sous-échantillonnage réduit le nombre d'échantillons des classes majoritaires en utilisant des méthodes telles que le sous-échantillonnage aléatoire ou NearMiss [193], qui sélectionne les échantillons de la classe majoritaire ayant des caractéristiques similaires à ceux de la classe minoritaire. Cependant, le sous-échantillonnage est plus approprié lorsque l'ensemble d'entraînement est suffisamment vaste, ce qui n'est pas toujours le cas en imagerie médicale. En revanche, le sur-échantillonnage vise à augmenter le nombre d'échantillons de la classe minoritaire, ce qui peut être réalisé par des méthodes telles que la duplication d'images ou des techniques d'augmentation et de synthèse d'images comme SMOTE [51] et ADASYN [116].

L'utilisation de fonctions de coût adaptées au déséquilibre de classes permet de réévaluer l'importance relative accordée aux différentes classes lors de l'entraînement. L'entropie croisée binaire (BCE) est fréquemment employée dans les problèmes de classification binaire. Cependant, en cas de déséquilibre de classes, elle traite chaque échantillon de manière égale, favorisant ainsi les échantillons des classes majoritaires. Pour relever ce défi, plusieurs fonctions de coût ont été proposées dans la littérature [188]. La BCE pondérée intègre des poids spécifiques pour chaque classe. Ces poids peuvent être définis manuellement ou appris pendant l'entraînement. Les classes minoritaires reçoivent des poids plus élevés, incitant ainsi le modèle à réduire sa préférence pour les classes majoritaires. De manière similaire, la fonction Focal [176], une variante de la BCE, introduit un coefficient de modulation pour réguler l'atténuation de coût pour les échantillons majoritaires. Elle est fréquemment utilisée dans des scénarios de grand déséquilibre, tels que dans la détection d'objets. Les fonctions de coût Dice et IoU se concentrent principalement sur la réduction des faux positifs, ce qui est particulièrement pertinent dans les ensembles de données déséquilibrés [31]. Enfin, le coût de Tversky [251] ajuste le compromis entre la précision et la sensibilité en intégrant deux coefficients dans le Dice pour contrôler l'importance des faux négatifs et des faux positifs.

Contraintes de calculs pour le traitement des données 3D

Les images cérébrales, souvent de grande dimension à $512 \times 512 \times 256$ voxels pour les images TOF-MRA, capturent la complexité anatomique du cerveau avec un haut niveau de détail. Cependant, l'entraînement direct des CNNs à partir de ces images 3D présente des défis, à la fois en termes de puissance de calcul et de quantité de données requises pour garantir la généralisation des modèles. Pour surmonter ces obstacles, la littérature explore deux approches principales : l'approche 2D et l'approche 3D.

L'approche 2D consiste à extraire plusieurs images 2D à partir de chaque image 3D en utilisant deux méthodes d'extraction possibles. La première méthode extrait des coupes 2D le long des plans anatomiques (axial, sagittal et coronal), mais elle suppose souvent une indépendance entre ces images 2D, ce qui est généralement incorrect et entraîne une perte d'informations dans la troisième dimension des images [65, 201]. De plus, cette approche complique la caractérisation des structures vasculaires et la distinction entre les vaisseaux sains et ceux présentant des anévrismes. La deuxième méthode utilise des images par projection d'intensité maximale (MIP) pour générer des images 2D mettant en évidence les régions les plus lumineuses des structures anatomiques selon un angle de projection. Cependant, l'approche MIP présente des limites, notamment une capacité réduite à différencier les structures superposées et des difficultés à évaluer précisément la tortuosité des vaisseaux [125].

En revanche, l'approche alternative consiste à extraire plusieurs sous-volumes 3D, appelés "patchs", à partir de chaque image initiale. Cette méthode capture les caractéristiques spatiales et contextuelles locales des structures anatomiques tout en atténuant la perte d'information tridimensionnelle associée aux images 2D [4]. Cette approche est cruciale en imagerie cérébrale, où la localisation et l'évaluation des structures anatomiques nécessitent une considération de l'information 3D [270, 65]. Cependant, elle entraîne des besoins computationnels plus élevés par rapport aux images 2D [25, 131]. De plus, le choix de la taille des patchs doit dépendre des pathologies étudiées. Ainsi, les anévrismes, en raison de leur petite taille pouvant être représentés au sein de patchs locaux. De plus, la configuration de divers paramètres, tels que la stratégie de sélection des patchs, le taux d'intersection, les dimensions, la résolution spatiale et le nombre de patchs extraits, est nécessaire pour mettre en place cette approche et doit être ajustée en fonction des caractéristiques des pathologies étudiées.

3.3 Détection et segmentation des anévrismes intracrâniens

Afin d'assister les radiologues dans le diagnostic des anévrismes intracrâniens, plusieurs systèmes CAD ont été présentés dans la littérature. Ces systèmes utilisent une variété de techniques, allant des méthodes classiques de traitement d'images aux algorithmes d'apprentissage automatique, notamment les réseaux de neurones. Parmi ces systèmes, deux approches principales se distinguent : la détection d'objets, utilisée par une minorité de travaux [27, 318], vise à localiser directement les anévrismes en employant des représentations spécifiques comme les boîtes englobantes. En revanche, la segmentation, qui est plus largement adoptée, permet d'extraire les localisations des anévrismes en se basant sur des images segmentées. Cette section présente une revue de ces différents systèmes, en mettant en évidence les techniques utilisées pour relever les défis associés à la détection des anévrismes intracrâniens.

3.3.1 Approches classiques de traitement d'images

Les premiers systèmes CAD dédiés à la détection des anévrismes se sont appuyés sur des méthodes conventionnelles de traitement d'image. Ces méthodes comprenaient des étapes de filtrage, de seuillage, et des algorithmes basés sur le raisonnement par cas, dans le but d'identifier des régions d'intérêt (RoIs) présentant des caractéristiques similaires à celles des anévrismes. Dans ce domaine de recherche, deux approches principales ont été employées pour l'identification de ces RoIs : la première se basait sur la forme sphérique ou ellipsoïdale des anévrismes anévrismes [10, 119, 11, 120, 161], tandis que la seconde considérait les anévrismes comme des anomalies au niveau de la surface des vaisseaux sanguins [320, 167].

Localisation des anévrismes par leur forme sphérique ou ellipsoïdale

En 2004, l'une des premières méthodes de détection des anévrismes a adopté un algorithme basé sur le raisonnement par cas [10]. Cette méthode a débuté en appliquant un filtre de forme sphérique pour identifier les structures dans les images présentant une similitude avec la forme des anévrismes. Ensuite, un processus de seuillage basé sur la taille de ces structures a été utilisé pour éliminer les régions dont la taille n'est pas similaire à celle des anévrismes, suivi d'un algorithme de croissance des régions pour accentuer les régions candidates. Afin de réduire les faux positifs (FP), un algorithme de raisonnement par cas, basé sur des règles prédéfinies liées aux caractéristiques des anévrismes telles que leurs valeurs d'intensité et leur morphologie, a été appliqué. L'efficacité de cette méthode a été testée sur un ensemble de 115 images MRA, comprenant 61 anévrismes, révélant une sensibilité de détection de 97% avec un taux de 3,8 FP/cas. D'autres études ont également utilisé des approches similaires en utilisant des filtres de rehaussement de forme sphérique [119, 11, 120].

Pour améliorer la détection des anévrismes, dont les tailles varient de moins de 1 mm à plus de 30 mm, l'étude [120] a employé un filtre de rehaussement sphérique doté de tailles variables. Dans le but de minimiser les faux positifs, une fonction discriminante linéaire a été conçue pour classer les régions d'intérêt identifiées, intégrant des paramètres tels que la forme, la taille et la position des anévrismes. Cette évaluation a été menée sur 151 images issues de diverses techniques d'imagerie. Les résultats ont indiqué une sensibilité de 95% avec 11,3 FP/cas pour les images TOF-MRA. Néanmoins, une variabilité dans les performances a été observée selon les modalités d'imagerie (CE-MRA : sensibilité de 95% et 8,2 FP/cas ;

CTA : sensibilité de 95% et 22,8 FP/cas). La majorité de ces faux positifs se situaient dans les zones courbées des vaisseaux, présentant des structures similaires aux anévrismes. Pour affiner la distinction entre les vaisseaux sains et ceux avec anévrismes, l'étude [147] a utilisé un filtre de rehaussement de forme ellipsoïdale, permettant de réduire les faux positifs de 51,4% par rapport au filtre sphérique. Toutefois, le taux de faux positifs demeurait élevé, s'établissant à 22,8 FP/cas, avec une sensibilité de 87%, sur la base de l'analyse de 30 images MRA comprenant au total 31 anévrismes.

La plupart de ces systèmes se sont principalement concentrés sur l'identification de régions d'intérêt correspondant aux anévrismes, sans établir de hiérarchie ou d'ordre de pertinence. Cependant, cette approche limite la capacité à classer et à évaluer ces détections en termes de fiabilité, un aspect potentiellement avantageux pour l'analyse des radiologues [151, 60]. Pour surmonter cette limitation, des stratégies de détection combinant à la fois la localisation et la classification ont été utilisées dans [161]. La méthode proposée attribue des scores de confiance, exprimés en pourcentage, à chaque région d'intérêt identifiée. Ces scores de confiance sont calculés en utilisant un algorithme basé sur le raisonnement par analogie, en se référant à un modèle d'artères saines. L'évaluation de cette approche a été réalisée sur un échantillon restreint de 16 images MRA, contenant un total de 19 anévrismes. Les résultats ont démontré une sensibilité de 100% avec 6,43 FP/cas.

Localisation des anévrismes par analyse de la surface des vaisseaux

Les approches fondées sur l'analyse de la surface des vaisseaux se basent sur la caractérisation des anévrismes comme renflements ou dilatations locales au sein de l'arbre vasculaire. Par exemple, dans [320], les anévrismes sont identifiés en tant que branches courtes au sein des vaisseaux sanguins. Le processus débute par la segmentation des vaisseaux afin de localiser ces branches d'intérêt. Ensuite, une phase de post-traitement, qui utilise un algorithme de raisonnement par cas, est appliquée pour réduire les faux positifs. L'évaluation de cette méthode a été réalisée sur un ensemble de 287 images MRA, comprenant un total de 147 anévrismes. Les résultats indiquent une sensibilité de 95% pour les grands anévrismes (> 5 mm), avec 9 FP/cas. Toutefois, pour les petits anévrismes (< 5 mm), les performances sont moins satisfaisantes, avec une sensibilité de 80% et 3 FP/cas. Une autre approche de localisation des anévrismes repose sur l'évaluation de la symétrie locale de l'arbre vasculaire, comme décrit dans [167]. Ce processus commence par la segmentation des vaisseaux et l'extraction de leur axe médian. Ensuite, les régions vasculaires sont analysées localement : les vaisseaux sains se caractérisent par des trajectoires rectilignes, cylindriques ou paraboliques, tandis que les anévrismes sont identifiés lorsque des régions présentent une asymétrie. Cette asymétrie est quantifiée par la tortuosité (nombre de Writhe [98]) de l'arbre vasculaire. L'évaluation de cette méthode a été effectuée sur un ensemble limité de 10 images 3DRA et CTA, montrant une sensibilité de 100%, avec un taux variable de faux positifs par modalité (0,66 FP/cas pour la 3DRA et 5,36 FP/cas pour la CTA).

Les premiers systèmes de détection d'anévrismes ont recouru à des techniques de traitement d'images traditionnelles, exploitant la forme sphérique ou ellipsoïdale des anévrismes ou les identifiant comme des anomalies locales au sein des vaisseaux sanguins. Bien que ces systèmes aient démontré une sensibilité de détection relativement élevée, leur applicabilité dans des contextes cliniques réels n'a pas été suffisamment explorée. Le processus manuel d'extraction de caractéristiques et d'établissement de règles a posé des défis en termes d'expertise et de généralisation. La présence de bruit dans les images médicales a entraîné une augmentation des faux positifs, notamment en raison de la similitude entre les anévrismes et les zones saines, en particulier dans les segments courbés [320]. Malgré l'utilisation de techniques de post-traitement, le taux de faux positifs est resté élevé. De plus, ces systèmes ont montré des performances moindres dans la détection des petits anévrismes (moins de 5 mm), avec des variations selon les modalités d'imagerie. Ces limitations mettent en évidence la complexité du problème de détection des anévrismes et soulignent la nécessité de mettre en place des méthodes plus adaptées.

3.3.2 Approches par apprentissage automatique

La détection des anévrismes par des méthodes classiques de traitement d'images repose sur des critères préétablis liés à leur forme, leur taille et leur emplacement. Cependant, la création et la formalisation de ces critères nécessitent de l'expertise, et leur applicabilité est souvent limitée en raison de la variabilité

des images résultant des différentes modalités et protocoles d'acquisition, ainsi que des variations intrinsèques aux caractéristiques morphologiques des anévrismes. Pour améliorer la capacité de discrimination de ces méthodes classiques, des tentatives ont été faites pour intégrer des algorithmes d'apprentissage automatique dans le processus de détection. Leur objectif est d'apprendre automatiquement les critères de classification en utilisant toujours des caractéristiques préalablement définies.

Filtres de rehaussement pour la proposition de régions d'intérêt

En 2012, l'une des premières approches d'apprentissage automatique pour la détection des anévrismes a choisi d'utiliser l'algorithme de clustering K-means en combinaison avec une méthode basée sur le filtrage [119]. Cette méthode débute par la localisation des régions d'intérêt à l'aide d'un filtre de rehaussement de forme sphérique. Ensuite, une variante de l'algorithme K-means est implémentée pour analyser et regrouper ces régions d'intérêt en se basant sur plusieurs critères extraites, tels que leurs valeurs d'intensité, leurs dimensions, leur forme, leur volume, et leur distance les séparant. L'évaluation de cette approche a été réalisée sur un ensemble de 93 images contenant un total de 80 anévrismes. Les performances varient selon la modalité d'imagerie, avec une sensibilité de 95% et 2,6 FP/cas pour la 3DRA, une sensibilité de 89% et 6,6 FP/cas pour la MRA, et une sensibilité de 95% et 37,6 FP/cas pour la CTA. Cette variabilité, en particulier en ce qui concerne les faux positifs, illustre l'impact des caractéristiques distinctes de ces modalités sur la capacité de généralisation de ces algorithmes.

De manière similaire, des classificateurs basés sur les forêts d'arbres décisionnels ont été associés à un filtre de rehaussement [142]. Dans cette méthode, des filtres de rehaussement de forme sphérique ont d'abord été utilisés pour localiser les régions d'intérêt susceptibles de contenir des anévrismes. Cependant, en raison de la similitude morphologique entre les irrégularités vasculaires et les petits anévrismes, un algorithme basé sur les forêts d'arbres décisionnels a été intégré pour réduire le nombre de faux positifs. Cet algorithme a été entraîné sur des caractéristiques extraites des sorties du filtre de rehaussement, en utilisant un ensemble de 15 images 3D-DSA pour l'entraînement. L'évaluation de cette approche a été menée sur 10 images, comprenant un total de 15 anévrismes. Les résultats ont montré une sensibilité de 100%, accompagnée de 0,4 FP/cas.

Approche par fenêtres glissantes pour la proposition de régions d'intérêt

Plutôt que de recourir à des filtres de rehaussement pour générer des régions d'intérêt candidates, une approche alternative basée sur des fenêtres glissantes a été utilisée dans [232]. Cette méthode commence par extraire des patchs (sous-images) de taille fixe avec un chevauchement à partir des images DSA. Les caractéristiques de ces patchs sont ensuite encodées sous forme de vecteurs en utilisant un descripteur basé sur les motifs binaires locaux (LBP). Par la suite, l'algorithme d'apprentissage automatique k-NN est entrainé en utilisant un ensemble annoté de ces vecteurs LBP, ce qui lui permet d'apprendre à les classifier en présence ou en absence d'anévrismes. Pour réduire le nombre de faux positifs, des opérations de post-traitement basées sur la morphologie des anévrismes ont été intégrées, conjointement avec le descripteur de Fourier générique (GFD) basé sur les contours. Cette approche a été évaluée dans la classification des patchs, montrant une sensibilité de 91%, une spécificité de 99%, et une précision de 63,21% sur un ensemble de 10 images DSA. Malgré l'utilisation d'opérations de post-traitement, cette méthode demeure sujette à un nombre élevé de faux positifs, mettant en évidence la complexité inhérente à la distinction entre les régions anévrismales et les vaisseaux sains dans les images DSA. Cela soulève également des interrogations quant à l'efficacité des motifs LBP pour l'extraction et la représentation des caractéristiques des images.

Malgré leur sensibilité élevée démontrée, ces méthodes d'apprentissage automatique présentent plusieurs limitations à prendre en considération. Les évaluations ont souvent été effectuées sur des ensembles de données restreints, dépourvus de métriques clairement définies ou inadaptées à la détection d'objets, entravant ainsi une comparaison complète des approches. De plus, ces algorithmes n'ont pas été intégrés à la phase de détection des anévrismes, mais utilisés uniquement en post-traitement pour réduire les faux positifs. Cette limitation découle principalement de la complexité computationnelle inhérente à ces algorithmes lorsqu'ils sont appliqués au traitement de données de haute dimension [142]. La plupart de ces travaux sont concentrés sur la localisation des anévrismes, sans fournir de scores de confiance, ce qui peut constituer un inconvénient dans les applications cliniques où l'évaluation de la fiabilité des détections peut être précieuse pour les radiologues [151, 60]. Face à ces défis et compte tenu de l'expansion de la

disponibilité des données d'imagerie, la transition vers des approches d'apprentissage profond a marqué un tournant significatif dans la détection des anévrismes. L'automatisation s'étend alors également à l'apprentissage des caractéristiques discriminantes liées aux anévrismes.

3.3.3 Approches par apprentissage profond

L'avènement de l'apprentissage profond a révolutionné le traitement d'images en exploitant les capacités des CNNs. Leur flexibilité dans le traitement de données de dimensions diverses, qu'il s'agisse d'images 2D ou 3D, représente une avancée majeure par rapport aux algorithmes classiques d'apprentissage automatique. De plus, ces réseaux éliminent le besoin d'intervention d'experts pour extraire et caractériser les anévrismes dans les images. Initialement axées sur les images 2D, les premières approches basées sur l'apprentissage profond ont progressivement évolué vers des méthodes 3D, exploitant ainsi les informations dans les trois dimensions des images.

Approches bidimensionnelles (2D)

Les premiers CNNs utilisés pour la détection des anévrismes intracrâniens ont favorisé l'utilisation d'images 2D. Ce choix découle de plusieurs considérations. La prépondérance de la modalité 2D DSA dans le diagnostic des anévrismes persistait avant que les modalités basées sur des coupes transversales, telles que la MRA et la CTA, ne gagnent progressivement en importance. De plus, les CNNs ont été initialement conçus pour traiter des images naturelles (2D), ce qui a ralenti leur adaptation au contexte des images 3D. L'entraînement des CNNs avec des images 3D nécessite des ressources de calcul et de vastes ensembles de données pour un apprentissage rapide et efficace, une accessibilité qui n'est devenue largement possible que ces dernières années [106]. Ces facteurs ont conduit à privilégier dans un premier temps les approches 2D ou 2.5D, notamment lors de l'utilisation d'images par projection d'intensité maximale (MIP).

Images DSA Les images de DSA, qui représentent les vaisseaux cérébraux sous forme d'images 2D en projection, ont constitué le pivot de premières approches exploitant des CNNs. En 2020, un CNN a été utilisé pour segmenter les anévrismes intracrâniens à partir de ces images [230]. Cette méthode a combiné l'encodeur VGG-16 [269] avec un décodeur composé de couches de convolution transposées pour segmenter les anévrismes sur des images DSA de 1024×1024 pixels. L'objectif principal était d'améliorer la distinction entre les vaisseaux sains et ceux présentant des anévrismes. Par conséquent, le réseau a été entraîné pour classer les pixels en trois catégories : anévrismes, vaisseaux sanguins et autres tissus cérébraux. L'entraînement a utilisé la somme de Dice et BCE comme fonction de coût, plus adaptée au déséquilibre des classes [188]. Pour éviter le sur-apprentissage avec un ensemble d'entraînement limité à 200 images, des techniques d'augmentation de données et la régularisation Dropout ont été appliquées. La méthode a obtenu les performances suivantes sur un ensemble de 100 images DSA : un score Dice de 90,3%, un score IoU de 82,3% et un score AUC-ROC de 79,1% pour la segmentation des anévrismes. Cependant, la segmentation des petits anévrismes (< 3 mm) et le nombre relativement élevé de faux positifs demeurent des limitations. De plus, cette étude s'est concentrée sur la tâche de segmentation, laissant ainsi inexplorée sa performance pour la tâche de détection.

Images de coupes canoniques La détection et la segmentation des anévrismes à partir d'images en coupes transversales (3D) présentent des défis significatifs en imagerie médicale. Contrairement aux images DSA, qui permettent une visualisation complète de l'arbre vasculaire, l'utilisation de coupes canoniques 2D dans les images en coupes transversales ne permet pas une distinction précise entre les anévrismes et les vaisseaux sains. Dans la pratique, les radiologues examinent plusieurs images de coupes adjacentes pour améliorer leur compréhension des éléments présents dans l'image. Une approche préliminaire [290] a exploré la détection directe des anévrismes à partir de plans de coupes 2D en utilisant un CNN basé sur ResNet-18 [118]. Cette méthode repose sur la sélection guidée de patchs extraits de ces plans, privilégiant les images offrant une visualisation optimale des anévrismes tout en incluant d'autres anomalies vasculaires telles que les sténoses et les bifurcations vasculaires. L'objectif était d'améliorer la capacité discriminative du CNN pour distinguer les anévrismes des irrégularités vasculaires. Pour chaque patch extrait, le réseau génère un score de classification indiquant la présence ou l'absence d'un anévrisme, sans spécifier sa localisation précise. L'évaluation de cette méthode sur 67 images TOF-MRA,

contenant au total 80 anévrismes, a montré une sensibilité de détection de 93%, avec 5 FP/cas. Cependant, la transition de l'approche de la classification d'images vers la détection des anévrismes, ainsi que la spécification des métriques d'évaluation, demeurent non précisées. De plus, le nombre élevé de faux positifs souligne l'importance de l'information contextuelle pour une meilleure distinction des anévrismes. Dans ce contexte, des méthodes 2.5D ont émergé en utilisant les images MIP.

Images MIP De manière similaire aux images DSA, les images MIP, qui offrent une représentation contextuelle des vaisseaux sanguins, ont été utilisées dans des approches analogues. Une première méthode de ce type a été introduite en 2018 [211]. Elle a employé un CNN peu profond composé de deux couches de convolution pour la détection des anévrismes. Cette méthode synthétisait et représentait un patch 3D de dimensions $16 \times 16 \times 16$ voxels ($7,5^3$ mm^3) sous forme de plusieurs images MIP à partir de 9 plans de projection. Les images MIP étaient extraites et concaténées à partir de ces plans. En plus des techniques standard d'augmentation de données telles que les rotations et les effets miroir, une stratégie de sous-échantillonnage était appliquée aux images négatives pour compenser la rareté des images positives contenant des anévrismes. L'évaluation réalisée sur un ensemble de 100 images, comprenant 104 anévrismes, a révélé une sensibilité de 94,2%, avec 2,9 FP/cas. Dans un contexte similaire, la méthode [274] a choisi d'utiliser un réseau de segmentation UNet [245] pour segmenter les anévrismes à partir d'images MIP de dimensions 256×256 pixels. Pendant la phase d'entraînement, une sélection stratégique des images d'entraînement a été effectuée en tenant compte des angles de projection autour des axes x et z, afin d'optimiser la visualisation des anévrismes par rapport à l'arbre vasculaire. L'ensemble de données d'entraînement, qui comprenait 500 images MIP, a bénéficié d'opérations standard d'augmentation de données. Le coefficient Dice a été utilisé comme fonction de coût lors de l'entraînement. L'évaluation de cette méthode sur un ensemble de 86 images TOF-MRA a révélé une sensibilité de classification d'images atteignant 98,8%, avec une AUC de 87% pour la courbe Précision-Sensibilité. Toutefois, l'efficacité de cette méthode dans le contexte de la tâche de détection d'objets reste inexplorée.

Les méthodes en 2D et 2.5D pour détecter les anévrismes rencontrent des défis en pratique, notamment à cause des taux élevés de faux positifs. Bien que les approches 2.5D intègrent des informations contextuelles, elles sont limitées par la superposition des structures vasculaires, augmentant ainsi les risques de faux positifs et de faux négatifs [211]. Cette problématique met en lumière la complexité de la structure vasculaire et l'importance des informations anatomiques en 3D pour distinguer efficacement les anévrismes. De plus, l'absence de précisions dans les méthodologies d'extraction de patchs avant la projection MIP et dans l'agrégation de prédictions pour construire une prédiction au niveau de l'image globale pose des défis. De plus, il y a un manque de détails dans la description des métriques d'évaluation, des architectures CNNs et des tailles de données utilisées pour l'entraînement, ce qui complique la comparaison objective des différentes approches, d'autant plus que certaines se sont centrées sur la tâche de classification d'images, d'autres sur la segmentation, et d'autres sur la détection.

Approches tridimensionnelles (3D)

Les approches utilisant les CNNs en 3D tirent parti des informations contextuelles présentes dans les différents plans anatomiques, à savoir axial, coronal et sagittal. Cette méthode favorise une meilleure compréhension des structures anatomiques et de leurs relations spatiales. Toutefois, ces approches sont limitées par des contraintes en termes de ressources de calcul et par la disponibilité limitée des données d'entraînement. Une solution couramment adoptée pour pallier ces limitations est l'utilisation de techniques basées sur les patchs. Cette stratégie implique l'extraction de multiples sous-volumes (patchs) de chaque image, augmentant de ce fait le nombre d'échantillons pour l'entraînement. Le choix des patchs, notamment leur taille, leur quantité, leur résolution spatiale, est un facteur pour optimiser la performance de ces approches. Lors de l'évaluation finale, les prédictions réalisées à l'échelle des patchs sont agrégées pour obtenir un résultat global sur toute l'image. Les sections suivantes examinent les méthodes de détection des anévrismes utilisant des approches 3D, y compris celles qui ont participé à des compétitions telles que ADAM [1] [285] et CADA [2] [137] en 2020. À noter que la plupart de ces méthodes ont été publiées entre 2019 et 2022, marquant ainsi les avancées récentes dans ce domaine. Nous nous rappelons que nos travaux de thèse ont débuté en novembre 2020.

1. Aneurysm Detection And segMentation (ADAM)
2. Cerebral Aneurysm Detection and Analysis (CADA)

Usage de réseaux existants et évaluation de l'apport clinique

En 2019, le réseau DeepMedic [153] a été appliqué à la segmentation des anévrismes à partir d'images TOF-MRA [268]. Ce réseau se distingue par l'intégration de deux voies de convolution parallèles, permettant l'extraction de caractéristiques à différentes résolutions de l'image d'entrée : $25\times25\times25$ voxels pour la résolution normale et $19 \times 19 \times 19$ voxels pour la résolution réduite. Face au problème du déséquilibre des classes, une stratégie d'échantillonnage attribue 50% des patchs d'entraînement à des échantillons contenant un anévrisme (positifs). L'entraînement s'est déroulé sur un ensemble de 58 images TOF-MRA, avec une optimisation basée sur la fonction de coût Dice. Cette approche a présenté une sensibilité de 90%, accompagnée de 6,1 FP/cas sur un ensemble de 17 images. Pour évaluer l'impact de cette méthode sur la performance des experts, une analyse clinique a été menée par deux experts agissant en tant que premiers relecteurs [91]. Évalué sur un ensemble de 85 images TOF-MRA contenant un total de 115 anévrismes, DeepMedic a amélioré leur sensibilité de détection. Le premier expert, avec 2 ans d'expérience, a augmenté sa sensibilité de 95 à 98%. De même, le deuxième expert, avec 12 ans d'expérience, a augmenté sa sensibilité de 94 à 97%. Cependant, malgré ces améliorations, un défi persiste, illustré par un taux élevé de faux positifs (6,1 FP/cas) et une sensibilité réduite à 38% pour les petits anévrismes (< 3 mm).

Également en 2019, le réseau de segmentation HeadXNet [222] illustre une approche similaire dans l'évaluation de l'impact des outils d'automatisation sur la pratique clinique, en se concentrant sur les images CTA. Employé par 8 experts, ce réseau a été spécifiquement entraîné pour segmenter les anévrismes sur des patchs de $16 \times 192 \times 192$ voxels, bénéficiant d'un pré-entraînement de son encodeur sur des images naturelles. Les paramètres du réseau HeadXNet ont été optimisés en utilisant une combinaison de fonctions de coût, incluant BCE, Dice et la perte Focal [176]. L'évaluation de ce modèle a révélé une amélioration de la sensibilité des experts, passant de 83,1 à 89%, sur un ensemble de 115 images CTA contenant 56 anévrismes. Cependant, il convient de noter que son évaluation se limite à des images comprenant exclusivement des anévrismes de taille supérieure à 3 mm, limitant ainsi son efficacité pour la détection des petits anévrismes, qui sont plus difficiles à repérer par les experts. Ces différentes approches soulignent la diversité des CNNs mis en œuvre pour automatiser et améliorer la détection et la segmentation des anévrismes. Malgré leur potentiel, ces outils doivent encore surmonter des défis importants, notamment la prévalence des faux positifs, qui peuvent prolonger le temps d'analyse des radiologues [81].

Stratégies de réduction des faux positifs

La quête de stratégies visant à réduire les taux élevés de faux positifs a conduit à l'exploration d'approches en deux étapes. En 2020, une méthode combinant un réseau de classification de patchs avec un réseau de segmentation a été proposée dans [150]. La première étape repose sur un réseau de classification de type ResNet [118], chargé d'identifier les RoIs. Cette étape est suivie de l'application d'un réseau 3D UNet [65] pour une segmentation d'anévrismes dans les RoIs positives. Après la segmentation, des boîtes englobantes de taille fixe à 10 mm^3 sont positionnées au centre de chaque composante connexe dans les images segmentées. L'ensemble d'entraînement englobe 468 images TOF-MRA, totalisant 551 anévrismes. Les patchs d'entraînement, de dimensions $16 \times 32 \times 32$ voxels, sont sélectionnés de manière aléatoire à partir du volume global, sous la contrainte que 66% contiennent des anévrismes. L'évaluation de cette approche démontre une sensibilité de 85,7% et une précision de 91,5% sur un ensemble de test composé de 56 images contenant un total de 63 anévrismes.

Une autre approche en deux étapes, centrée sur la segmentation des vaisseaux sanguins, a été exposée dans une étude distincte [52]. Dans un premier temps, la segmentation des vaisseaux sanguins est réalisée au moyen de techniques classiques de traitement d'image, comprenant le seuillage, l'histogramme et les algorithmes de croissance de région. Utilisant ensuite des patchs de taille $128 \times 128 \times 128$ voxels, la segmentation des anévrismes est effectuée à l'aide d'une version modifiée du réseau 3D UNet [65], intégrant une supervision profonde pour améliorer la reconstruction de l'image à sa résolution initiale. L'entraînement s'effectue sur un ensemble de 76 images TOF-MRA comprenant un total de 80 anévrismes. Après l'extraction des composantes connexes des images segmentées, les anévrismes sont localisés et marqués par des sphères de taille fixe. Cette approche atteint une sensibilité de 82,9%, avec un taux de 0,86 FP/cas, sur un ensemble de 35 images contenant un total de 35 anévrismes. Cependant, l'efficacité de cette méthode dépend essentiellement de la qualité de la segmentation préalable des vaisseaux sanguins, et l'utilisation de patchs de grande taille nécessite des ressources de calcul importantes. De plus, l'adoption

d'une taille fixe pour les sphères ou les boîtes englobantes ne tient pas compte de la variabilité des tailles couramment observées pour les anévrismes, allant de moins de 1 à plus de 30 mm.

Réseaux CNNs plus sophistiqués

Les avancées dans les méthodes de détection des anévrismes ont conduit à l'exploration d'architectures de réseaux plus avancées, mettant principalement l'accent sur des mécanismes d'attention. En 2020, le réseau DAResUNet [263] a été employé pour segmenter les anévrismes à partir de patchs de $80 \times 80 \times 80$ voxels. Cette architecture intègre des blocs résiduels et des blocs d'attention double [96] au niveau des connexions "skip" d'un réseau 3D UNet [65]. L'ajout de blocs résiduels vise à améliorer la stabilité de l'apprentissage des réseaux profonds, tandis que le bloc d'attention permet d'extraire des informations contextuelles à longue portée, générant des cartes de caractéristiques plus pertinentes. L'ensemble d'entraînement se compose de 927 images CTA, et la fonction de coût est une combinaison de BCE et de Dice utilisée pour optimiser les poids du réseau. La sélection des patchs d'entraînement s'est effectuée en choisissant la moitié contenant des anévrismes. L'évaluation de cette approche a été menée sur un ensemble de données composé de 211 images CTA, comprenant un total de 46 anévrismes. Les résultats montrent une sensibilité de détection de 76,1%, avec 0,27 FP/cas. Une caractéristique notable de cette méthode réside dans la qualité de la vérité terrain utilisée, où la présence d'anévrismes a été doublement vérifiée par la modalité DSA.

Contrairement à la majorité des travaux existants qui s'appuient sur des réseaux de segmentation, une approche basée sur un réseau de détection d'objets utilisant des ancres a été introduite en 2021 [318]. Le réseau proposé représente une amélioration du réseau RPN [239], incorporant des modules d'attention CBAM [310] et des couches de convolution dilatées [55]. Le module CBAM, englobant à la fois l'attention spatiale et l'attention de canal, améliore la discernement des caractéristiques et des régions anévrismales dans les images. Parallèlement, l'utilisation de convolutions dilatées facilite l'intégration d'un contexte plus large lors de l'extraction des caractéristiques. Structurée autour de trois ancres, cette architecture divise les images d'entrée de dimensions $128 \times 128 \times 128$ voxels en cellules de taille $4 \times 4 \times 4$ voxels, prédisant les détections d'anévrismes sous forme de boîtes englobantes 3D. La phase d'entraînement a impliqué 534 images CTA, comprenant au total 688 anévrismes, suivie d'une évaluation sur un ensemble de 400 images contenant au total 144 anévrismes. De manière significative, cette approche a démontré une amélioration substantielle de la sensibilité de détection des radiologues, passant de 79,1 à 88,9%. Cependant, malgré la sensibilité accrue et l'ampleur des données d'entraînement, cette méthode présente un taux élevé de faux positifs (13,8 FP/cas), soulignant le défi complexe de trouver un équilibre optimal entre une sensibilité élevée et des faux positifs réduits. De plus, il convient de noter la complexité inhérente de l'architecture proposée, caractérisée par un nombre significatif de cellules à traiter et un grand déséquilibre entre les cellules positives et négatives. En outre, l'efficacité et la contribution des mécanismes d'attention dans ces architectures n'ont pas été entièrement démontrées.

Évaluation et comparaison des méthodes

Les méthodes 3D pour la détection des anévrismes, comme rapporté dans la littérature, ont démontré des sensibilités comparables à celles des experts, mais souffrent d'un taux élevé de faux positifs, restreignant ainsi leur utilité clinique. L'évaluation et la comparaison de ces méthodes sont compliquées par divers facteurs, notamment le manque de bases de données partagées, de métriques standardisées, et de descriptions détaillées des opérations de post-traitement, surtout dans les méthodes basées sur des réseaux de segmentation d'images. Pour relever ces défis et permettre une comparaison objective, deux compétitions ont été organisées lors de la conférence MICCAI 2020. La compétition ADAM a porté sur la détection et la segmentation des anévrismes à partir de l'imagerie TOF-MRA, tandis que la compétition CADA a traité de la détection, la segmentation et l'estimation du risque de rupture des anévrismes en utilisant des images 3DRA. Chaque compétition a utilisé deux ensembles de données : un ensemble d'entraînement (ADAM : 113 images avec 129 anévrismes ; CADA : 109 images avec 127 anévrismes) et un ensemble privé pour l'évaluation finale (ADAM : 141 images avec 153 anévrismes ; CADA : 22 images avec 30 anévrismes). On remarque la participation de 11 équipes pour ADAM et de 8 équipes pour CADA, mais seule une minorité d'entre elles a fourni des descriptions détaillées de leurs solutions. Les résultats de ces compétitions dans la détection des anévrismes seront analysés pour mettre en lumière les avancées et les défis subsistants. Un résumé de ces résultats est présenté dans la table 3.2.

Compétitions	Méthodes	Données d'entraînement		Réseau CNN	Fonction de coût	Évaluation		
		Dimension (voxels)	Sélection			Sensibilité (%)	FP/cas	Score F2 (%)
ADAM	abc [326]	128 × 128 × 128	Aléatoire	UNet [65]	BCE + Dice	68	0,40	
	xlim [250]	16 × 144 25 × 25 × 25	Vaisseaux	Double entrée	BCE	70	4,03	
	mibaumgartner [26]	224 × 224 × 56	Aléatoire (50% *)	nnDetection [27]	BCE + IoU	67	0,13	
	unil-chuv [80]	64 × 64 × 64	Polygone Willis	UNet [65]	BCE + Dice	68	2,50	
	joker [321]	192 × 224 × 56	(33% *)	UNet [65]	Dice	63	0,16	
	uw2 [197]	256 × 224 × 56	Aléatoire (33% *)	nnUNet [135]	BCE + Dice + TopK	63	1,08	
	JunMa [186]	256 × 224 × 56	Aléatoire (33% *)	nnUNet [135]	BCE + Dice	61	0,18	
	kubiac [76]	64 × 64 × 64	(54% *)	UNet [65]	BCE + Dice	60	0,36	
CADA	Mediclouds [145]	56 × 56 × 56	Vaisseaux	Attention UNet [218]	BCE + Dice			91,8
	ibbm [265]	128 × 128 × 128	(75% *)	UNet [65]	BCE + Dice			85,6
	jmkuhnigk [138]	28 × 28 × 28	Vaisseaux	UNet [65]	BCE			83,9
	mibaumgartner [26]	224 × 224 × 56	Aléatoire (50% *)	nnDetection [27]	BCE + GIoU			77,7
	JunMa [186]	256 × 224 × 56	Aléatoire (33% *)	nnUNet [135]	BCE + Dice			85,0

TABLE 3.2 – Analyse comparative des compétitions ADAM (en haut) et CADA (en bas) pour la tâche de détection d'anévrismes : Prédominance des approches en 3D utilisant des réseaux de segmentation d'images et le coefficient de Dice comme fonction de coût. L'astérisque (*) indique le rapport entre le nombre de patchs positifs et le nombre total de patchs dans l'ensemble d'entraînement.

Échantillonnage et augmentation de données Face au déséquilibre entre la classe minoritaire des anévrismes et la classe majoritaire des autres tissus cérébraux, différentes stratégies d'échantillonnage et d'augmentation des données ont été utilisées. Les patchs positifs sont généralement concentrés sur les anévrismes en raison de leur faible prévalence. La sélection des patchs négatifs varie, allant d'une sélection aléatoire pour la majorité des méthodes, à une sélection centrée sur l'arbre vasculaire [76], voire une sélection plus sophistiquée guidée par des a priori médicaux, tels que les régions vasculaires fréquemment touchées par des anévrismes, notamment le polygone de Willis [80]. Le choix du nombre de patchs positifs et négatifs pendant l'entraînement peut influencer le processus d'apprentissage en se concentrant sur la classe majoritaire. Pour remédier à ce déséquilibre, différentes méthodes d'échantillonnage entre les patchs positifs et négatifs ont été examinées. Par exemple, 33% des patchs ont été sélectionnés comme positifs, pouvant atteindre jusqu'à 50% voire même être étendus jusqu'à 75% d'échantillons positifs.

Afin d'accroître à la fois la quantité et la diversité des données d'entraînement tout en prévenant le problème de sur-apprentissage, diverses opérations d'augmentation de données aléatoires ont été utilisées, incluant la rotation, l'effet miroir et la translation. De plus, une variabilité dans les dimensions des images a été observée. L'exploration de la combinaison entre de petits patchs 3D ($25\times25\times25$ voxels) et les images MIP (16×144 pixels) a montré un score élevé de sensibilité, cependant, le taux de faux positifs reste une limitation majeure de cette approche, atteignant 4.03 FP/cas [250]. D'autre part, l'utilisation de patchs contextuels de grande dimension, tels que $128\times128\times128$ voxels et $224\times224\times56$ voxels, a présentée des performances supérieures en termes d'équilibre entre la sensibilité et les faux positifs. Il convient cependant de noter que cette approche nécessite une puissance de calcul importante pour l'entraînement et des données de patients sains pour garantir une modélisation fiable entre les anévrismes, les vaisseaux sanguins et les autres tissus cérébraux. Extraire des patchs sans anévrismes de grande dimension est plus difficile chez les patients porteur d'anévrismes.

Architectures de réseaux et fonctions de coût La plupart des architectures CNNs utilisées dans les compétitions ADAM et CADA sont des réseaux de segmentation d'images, avec une prédominance du réseau 3D UNet [245]. Un framework auto-configurable, nnUNet [135], est largement adopté en raison de sa simplicité d'utilisation et de son adaptation automatique aux données d'entraînement. Il permet de personnaliser et d'entraîner une variante du réseau 3D UNet en se basant sur des règles et des heuristiques développées à partir de l'expertise en plusieurs problèmes de segmentation médicale (voir Figure 3.8). La personnalisation inclut des modifications dans le nombre de couches de convolution et de sous-échantillonnage, ainsi que des changements dans la résolution spatiale et les dimensions des patchs d'entraînement. Le nnUNet utilise des techniques avancées, telles que la supervision profonde, pour améliorer la précision de la segmentation et l'algorithme de descente de gradient stochastique (SGD) pour

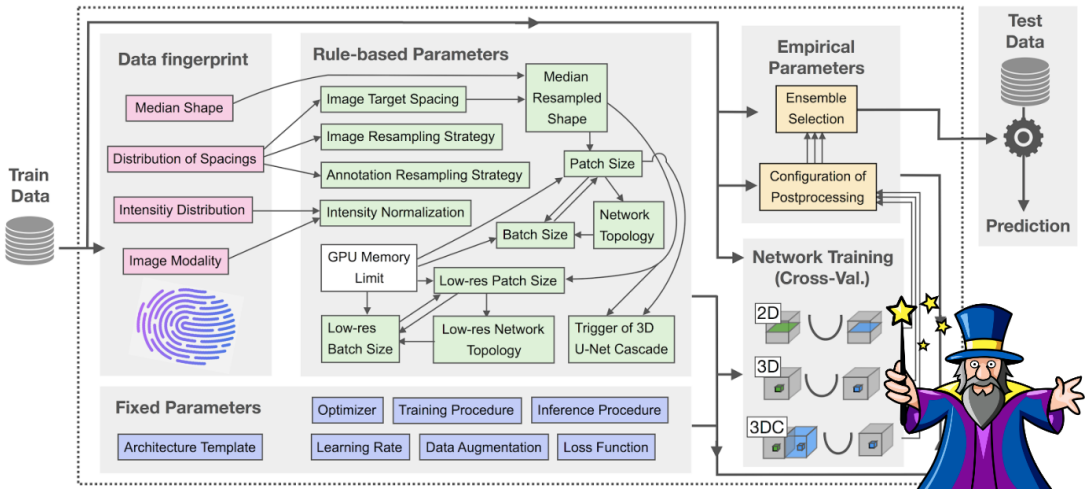


FIGURE 3.8 – nnUNet [135] : un framework auto-configurable pour la segmentation d'images médicales, présentant une configuration automatique adaptée aux besoins de chaque ensemble de données.

l'optimisation sur 1000 époques avec un lot d'entraînement composé de 4 patchs. Sa fonction de coût combine la BCE et le score Dice. De plus, le nnUNet intègre des processus de pré-traitement automatisés, tels que le recadrage des données aux zones de valeurs non nulles et la normalisation z-score des images. Des techniques d'augmentation de données, telles que les rotations, les effets miroir et les corrections gamma, sont également appliquées pendant l'entraînement. Ces caractéristiques font du nnUNet un outil robuste et flexible adapté à un large éventail de problèmes en segmentation médicale.

La préférence pour les réseaux de segmentation dans les tâches de détection découle de la complexité inhérente aux réseaux de détection d'objets. Ces derniers requièrent la configuration de multiples fonctions de coût pour simultanément traiter les paramètres de localisation (MAE, MSE) et la classification (BCE, Focal) des boîtes englobantes. De plus, leur implémentation implique l'utilisation de stratégies d'échantillonnage avancées, telles que le "hard negative mining," la suppression non maximale (NMS), ainsi que la gestion des boîtes d'ancres en termes de taille, de nombre et de localisation [139].

Une seule méthode [26] a adopté une approche de détection d'objets au sein des compétitions ADAM et CADA. Elle a utilisé le framework auto-configurable nnDetection [27] publié en 2021, pour détecter les anévrismes avec des boîtes englobantes 3D. Simaire à nnUNet, ce framework se base sur une série de règles et d'heuristiques pour sélectionner les données d'entraînement et configurer le réseau de détection (voir Figure 3.9). Il utilise le réseau RetinaUNet [139] qui génère des prédictions multi-échelles et emploie un ensemble de 27 boîtes d'ancres. Ce réseau combine les fonctionnalités de détection d'objets propres à RetinaNet [176] avec une tâche auxiliaire de segmentation, ce qui lui permet d'intégrer des informations sémantiques lors de la détection des anévrismes. L'optimisation de la classification des boîtes englobantes se fait via la BCE, tandis que la localisation utilise la fonction de coût IoU généralisée (GIoU) [242], avec un entraînement réparti sur 60 époques et des lots de 4 patchs pour un total de 2500 itérations par époque. Par ailleurs, la majorité des méthodes ont privilégié l'utilisation de techniques d'apprentissage ensembliste afin de diminuer la variabilité des données d'entraînement.

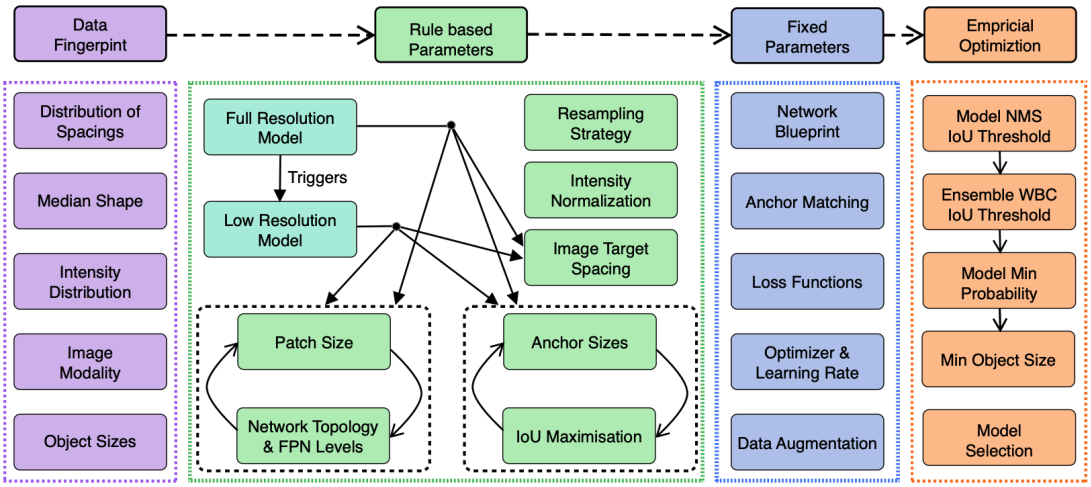


FIGURE 3.9 – Framework nnDetection [27] pour la détection d'objets : choix automatiquement des données d'apprentissage et ajustement de la structure du réseau selon plusieurs règles.

Métriques d'évaluation La compétition ADAM se concentre sur la localisation des anévrismes via leurs coordonnées de centre, tandis que CADA se concentre sur leur détection à l'aide de boîtes englobantes. Dans la compétition ADAM, l'évaluation repose sur la distance euclidienne entre les coordonnées prédites du centre et la vérité terrain, exprimée sous forme de sphère. Une prédiction est considérée comme positive (TP) si son centre se trouve à l'intérieur d'une sphère de vérité terrain. Cependant, l'utilisation de réseaux de segmentation pour de telles tâches de détection comporte des limitations. Ces réseaux nécessitent des étapes de post-traitement pour obtenir les coordonnées et les dimensions des objets détectés, comprenant des opérations telles que le seuillage, l'extraction et l'analyse des composantes connexes (CC). De plus, ces opérations requièrent la sélection du score de confiance par CC parmi plusieurs valeurs de voxels, ce qui peut potentiellement avoir un impact négatif sur les métriques d'éva-

luation et les performances par rapport aux méthodes directes de détection d'objets qui intègrent ces étapes dans leur procédure d'apprentissage [139]. Par exemple, en se basant sur les critères d'évaluation d'ADAM, plusieurs prédictions de petite taille peuvent être classées comme vrais positifs, tout comme une prédiction de grande taille centrée sur la vérité terrain, sous la condition que la distance entre le centre de l'anévrisme et les prédictions soit inférieure au rayon de l'anévrisme. Par ailleurs, augmenter le seuil de classification peut paradoxalement augmenter le nombre de CC et donc les détections (voir Figure 3.7c).

En revanche, l'utilisation de mesures basées sur l'indice d'intersection sur union (IoU), similaire à celles adoptées par CADA, se révèle plus appropriée. Ces mesures se fondent sur l'intersection entre les prédictions et la vérité terrain, représentée sous la forme de boîtes englobantes. Dans ces approches, un TP est identifié lorsque cette intersection dépasse un seuil prédéfini (IoU$>$ 0% dans la compétition CADA). Une autre considération importante lors de l'évaluation de ces deux compétitions réside dans l'utilisation de métriques par rapport à un seuil de classification fixe. Néanmoins, l'usage de ces métriques peut restreindre la comparaison équitable entre différentes méthodes. Ainsi, le recours à des métriques basées sur des représentations graphiques, telles que la précision moyenne, offre une comparaison plus objective et globale des différentes approches.

Bilan des compétitions Une variété de stratégies ont été utilisées pour aborder les problèmes liés au déséquilibre des classes et à la limitation des données dans la détection des anévrismes intracrâniens. L'adoption de patchs de grande taille a favorisé un meilleur équilibre entre sensibilité et taux de faux positifs, nécessitant cependant des ressources importantes en calcul et en données. Les réseaux de segmentation d'images se sont distingués comme l'approche prédominante, tandis que les réseaux de détection d'objets ont révélé leur efficacité, suggérant la nécessité d'une investigation plus poussée. Malgré certaines limitations dans les métriques d'évaluation, la compétition ADAM s'est distinguée par son ensemble de données d'évaluation plus vaste, incluant 141 images révélant 153 anévrismes, ce qui a permis une analyse plus approfondie en comparaison avec CADA.

La méthode classée en tête, abc [326], a démontrée une sensibilité médiocre de 68% et un bon taux de faux positifs de 0,4 FP/cas. Cette performance souligne les défis à relever pour l'application clinique, notamment pour la détection des anévrismes de petite taille ($<$ 5 mm). Une étude distincte [308] a indiqué que les experts atteignaient une sensibilité globale de 87% dans la détection des anévrismes, mais seulement 38% pour ceux de moins de 3 mm et 94% pour ceux de plus de 3 mm. En comparaison, la méthode xlim [250] qui a obtenu le meilleur score de sensibilité lors de la compétition ADAM a présenté une sensibilité de 71% pour les anévrismes mesurant entre 3,54 mm et 4,98 mm, tandis qu'elle a atteint 95% pour ceux dépassant la dimension de 4,98 mm.

3.4 Bases de données et annotation des anévrismes intracrâniens

Les bases de données jouent un rôle essentiel dans le développement, l'entraînement et l'évaluation des algorithmes d'apprentissage. La performance des modèles dépend de divers facteurs, notamment la quantité, la qualité, la diversité et la représentativité des données disponibles. Des initiatives notables, telles que ImageNet [78] en 2009 et COCO [177] en 2014, ont été à l'origine d'avancées significatives en mettant à disposition des millions d'images naturelles annotées. Cependant, l'accès à des ensembles de données annotées en imagerie médicale demeure une problématique. La compétition BRATS [1] a constitué l'une des bases de données annotées les plus vastes en imagerie médicale, comprenant un total de 4 500 images IRM résultant d'une collaboration à grande échelle. Cependant, cette quantité reste modeste en comparaison avec les millions d'images d'ImageNet. De plus, les ensembles de données spécifiquement dédiés aux anévrismes restent extrêmement limités. Même les bases de données privées utilisées dans les travaux existants ne comprennent que quelques centaines d'images. Parallèlement à la généralisation des techniques d'imagerie médicale et à la croissance continue de l'apprentissage profond, ces dernières années ont vu l'émergence d'initiatives visant à fournir des ensembles de données annotées spécifiquement dédiées à l'imagerie des vaisseaux cérébraux, y compris les anévrismes.

3.4.1 Bases de données disponibles

Dans cette section, nous explorons les bases de données clés dans le domaine de la détection et de la segmentation des anévrismes intracrâniens. L'année 2020 a été marquée par l'organisation des compétitions ADAM [285] et CADA [137], qui se sont concentrées sur l'évaluation de diverses méthodes dans ce domaine en utilisant des images TOF-MRA et 3DRA. Une avancée significative a été réalisée fin 2022 avec la publication d'un ensemble de données TOF-MRA dédié à la détection des anévrismes par l'hôpital universitaire de Lausanne (CHUV) [80]. De plus, l'ensemble de données IntrA [319] propose une approche représentant les vaisseaux sanguins et les anévrismes sous forme de nuages de points et de maillages. Enfin, nous présenterons également la base de données IXI [68], qui regroupe des images MRA cérébrales de patients sans anévrismes.

Compétition ADAM

La compétition ADAM a utilisé un ensemble d'images TOF-MRA du cerveau collectées aux Pays-Bas entre 2001 et 2019. Ces images ont été obtenues à des intensités de champ magnétique de 1T, 1,5T ou 3T, avec une variation de l'espacement des voxels dans le plan de 0,195 à 1,04 mm et une épaisseur de coupe de 0,4 à 0,7 mm. Les dimensions des images sont d'environ $512 \times 512 \times 140$ voxels, avec un prétraitement incluant la correction N4 [289] pour améliorer leur qualité. Chaque image TOF-MRA est accompagnée d'une image IRM anatomique, en séquence T1, T2 ou FLAIR. Concernant la démographie des patients, l'âge médian est de 55 ans, et les femmes représentent 75% de l'échantillon.

L'ensemble de données utilisé dans la compétition se compose de divers cas, incluant des anévrismes non traités, des anévrismes traités (par coils ou clips), ainsi que des images sans anévrismes. Au total, cet ensemble compte 254 images. Parmi ces images, on dénombre 282 anévrismes non traités répartis sur 208 images, tandis que 46 images ne contiennent pas d'anévrismes. De ces 208 images, 156 forment un sous-ensemble spécifique, comprenant des images appariées provenant du même patient. Chaque paire d'images est composée d'une image de référence et d'une image de suivi acquise 6 mois plus tard. Par ailleurs, 52 images présentent des cas d'anévrismes non traités multiples, tandis que 59 images contiennent des anévrismes ayant été traités. Les anévrismes non traités observés dans ces images sont principalement de petite taille, avec un diamètre variant de 1,0 à 15,9 mm, et une médiane de 3,6 mm. Ils sont situés à diverses localisations, notamment au niveau des artères cérébrales antérieures (ACA), des artères communicantes antérieures (ACom), des artères carotides internes (ICA), des artères communicantes postérieures (PCom), des artères cérébrales moyennes (MCA), et de la circulation postérieure.

Cet ensemble de données est divisé en un ensemble d'entraînement et un ensemble de test privé, avec une distribution équilibrée entre les images sans et avec anévrismes. L'ensemble d'entraînement comprend 113 images, parmi lesquelles 93 présentent au moins un anévrisme non traité (composées de 35 images de référence, 35 images de suivi et 23 images uniques), tandis que 20 images ne contiennent pas d'anévrisme. De manière similaire, l'ensemble de test est composé de 141 images, dont 115 présentent au moins un anévrisme non traité (composées de 43 cas de référence, 43 cas de suivi et 29 images uniques), tandis que 26 images ne présentent pas d'anévrisme. Il convient de noter que cet ensemble de données propose 58 images indépendantes pour l'entraînement, certaines étant obsolètes en ce qui concerne les normes actuelles en matière de scanners 3T à haute résolution. Il est important de mentionner que le site web de la compétition n'est plus actif, ce qui restreint l'accès à la base de données et empêche toute participation ultérieure à cette compétition.

Compétition CADA

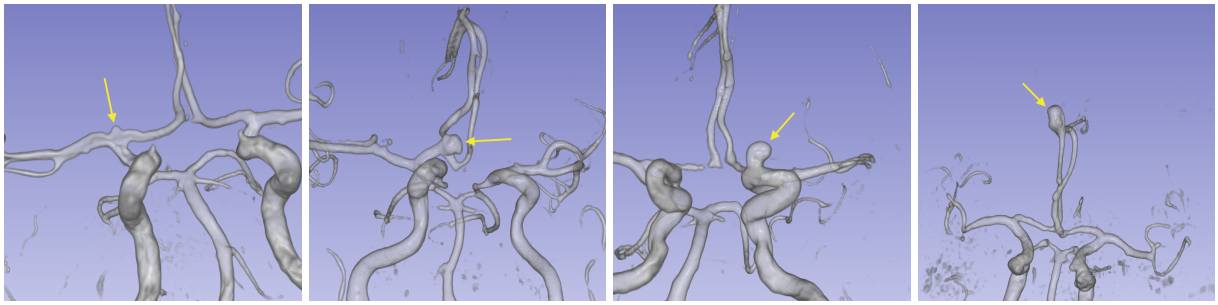
L'ensemble de données de la compétition CADA se compose de 131 images 3DRA obtenues à partir de divers patients, excluant les examens de suivi, les patients ayant subi un traitement d'anévrismes, ainsi que ceux sans anévrismes. Ces images ont été acquises au service de neurochirurgie de Helios Klinikum Berlin-Buch en Allemagne. Une injection manuelle d'un agent de contraste a été réalisée soit dans la carotide interne gauche ou droite pour les anévrismes antérieurs, soit dans l'artère vertébrale pour les anévrismes postérieurs. La reconstruction d'un volume d'intérêt spécifié par un neurochirurgien a généré une série de 440 coupes d'images, chacune avec une résolution de 512×512 pixels. Ces coupes forment un volume de $512 \times 512 \times 440$ voxels, avec une taille de voxel isotrope de 0,25 mm^3. Le temps d'acquisition par image était de 5 secondes.

Les 131 images 3DRA présentent un nombre variable d'anévrismes, variant de 1 à 7, avec environ 10% des images contenant plusieurs anévrismes. Une division de l'ensemble de données a été effectuée, créant un ensemble d'entraînement composé de 109 images contenant 127 anévrismes, et un ensemble de test comprenant 22 images avec 30 anévrismes. La taille moyenne des anévrismes dans l'ensemble d'entraînement est de 9 ± 5 mm pour les femmes et de 8,7 ± 3,8 mm pour les hommes. Dans l'ensemble de test, les anévrismes ont un diamètre moyen de 9 ± 6,3 mm pour les femmes et de 10,1 ± 6,1 mm pour les hommes. Il est important de noter que cet ensemble de données comprend un nombre limité de petits anévrismes (< 5 mm). La compétition se concentre principalement sur les anévrismes de plus grande taille, qui sont considérés à haut risque de rupture. De plus, il est à noter que la modalité choisie (3DRA) diffère considérablement de la TOF-MRA et n'est pas une modalité diagnostique.

Base de données CHUV

En 2022, la base de données TOF-MRA de l'hôpital universitaire de Lausanne (CHUV) en Suisse a été rendue publique [80]. Issue d'une étude menée entre 2010 et 2015, cette base regroupe les données de 284 patients consécutifs ayant subi des examens d'imagerie TOF-MRA. Des critères d'exclusion ont été appliqués pour éliminer les patients présentant des anévrismes rompus, traités, ainsi que d'autres anomalies vasculaires telles que des anévrismes thrombosés et infundibulaires. Les données d'imagerie ont été acquises à des intensités de champ magnétique de 1,5T ou 3T, avec des résolutions spatiales variables, allant de $0,27 \times 0,27 \times 0,5$ mm^3 à $0,4 \times 0,4 \times 0,7$ mm^3. Chaque image TOF-MRA est accompagnée d'une image pondérée en séquence T1.

Cet ensemble de données comprend 296 images TOF-MRA, réparties en deux catégories distinctes. Parmi elles, 168 images (provenant de 157 patients différents) montrent un ou plusieurs anévrismes, tandis que les 128 autres images ne présentent aucun anévrisme (provenant de 127 patients distincts). Le nombre total d'anévrismes dans cet ensemble s'élève à 198, comprenant 178 anévrismes de type sacculaire et 20 de type fusiforme. Les patients atteints d'anévrismes ont en moyenne 56 ans, avec une nette prédominance des femmes, représentant 66% de la population étudiée. Les dimensions des anévrismes varient de 1 à 20 mm (voir Figure 3.10) et leur classification est basée sur le score PHASES pour l'estimation du risque de rupture [34]. Parmi les 198 anévrismes identifiés, la plupart mesurent moins de 7 mm, soit 180 cas au total. Il y a également 7 anévrismes mesurant entre 7 mm et 10 mm, 10 anévrismes mesurant entre 10 mm et 19 mm, ainsi qu'un unique anévrisme mesurant 20 mm ou plus. En ce qui concerne leur emplacement, 29,8% des anévrismes se trouvent au niveau de l'artère carotide interne (ICA), tandis que 28,8% sont situés au niveau de l'artère cérébrale moyenne (MCA), les autres étant répartis entre l'artère cérébrale antérieure (ACA), l'artère communicante postérieure (PCom) et l'artère cérébrale postérieure (PCA). Cette base de données est donc particulièrement pertinente dans notre contexte applicatif.



(a) 1,85 mm (b) 4,41 mm (c) 6,20 mm (d) 7,75 mm

FIGURE 3.10 – Exemples d'images TOF-MRA de la base de données CHUV [80] : anévrismes de tailles variées sont indiqués.

Base de données IntrA

L'ensemble de données IntrA [319] aborde les questions éthiques liées au partage de données médicales en mettant à disposition des données concernant les vaisseaux sanguins et les anévrismes sous forme de nuages de points et de maillages. Son objectif est de faciliter l'utilisation des données pour le

développement d'algorithmes de segmentation et de planification d'interventions chirurgicales liées aux anévrismes, telles que la pose de clips au niveau du collet de l'anévrisme. Cette ressource comprend 103 modèles 3D complets de vaisseaux cérébraux, générés à partir de 103 images TOF-MRA de dimensions $512 \times 512 \times 180$ voxels avec une résolution spatiale de $0,469 \times 0,469 \times 1$ mm^3. Chaque image produit environ 300 images de coupe 2D, qui subissent ensuite un processus de reconstruction, d'annotation par des experts et de réduction du bruit pour créer les modèles 3D des vaisseaux sanguins. À partir de ces modèles 3D, la base de données IntrA génère automatiquement 1909 segments de vaisseaux, comprenant 1694 segments de vaisseaux sains et 215 segments contenant des anévrismes.

Cet ensemble de données a été moins fréquemment utilisé. Par exemple, dans [327], il a été exploité pour deux tâches principales : la classification des segments en fonction de la présence ou de l'absence d'anévrismes et la segmentation des vaisseaux-parents et des anévrismes. Cependant, cet ensemble de données présente certaines limites, notamment en ce qui concerne les petits anévrismes, dont le signal diminue au cours du processus de reconstruction, rendant leur annotation difficile. De plus, certaines annotations d'anévrismes peuvent s'avérer incomplètes et difficiles à distinguer des vaisseaux sains.

Base de données IXI

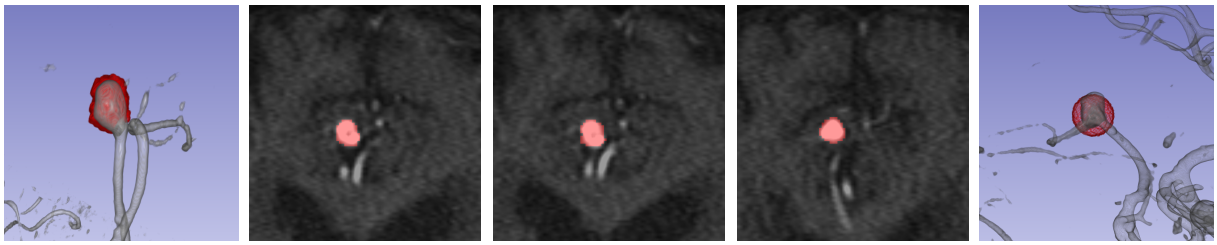
En 2015, la base de données IXI [68] a été rendue publique, comprenant un ensemble d'images IRM du cerveau provenant de patients en bonne santé. Ces données, totalisant 584 images, ont été acquises entre 2005 et 2006 dans trois hôpitaux distincts de Londres, utilisant des scanners de 1,5T et 3T. Chaque ensemble de données de patient englobe des séquences pondérées en T1, T2, PD, une image d'angiographie MRA, et une image pondérée en diffusion. Les images présentent une résolution spatiale d'environ $0.9375 \times 0.9375 \times 1.2$ mm^3, avec des dimensions de $160 \times 192 \times 224$ voxels. Cette base de données se compose d'images cérébrales de patients en bonne santé, restreignant ainsi son applicabilité dans le contexte des anomalies vasculaires, telles que les anévrismes intracrâniens. Sa principale utilisation s'est orientée vers des applications spécifiques, notamment l'amélioration de la résolution spatiale [100], la transformation entre différentes modalités d'imagerie [74], le recalage d'images [59], et la reconstruction d'images [13]. Cette base pourrait cependant constituer une base de patients sains.

3.4.2 Techniques d'annotation

Outre la confidentialité des données des patients, un défi majeur associé au manque d'ensembles de données annotées à grande échelle réside dans la complexité de l'annotation des images médicales [236]. Pour les anévrismes, cette procédure exige un étiquetage voxel par voxel précis et détaillé pour délimiter les zones affectées dans des images 3D riches en informations visuelles. Cette tâche est réalisée manuellement par les radiologues, qui examinent les coupes 2D et annotent les anévrismes sur les images les contenant. Il est important de souligner que cette procédure d'annotation est intrinsèquement complexe, chronophage et exige une forte expertise. Par exemple, annoter un anévrisme de 5 mm peut nécessiter environ 93 secondes d'attention continue en utilisant des logiciels d'annotation dédiés [80].

L'annotation des anévrismes est sujette à une variabilité significative en raison des variations de taille, de nombre et de localisation de ces anévrismes dans les images 3D. Cette variabilité se traduit par deux défis principaux : d'une part, le risque de non-détection, entraînant par conséquent une absence d'annotation pour certains anévrismes, et d'autre part, la possibilité d'annotations imprécises, notamment en ce qui concerne les contours et les limites de résolution (voir Figure 3.11). Ces défis ont un impact direct sur la fiabilité des données de référence utilisées pour l'entraînement et l'évaluation des algorithmes. Pour atténuer cette variabilité, l'approche courante consiste à impliquer plusieurs annotateurs afin d'obtenir un consensus, réduisant ainsi les divergences et garantissant une annotation plus fiable. Cependant, cette approche est rarement adoptée, et la documentation des mesures de variabilité inter-annotateurs est souvent omise dans la littérature concernant les anévrismes. Par ailleurs, d'autres facteurs viennent complexifier la fiabilité du processus d'annotation, tels que la qualité variable des images en raison des différences d'équipement d'imagerie, la présence de bruit ou d'artefacts, ainsi que les variations anatomiques entre les patients. Enfin, le coût élevé et la disponibilité limitée d'experts qualifiés pour effectuer ces annotations constituent une autre barrière significative à la création de vastes ensembles de données annotées de manière fiable et précise.

En 2022, une approche alternative a été introduite dans la base de données CHUV pour rationaliser et accélérer le processus d'annotation des anévrismes [80]. Cette approche repose sur l'approximation de chaque anévrisme par une sphère englobante, indépendamment de sa forme géométrique (voir Figure 3.11e). L'objectif principal était de réduire considérablement le temps requis pour l'annotation, passant ainsi de 93 secondes à 23 secondes, ce qui a facilité la création d'une base de données comprenant 296 images. Cette méthode d'annotation a également démontré des performances comparables à l'annotation voxel par voxel pour les algorithmes de détection et de segmentation des anévrismes. Toutefois, une limitation inhérente à cette approche réside dans le fait que les annotations peuvent englober les vaisseaux parents avec les anévrismes, ce qui pourrait entraver la capacité des algorithmes d'apprentissage à différencier efficacement les vaisseaux sains des anévrismes. Il est à noter qu'au début de la thèse en 2020, nous avons opté pour ce type d'annotation [15] (voir également Chapitre 4).



(a) (b) (c) (d) (e)

FIGURE 3.11 – Variabilité d'annotation voxel par voxel d'un anévrisme intracrânien de 7,75 mm : (a) vue 3D et (b, c, d) exemples de coupes axiales. (e) Annotation approximative d'un anévrisme de 5,76 mm par une sphère englobante.

Conclusion

Ce chapitre a passé en revue l'état de l'art de la détection des anévrismes intracrâniens, en examinant diverses techniques, des méthodes classiques de traitement d'images aux approches plus récentes basées sur les réseaux de neurones en 3D. L'analyse a révélé des avancées notables dans ce domaine depuis le début de la thèse, soulignant l'importance de rester à jour avec ces développements. Deux approches principales pour la localisation des anévrismes ont été identifiées : la segmentation et la détection d'objets. Les méthodes de segmentation prédominent et exigent des étapes de post-traitement spécifiques, ainsi que l'utilisation de mesures d'évaluation appropriées pour garantir une évaluation fiable. En revanche, les méthodes de détection d'objets reposent sur des représentations telles que les boîtes englobantes, nécessitant des fonctions de coût spécifiques et des configurations d'ancres adaptées aux caractéristiques des données.

Bien que les méthodes existantes atteignent des niveaux de sensibilité comparables à ceux des experts pour les anévrismes de taille moyenne à grande, elles présentent des limitations lorsqu'il s'agit de détecter de petits anévrismes (moins de 5 mm) et sont sujettes à des taux élevés de faux positifs, ce qui peut limiter leur utilisation en milieu clinique. De plus, la qualité et la quantité des ensembles de données annotés, cruciales pour l'entraînement et l'évaluation des algorithmes, sont compromises par le processus exigeant d'annotation et par la variabilité inhérente des annotations, influencée par les modalités d'imagerie utilisées et l'expertise des annotateurs. Dans les chapitres à venir, nous exposerons nos contributions visant à résoudre ces défis découlant de la disponibilité limitée des données, de la rareté des cas d'anévrismes et des contraintes liées à l'évaluation des méthodes. Le chapitre 4 présente une méthode de détection basée sur la segmentation, tandis que le chapitre 5 se concentre sur une approche basée sur un réseau de détection d'objets. Enfin, le chapitre 6 présentera une extension de ces travaux à la question innovante de l'estimation de la pose des anévrismes détectés.

Chapitre 4

Stratégie de données efficace pour la détection d'anévrismes intracrâniens

Sommaire

4.1 Approches de détection d'anévrismes par segmentation 76
4.2 Stratégie d'échantillonnage et de génération de données 79
4.2.1 Collection et annotation de données 79
4.2.2 Approche par petits patchs 80
4.2.3 Sélection et synthèse de données 81
4.3 Réseau de neurones pour la détection des anévrismes 82
4.3.1 Architecture du réseau de segmentation UNet 82
4.3.2 Détails d'implémentation 83
4.3.3 Métriques et méthodologies d'évaluation 84
4.4 Étude expérimentale et discussion 85
4.4.1 Étude par ablation 85
4.4.2 Évaluation par validation croisée 87
4.4.3 Synthèse et analyse des résultats 89
4.5 Données de référence et intégration de l'a-priori médical 91
4.5.1 Consolidation des annotations et reproductibilité 91
4.5.2 Accélération de l'entraînement et optimisation des hyper-paramètres 92
4.5.3 Intégration de l'a-priori médical pour améliorer la détection 94

Ce chapitre présente notre première contribution à la détection des anévrismes à partir d'images TOF-MRA, dont une partie a été publiée dans notre article [15]. Notre approche repose sur une stratégie efficace de préparation et de gestion des données pour cette tâche. La section 4.1 synthétise les techniques de segmentation utilisées dans le contexte de la compétition ADAM [285]. Ensuite, la section 4.2 décrit notre stratégie de gestion des données, conçue pour relever les défis liés à la disponibilité limitée des données, au déséquilibre des classes et à la rareté des anévrismes dans les images. Nous y détaillons notre processus de collecte et d'annotation des données, l'utilisation de petits patchs, ainsi que nos méthodes de sélection et de synthèse des données pour l'entraînement des réseaux de neurones. La section 4.3 présente l'architecture du réseau 3D UNet choisie pour évaluer notre stratégie, en fournissant des informations sur ses composants, les détails d'implémentation, et les métriques d'évaluation adaptées à la tâche de détection d'objets. Par la suite, la section 4.4 se concentre sur l'étude expérimentale et les résultats obtenus. Nous menons une étude par ablation et une validation croisée pour évaluer l'efficacité de notre méthode, accompagnées d'une discussion analytique mettant en évidence les forces et les faiblesses de notre approche. Enfin, la section 4.5 traite des défis techniques rencontrés et des améliorations apportées à notre méthode, en mettant particulièrement l'accent sur les questions de reproductibilité et l'intégration de connaissances a priori pour guider et optimiser la détection des anévrismes.

4.1 Approches de détection d'anévrismes par segmentation

L'introduction des CNNs dans le domaine de la détection des anévrismes a marqué une avancée significative. Plusieurs systèmes, principalement axés sur des approches de segmentation d'images, ont montré des résultats prometteurs [81]. Toutefois, l'évaluation de ces systèmes a été souvent complexe en raison de l'usage de métriques et de méthodologies d'évaluation peu appropriées, appliquées sur des ensembles de données privés. Cette situation a compliqué toute comparaison objective entre les diverses méthodes. En fin d'année 2020, la compétition ADAM [285] a été organisée dans le but de favoriser le développement et l'évaluation de méthodes en se basant sur un ensemble de données d'imagerie TOF-MRA. La synthèse des méthodes proposées est présentée dans le table 4.1.

Dimensions de données

Dans le cadre de la compétition ADAM, la plupart des méthodes reposent sur des approches en 3D, exploitant les informations contextuelles le long des trois axes anatomiques. Un élément crucial de ces méthodes réside dans la détermination de la dimension des patchs utilisés pour l'entraînement. Bien que les dimensions de ces patchs varient, les stratégies les plus efficaces, qui maintiennent un équilibre entre sensibilité et réduction du taux de faux positifs, privilégient l'utilisation de patchs de grande dimension. Par exemple, la méthode gagnante, abc [326], a opté pour des patchs de $128 \times 128 \times 128$ voxels, tandis que la méthode joker [321], classée quatrième, a sélectionné des patchs de $192 \times 224 \times 56$ voxels. Il est essentiel de noter que l'utilisation de patchs de grande dimension exige d'importantes capacités de traitement graphique (GPU) et la nécessité d'inclure des données provenant de patients sains pour modeler adéquatement les zones exemptes d'anévrismes. Cette exigence peut restreindre l'application de telles méthodes dans des contextes où les ressources sont limitées. À l'inverse, l'usage de patchs de petite taille, adaptés aux environnements avec des capacités de calcul plus modestes, se traduit par une réduction des informations contextuelles disponibles, ce qui accroît le risque de faux positifs. Par exemple, les méthodes xlim [250] et unil-chuv [80] ont employé des patchs de $25 \times 25 \times 25$ voxels et $64 \times 64 \times 64$ voxels respectivement, entraînant des moyennes de 4,03 et 2,5 FP/cas.

Sélection et augmentation de données

Une variabilité a également été observée dans les approches de sélection des patchs d'entraînement à partir des données de chaque patient. Pour les patchs positifs, contenant des anévrismes, ils sont généralement centrés sur l'anévrisme lui-même. En revanche, les méthodes diffèrent dans leur approche pour la sélection des patchs négatifs, qui n'incluent pas d'anévrisme. Plusieurs méthodes optent pour une sélection aléatoire de ces patchs négatifs, à l'instar des méthodes uw2 [197] et junma [321] qui recourent au framework nnUNet [135]. D'autres méthodes préfèrent une approche plus ciblée, en utilisant les structures vasculaires comme guide pour placer les patchs négatifs le long des vaisseaux sanguins. Par exemple, la méthode unil-chuv [80] utilise des repères issus d'un atlas vasculaire recalé pour positionner les patchs dans des régions à risque d'anévrismes, telles que le polygone de Willis. Toutefois, l'efficacité de cette approche est fortement tributaire de la précision de l'étape de recalage des images. De plus, afin d'enrichir les données d'entraînement, diverses techniques d'augmentation sont employées, telles que les rotations,

Méthodes	Approche	Réseau	Patchs (voxels)	Ensemble	Post-traitement	Sensibilité (%)	FP/cas
abc [326]	Segmentation	3D UNet [65]	$128 \times 128 \times 128$	✓		68	0,40
xlim [250]	Segmentation	Double-Entrée	$25 \times 25 \times 25$		T	70	4,03
mibaumgartner [26]	Détection	nnDetection [27]	$224 \times 224 \times 56$	✓	T	67	0.13
unil-chuv [80]	Segmentation	3D UNet [65]	$64 \times 64 \times 64$		N, L	68	2,50
joker [321]	Segmentation	3D UNet [65]	$192 \times 224 \times 56$	✓		63	0,16
uw2 [197]	Segmentation	nnUNet [135]	$256 \times 224 \times 56$	✓	T	63	1,08
junma [186]	Segmentation	nnUNet [135]	$256 \times 224 \times 56$	✓	T	61	0,18
Kubiac [76]	Segmentation	3D UNet [65]	$64 \times 64 \times 64$	✓	L	60	0,36

TABLE 4.1 – Performances des meilleures méthodes de la compétition ADAM [285] : Prédominance des approches de segmentation et d'apprentissage ensembliste. Diverses opérations de post-traitement sont employées : filtrage des détections par nombre (N), par taille/volume (T) et par localisation (L).

les zooms, l'ajout de bruit gaussien et les effets miroirs. Cependant, l'évaluation de l'impact individuel de ces techniques de sélection et d'augmentation de données est souvent négligée par ces méthodes. De plus, l'utilisation de données synthétiques dans le contexte spécifique des anévrismes intracrâniens demeure peu explorée.

Problème de déséquilibre des classes

Le réseau de segmentation UNet [65] est utilisé par la majorité des méthodes proposées pour la détection des anévrismes. Cependant, ce choix se confronte au défi inhérent au déséquilibre entre les voxels représentant les anévrismes et ceux du fond dans l'ensemble d'entraînement. Diverses stratégies ont été déployées pour atténuer ce déséquilibre. Par exemple, dans certaines méthodes [321, 186], jusqu'à 33% des patchs d'entraînement ont été sélectionnés comme étant des patchs positifs, tandis que dans d'autres cas, cette proportion peut atteindre jusqu'à 50% [76]. En ce qui concerne les fonctions de coût, le coefficient de Dice, qui évalue la similitude entre la segmentation prédite et la vérité terrain, s'est révélé particulièrement efficace lorsque le déséquilibre des classes est prononcé [188, 30]. Par conséquent, plusieurs architectures de segmentation ont opté pour l'utilisation d'une combinaison de fonctions de coût, englobant la BCE, le coefficient de Dice, et parfois la fonction TopK [30]. Pour obtenir une segmentation encore plus précise, notamment au niveau des contours des anévrismes, certaines méthodes, comme Kubiac [76], ont introduit le coût de frontière [159], en complément de la BCE et du Dice. Cependant, l'efficacité de ces approches a été mise en question par des méthodes plus simples mais tout aussi efficaces, comme la méthode abc [326], qui a remporté la compétition en utilisant une combinaison de BCE et de Dice.

Modèles d'ensemble

L'approche ensembliste de l'apprentissage, adoptée par de nombreux participants à la compétition ADAM, vise à améliorer la fiabilité et la précision des modèles d'apprentissage. Elle consiste à entraîner divers modèles en utilisant différents ensembles de données, fonctions de coût, et architectures de CNNs. Les prédictions générées par ces modèles variés sont ensuite combinées à l'aide de techniques de post-traitement, ce qui conduit à des prédictions plus robustes. Un exemple de cette approche est la méthode Kubiac [76], qui a élaboré un ensemble de 18 modèles de segmentation utilisant diverses configurations, telles que BCE, Dice et le coût de frontière. Cette collection comprend des réseaux à double résolution, dont un réseau FCN [184] et deux réseaux 3D UNet. De même, la méthode joker [76] a généré des prédictions à partir de quatre modèles UNet entraînés séparément, puis les a combinées en utilisant un vote majoritaire. Cependant, les détails du post-traitement utilisé pour la combinaison des prédictions ne sont pas explicitement précisés par ces méthodes. De plus, cette approche exige des ressources importantes en termes de données et de puissance de calcul, ce qui peut limiter son utilisation dans des contextes à ressources limitées.

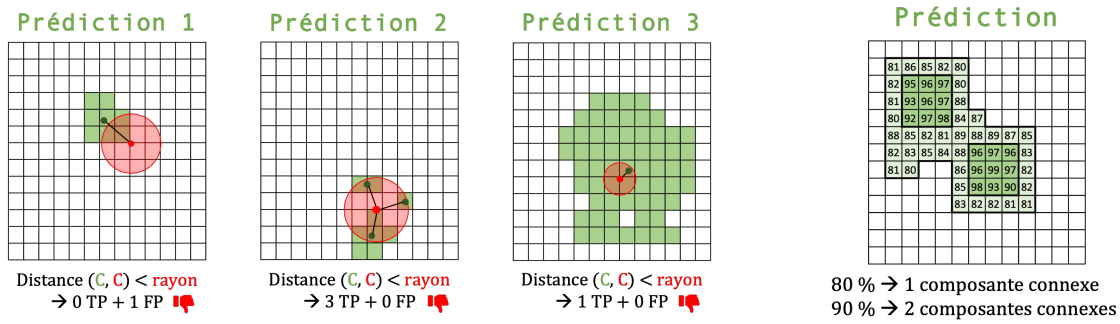
Opérations de post-traitement et métriques d'évaluation

Les réseaux de segmentation d'images produisent des cartes de probabilités à partir desquelles les prédictions sont extraites sous forme de composantes connexes (CC) en utilisant une opération de seuillage. Par la suite, un processus de filtrage est souvent employé pour analyser et sélectionner les CC, en prenant notamment en compte leurs dimensions et leur volume. De plus, chaque CC se voit attribuer un score de confiance basé sur plusieurs scores de probabilité de voxels. Par exemple, la méthode xlim [250] exclut les CC contenant moins de 10 voxels, tandis que la méthode unil-chuv [80] conserve uniquement les CC situées à une distance prédéterminée par rapport à des points de repère recalés et sélectionne 4 CC par image. Cependant, ces opérations de post-traitement peuvent influencer les détections extraites, nécessitant ainsi des ajustements adaptés aux caractéristiques des données et des objets analysés. De plus, une augmentation du seuil de classification appliqué aux cartes de probabilité a tendance à produire des CC de taille réduite et peut entraîner une augmentation du nombre de CC. Les CC ainsi extraites sont comparées aux données de vérité terrain en vue du calcul de métriques d'évaluation.

Les métriques d'évaluation adoptées dans le cadre de la compétition ADAM reposent sur la mesure de la distance euclidienne entre les centres des prédictions et ceux de la vérité terrain, en représentant les anévrismes par des sphères. Les coordonnées des centres de gravité (x, y, z) sont extraites à partir des CC : un vrai positif (TP) est identifié lorsque la sphère de vérité terrain inclut le centre d'une CC, tandis qu'un faux négatif (FN) se manifeste lorsque la sphère de vérité terrain ne contient aucun centre de

CC, et un faux positif (FP) se produit lorsque le centre d'une CC ne correspond à aucune vérité terrain. Cependant, il est important de noter que ces critères peuvent être moins précis dans certaines situations, notamment lors de la détection de multiples petites entités à l'intérieur d'une sphère de vérité terrain ou lorsqu'une grande prédiction est centrée sur l'anévrisme, ces deux cas pouvant tous deux être considérés comme des TP (voir Figure 4.1a). En revanche, des critères basés sur l'indice d'intersection sur l'union (IoU) semblent plus pertinents pour des évaluations plus précises [192].

L'évaluation de la compétition ADAM a été réalisée sur un ensemble de 141 images, contenant au total 153 anévrismes. Pour la comparaison et le classement des méthodes, les métriques de sensibilité et le nombre moyen de faux positifs ont été utilisés. Cependant, ces métriques ont été calculées en se basant sur un seuil de classification fixe, ce qui peut restreindre la précision de l'évaluation, étant donné que les méthodes réagissent de manière variable aux seuils de classification. De plus, ce seuil influence les caractéristiques des CC en termes de nombre et de dimensions (voir Figure 4.1b). Dans le but d'une évaluation plus objective, l'utilisation de métriques basées sur des représentations graphiques, telle que la précision moyenne (AP), évaluant les performances à divers seuils, s'avère plus appropriée et permet d'évaluer chaque méthode dans sa configuration optimale. Une autre limitation de l'évaluation adoptée dans la compétition ADAM est l'absence des scores de variabilité des performances pour chaque méthode. L'intégration de tels scores de variabilité serait bénéfique pour une comparaison plus approfondie des méthodes, en particulier dans un contexte de données limitées. Dans ce contexte, l'utilisation de la validation croisée offre une évaluation plus exhaustive et représentative de la robustesse des différentes méthodes sur des sous-ensembles de données.



(a) (b)

FIGURE 4.1 – Défis liés à l'utilisation de réseaux basés sur la segmentation pour les tâches de détection d'objets. (a) Métriques de la compétition ADAM : évaluation de la distance entre les centres des prédictions (en vert) et les annotations (en rouge). (b) Le seuil de classification affecte le nombre et les dimensions des composantes connexes extraites.

Contributions

Au cours des quatre dernières années, le domaine de la détection d'anévrismes à partir des images TOF-MRA a connu des progrès notables, impulsés en partie par la compétition ADAM. Malgré ces avancées, plusieurs défis demeurent, notamment la disponibilité restreinte de données, le déséquilibre des classes, ainsi que la diversité en taille, forme et emplacement des anévrismes. Pour y faire face, les approches précédentes ont largement adopté des réseaux CNNs avec l'utilisation de grands patchs, combinés à des techniques d'ensemble et diverses méthodes de post-traitement. Notre méthode se distingue par sa focalisation sur la gestion des données, adaptée spécifiquement aux particularités des anévrismes. Elle inclut une annotation rapide pour enrichir les ensembles de données et emploie des stratégies de synthèse et d'échantillonnage guidées, utilisant des patchs plus petits pour optimiser l'exploitation des données disponibles. L'efficacité de notre stratégie a été validée à travers l'utilisation d'un réseau UNet pour la segmentation, associé à des métriques et protocoles d'évaluation ciblés. Des tests d'ablation ont précisé l'impact de chaque élément de notre approche, et une validation croisée sur un ensemble élargi de données a permis une évaluation approfondie, y compris des analyses de variabilité.

4.2 Stratégie d'échantillonnage et de génération de données

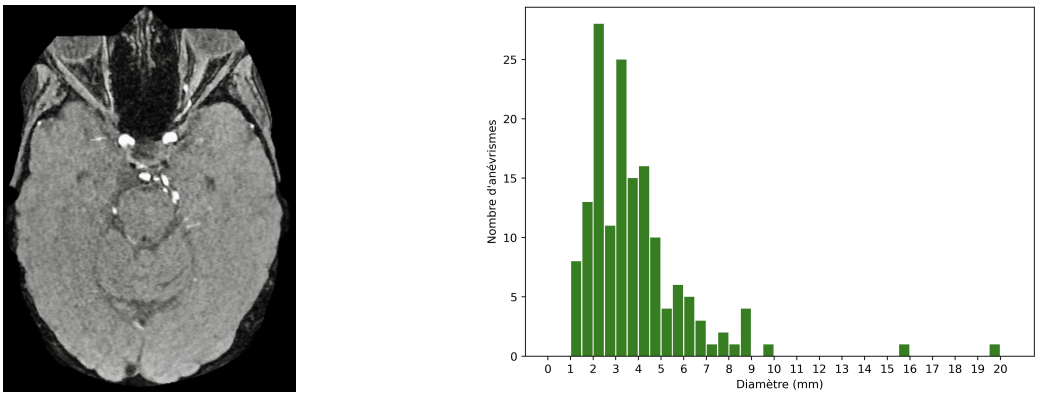
4.2.1 Collection et annotation de données

Collection de données

Un ensemble de 111 examens TOF-MRA a été collecté rétrospectivement au sein du service de neuroradiologie interventionnelle du CHRU de Nancy, couvrant la période d'avril 2015 à janvier 2020. Ces examens comprenaient des données provenant de 56 femmes et 55 hommes. Les critères d'inclusion exigeaient la présence d'anévrismes de type sacculaire, tandis que les critères d'exclusion englobaient les anévrismes déjà traités et ceux de grande taille (> 20 mm), en raison de leur rareté et de leur facilité de détection par les experts. Les examens d'imagerie ont été réalisés à l'aide d'une machine IRM de GE Healthcare (Discovery MR750w) fonctionnant à une intensité de champ magnétique de 3 Tesla (3T). Le protocole employé correspond à celui utilisé de manière standard en pratique clinique.

Les paramètres d'acquisition incluaient un temps de répétition (TR) de 28 ms, un temps d'écho (TE) de 3,4 ms, une épaisseur de coupe de 0,8 mm, un champ de vision (FOV) de 24, et un angle d'inclinaison de 17 degrés, organisés en 4 blocs de 54 coupes chacun. La durée totale était d'environ 6 minutes et 28 secondes. Ces réglages ont facilité la génération de volumes d'images de haute résolution avec des dimensions de $512 \times 512 \times 254$ voxels et une taille de voxel de $0,47 \times 0,47 \times 0,4$ mm^3. Chaque fichier DICOM a été anonymisé selon les standards de protection de la vie privée et converti en format Nifty sur le site clinique avant tout traitement. Cette procédure est conforme à la politique du CHRU de Nancy qui autorise l'utilisation anonyme des données d'imagerie pour la recherche, en respectant le principe de non-opposition du patient.

Les images TOF-MRA ont été prétraitées pour améliorer leur qualité. Une méthode de suppression du crâne, basée sur l'approche de Brett [42], a été utilisée pour éliminer les structures crâniennes et adjacentes. Cependant, cette méthode avait tendance à éliminer des informations vasculaires, en particulier autour du siphon carotidien. Par conséquent, nous avons mis en œuvre une méthode simple mais approximative : l'image de l'amplitude du gradient a été seuillée à son 80e centile et un masque a été créé entre le premier et le dernier voxel non nul de chaque ligne dans le volume. Ce masque a ensuite été érodé 20 fois à l'aide d'une boule 3D d'un rayon de 2 voxels (voir Figure 4.2a). Nous avons également évalué l'efficacité de la correction de biais N4, une technique couramment utilisée pour corriger les variations d'intensité dues aux artefacts de biais dans les images médicales [289]. Cependant, elle n'a pas apporté d'améliorations significatives aux images de notre ensemble de données, qui étaient déjà de haute qualité. Par conséquent, nous l'avons exclue de notre étude.



(a) (b)

FIGURE 4.2 – Caractéristiques de la base de données du CHRU de Nancy : (a) exemple d'image après la suppression du crâne, (b) distribution des tailles des anévrismes.

Annotation de données

Dans le cadre de la pratique clinique, un total de 155 anévrismes a été diagnostiqué à partir des 111 images de la base de données. La répartition de ces anévrismes par examen varie de 1 anévrisme par examen dans 81 cas à un maximum de 5 anévrismes dans un seul examen. Le diamètre moyen de ces anévrismes était de 3,86 mm, avec un écart-type de 2,39 mm. Le diamètre médian s'est établi à 3,36 mm, avec des valeurs minimales de 1,23 mm et des valeurs maximales de 19,63 mm. La majorité des anévrismes étaient de petite taille, avec 60 d'entre eux mesurant moins de 3 mm, tandis que 66 se situaient dans la plage de 3 à 5 mm (voir Figure 4.2b). Cette diversité en termes de tailles a enrichi la base de données, mais également pose des défis aux méthodes de détection automatique, qui présentent des limitations dans la détection de cette catégorie spécifique d'anévrismes [81].

Les travaux antérieurs en détection et segmentation des anévrismes ont généralement utilisé une annotation voxel par voxel des anévrismes. Cette approche s'est révélée chronophage, exigeante en termes d'expertise, et sujette à une variabilité dans l'annotation, bien que cette dernière soit rarement mesurée dans ces travaux. La variabilité se manifeste principalement dans la délimitation des contours des anévrismes. Cependant, étant donné que notre objectif principal était la détection des anévrismes plutôt que leur segmentation précise, nous avons opté pour un processus d'annotation moins précis, mais considérablement plus rapide et plus simple. Dans notre approche, chaque anévrisme a été approximé par une sphère définie par deux points stratégiques : l'un positionné au centre du collet de l'anévrisme et l'autre au sommet de son dôme. Cette procédure d'annotation a été réalisée par un radiologue ayant 10 ans d'expérience. Nous avons développé un plugin pour le logiciel 3D Slicer [92] afin de positionner ces points dans une vue de rendu en volume en 3D, avec des ajustements effectués dans les plans de coupes canoniques (voir Figure 4.3). Cette approche a réussi à équilibrer la précision de l'annotation avec la faisabilité pratique, ce qui facilite la collecte de données supplémentaires.

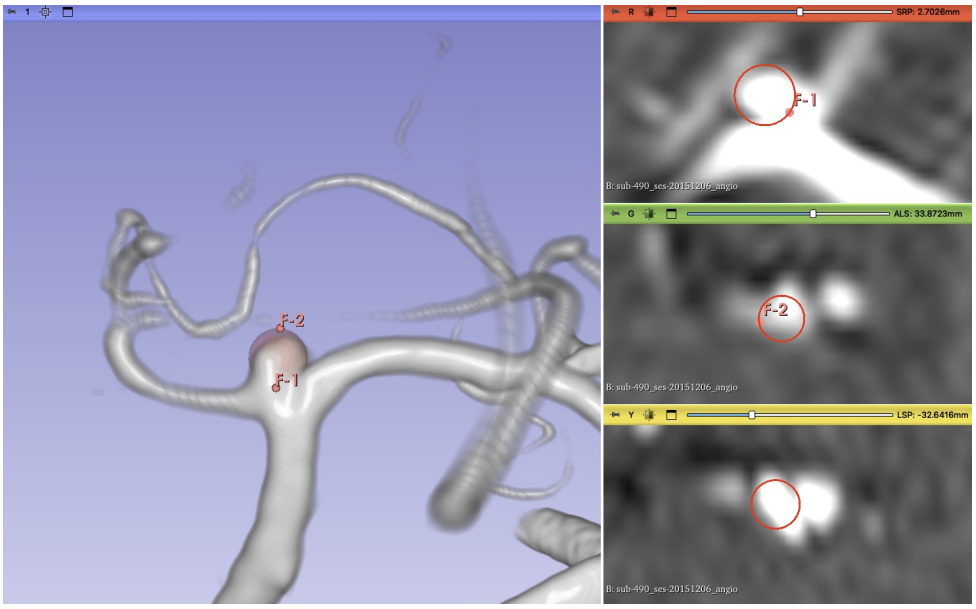


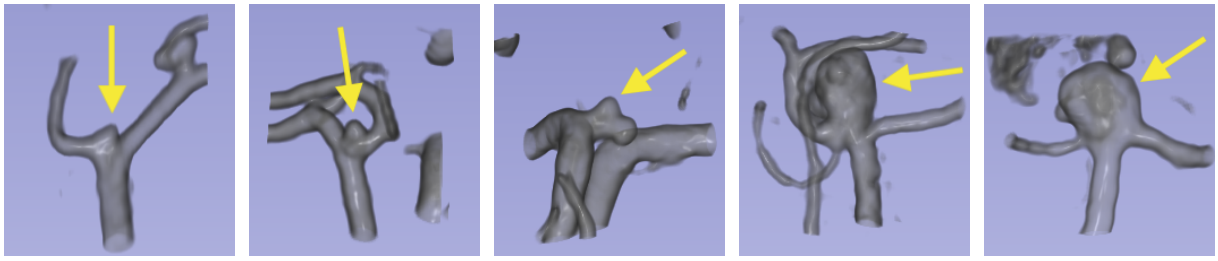
FIGURE 4.3 – Plugin 3D Slicer développé pour l'annotation des anévrismes : Chaque anévrisme est approximé par une sphère définie par deux points (F1, F2).

4.2.2 Approche par petits patchs

La performance d'un classificateur dépend de sa capacité à modéliser à la fois les anévrismes (objets d'intérêt) et leur environnement contextuel (zones sans anévrismes). Le choix de la dimension et de la sélection des patchs utilisés pour l'entraînement jouent un rôle important dans ce processus [281]. Utiliser des patchs de grande dimension, bien que cela permette d'intégrer plus d'informations contextuelles, présente des problèmes. L'un de ces problèmes est qu'il y a plus de chances d'inclure des anévrismes

dans la plupart des patchs, ce qui peut biaiser l'apprentissage du modèle en le rendant moins capable de distinguer correctement les régions sans anévrismes. Pour maintenir un équilibre entre les patchs contenant des anévrismes et ceux sans anévrismes, il est important d'inclure également des images de patients en bonne santé, à partir desquelles on peut extraire des patchs sans anévrismes. En outre, l'utilisation de grands patchs nécessite une puissance de calcul significative pour l'entraînement des CNNs.

Pour surmonter ces limitations, notre approche repose sur l'utilisation de petits patchs non intersectants mesurant $48 \times 48 \times 48$ voxels, avec une résolution voxel de 0,4 mm, qui est proche de la résolution nominale. Cela signifie que chaque patch a une taille spatiale d'environ $19.2 \times 19.2 \times 19.2$ mm^3. Cette méthode s'adapte de manière efficace aux anévrismes, dont les dimensions varient généralement de 1 à 20 mm, et ces petits patchs permettent une représentation adéquate de ces anévrismes (voir Figure 4.4). De plus, en utilisant ces petits patchs, nous obtenons un plus grand nombre d'échantillons à partir des données des patients, ce qui enrichit considérablement la variabilité des échantillons d'entraînement, un aspect essentiel dans le domaine de l'apprentissage profond [340]. Un avantage supplémentaire réside dans l'utilisation de patchs non superposés, qui sont alors considérés comme indépendants les uns des autres. Chaque patch représente des informations distinctes des données, ce qui élimine la redondance et renforce la diversité des échantillons.



(a) 1,97 mm (b) 2,50 mm (c) 4,88 mm (d) 7,52 mm (e) 15,58 mm

FIGURE 4.4 – Anévrismes intracrâniens de différentes tailles au sein de patchs de $48 \times 48 \times 48$ voxels.

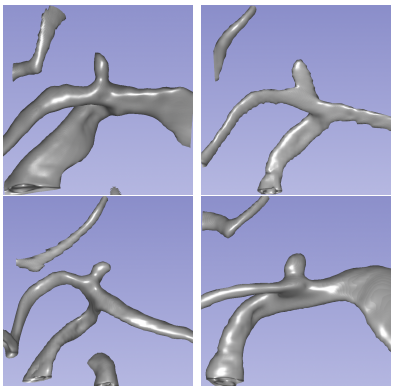
4.2.3 Sélection et synthèse de données

L'utilisation de petits patchs permet d'extraire plusieurs échantillons à partir de chaque image TOF-MRA. Cependant, en raison de la rareté des anévrismes dans les images, il se produit un déséquilibre important entre les classes. Les patchs positifs, qui sont centrés sur un anévrisme, sont en nombre limité, tandis que de nombreux patchs négatifs peuvent être extraits. Pour résoudre ce déséquilibre, des stratégies d'échantillonnage sont couramment utilisées [148]. Notre première approche consiste à utiliser des techniques de synthèse d'images. Les patchs positifs, centrés sur les anévrismes, sont dupliqués 50 fois et soumis à des déformations aléatoires pour créer différentes formes d'anévrismes. Pour réaliser cette synthèse, un treillis de points de contrôle de taille $3 \times 3 \times 3$ est placé autour de chaque patch positif. Chaque point de ce treillis, à l'exception du point central, subit un déplacement aléatoire de 3 mm dans les trois dimensions de l'espace. Le champ de distorsion 3D est généré par interpolation spline cubique en utilisant la fonction `scipy.ndimage.map_coordinates`. Des exemples de synthèses obtenues à partir d'un anévrisme de taille 2,92 mm sont présentés dans la Figure 4.5a. À noter que les points d'annotation subissent la même déformation que le patch.

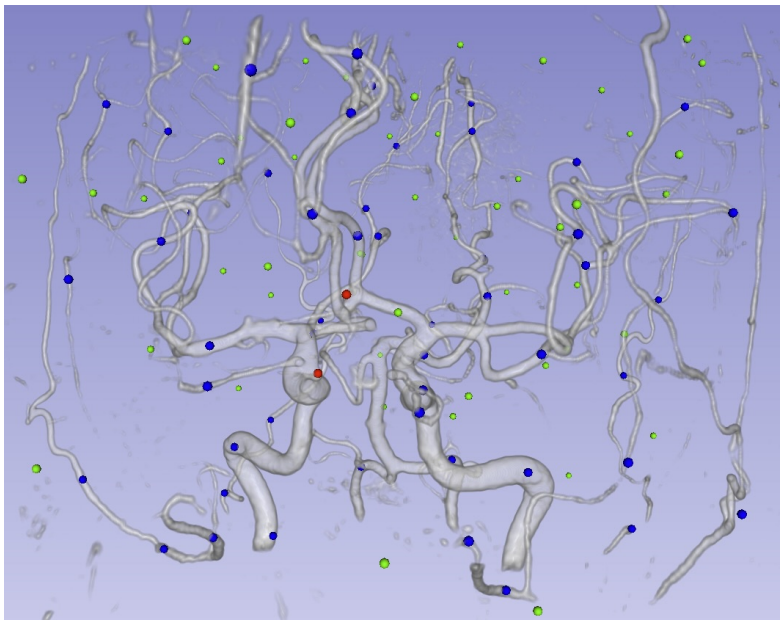
Un autre déséquilibre des classes se manifeste entre les vaisseaux sanguins et les autres tissus cérébraux, où les informations vasculaires ne représentent qu'une proportion limitée de l'image, approximativement de 3 à 5%. Pour améliorer la capacité du modèle à discriminer entre les vaisseaux sains et ceux présentant des anévrismes, une seconde stratégie est mise en œuvre. Cette approche implique la sélection de la moitié des patchs négatifs centrés sur les vaisseaux sanguins et l'autre moitié dans le parenchyme cérébral. Pour ce faire, les 100 voxels les plus lumineux sont choisis comme centres des patchs centrés sur les vaisseaux en respectant la non-intersection des patchs, tandis que les centres des patchs négatifs sont sélectionnés de manière aléatoire parmi les voxels dont les valeurs se situent entre le 20e et le 80e percentile de l'histogramme de l'image. Une distance minimale de 20 mm est maintenue entre les centres des patchs

positifs et négatifs. En conséquence, un total de 200 patchs négatifs est extrait de chaque image TOF-MRA. La répartition de ces centres de patchs dans une image de patient est illustrée dans la Figure 4.5b.

Pendant la phase d'entraînement, des techniques classiques d'augmentation de données sont appliquées à l'ensemble des patchs (positifs et négatifs), notamment des rotations aléatoires allant de 0 à 180 degrés et des translations aléatoires allant de 0 à 10 mm dans les trois dimensions des images. Les patchs positifs sont ré-extrait pour chaque instance de transformation, qu'il s'agisse de translation, de rotation ou éventuellement de déformation.



(a) Synthèse de patchs positifs.



(b) Sélection de patchs.

FIGURE 4.5 – (a) Exemple d'anévrismes synthétisés par distorsion aléatoire à partir d'un anévrisme de taille 2,92 mm. (b) Localisations des centres de patchs positifs (en rouge), des patchs négatifs (en bleu pour les vaisseaux sanguins et en vert pour le parenchyme).

4.3 Réseau de neurones pour la détection des anévrismes

4.3.1 Architecture du réseau de segmentation UNet

Pour évaluer l'efficacité de notre approche de gestion des données de manière indépendante de toute autre technique, nous avons utilisé le réseau de segmentation 3D UNet [65] sans chercher à optimiser excessivement les hyperparamètres. Ce réseau prend des patchs de dimension $48 \times 48 \times 48$ voxels en entrée et les traite à travers plusieurs couches pour générer une segmentation de sortie de la même dimension. Comme illustré dans la Figure 4.6, le réseau UNet se compose de deux parties principales : un chemin de contraction (encodeur) et un chemin d'expansion (décodeur).

Le chemin de contraction extrait et représente les caractéristiques de l'image d'entrée en cartes de caractéristiques sémantiques. Cette représentation s'opère via plusieurs couches de convolution utilisant des filtres de taille $3 \times 3 \times 3$, suivies de normalisations par lots (BN) et d'activations non linéaires ReLU. À noter que le terme de biais a été omis des couches de convolution avant l'application des couches de normalisations par lots, car ces dernières incluent déjà l'ajout de ce terme [133]. Pour réduire la dimension des cartes de caractéristiques tout en préservant les informations pertinentes, des couches de max-pooling sont employées. Le chemin d'expansion, de son côté, vise à restaurer la résolution spatiale en sur-échantillonnant les caractéristiques extraites à travers des couches de déconvolution, tout en intégrant des informations spatiales par le biais de connexions skip. Cette approche permet la fusion des caractéristiques à différentes échelles spatiales. La segmentation finale est obtenue à l'aide d'une couche

de convolution $1 \times 1 \times 1$ suivie d'une activation sigmoïde. Chaque voxel reçoit un score compris entre 0 et 1, indiquant sa probabilité d'appartenir à un anévrisme.

Les couches profondes du réseau UNet sont efficaces pour extraire des caractéristiques de haut niveau, mais elles posent des défis lorsque ils sont appliquées à la segmentation de petits objets. Nous avons limité la profondeur à 4 niveaux (après 3 couches de sous-échantillonnage) en raison de la petite taille des anévrismes. Par exemple, un anévrisme de 4 mm, au départ représenté par une zone d'environ 10 voxels de diamètre, perd progressivement son signal lors des étapes de sous-échantillonnage, atteignant environ 1,25 voxel au 4 niveau. Une augmentation de la profondeur pourrait entraîner une perte d'information significative, même pour les anévrismes de plus taille. Par ailleurs, les connexions skip permet de réduire l'impact de cette perte en fusionnant les données des couches profondes, qui capturent le contexte global, avec celles des couches initiales, facilitant ainsi la restauration de détails des anévrismes.

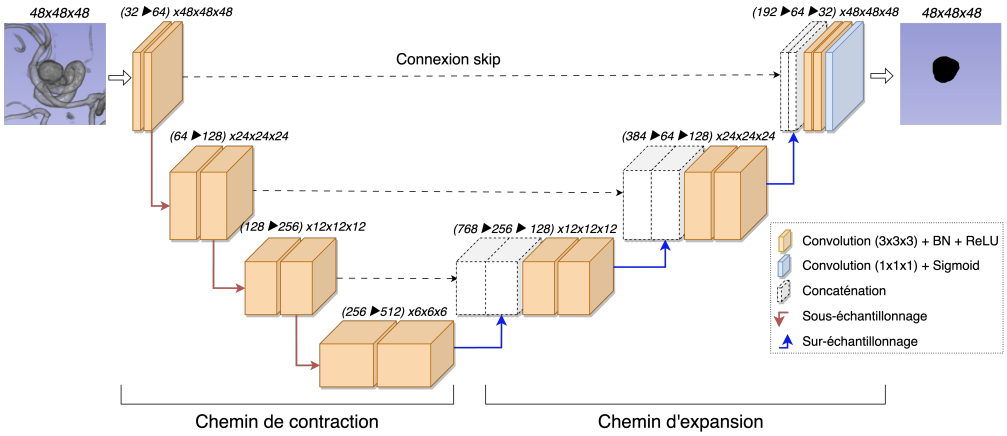


FIGURE 4.6 – Architecture du réseau 3D UNet [65] employé pour la segmentation d'images de dimensions $48 \times 48 \times 48$ voxels. Les dimensions des images et des cartes de caractéristiques sont illustrées.

4.3.2 Détails d'implémentation

Notre méthode a été implémentée en utilisant le langage de programmation Python (version 3.8.5) conjointement avec le framework Keras (version 2.4.3). Le réseau 3D UNet que nous avons employé compte environ 19 millions de paramètres, nécessitant une mémoire de 76,31 Mo. Pour optimiser ces paramètres, nous avons sélectionné les hyper-paramètres suivants : un total de 100 époques d'entraînement, avec un taux d'apprentissage constant fixé à 10^{-4}, utilisant la fonction de coût BCE en combinaison avec l'optimiseur Adam (β_1=0.9, β_2=0.999) [160]. La taille du lot d'entraînement a été fixée à 10, en tenant compte des contraintes de mémoire GPU. Le processus d'entraînement a été effectué sur la plateforme de calcul Grid5000 [21], en utilisant une carte graphique NVIDIA RTX 1080Ti avec 11 Go de mémoire. Chaque itération d'entraînement du modèle a nécessité une allocation de mémoire GPU de 7,8 Go. Chaque époque d'entraînement a duré environ 1h30min, ce qui a entraîné une durée totale d'entraînement d'environ 6 jours pour chaque modèle de notre méthode.

Pendant la phase d'inférence sur l'ensemble de test, nous avons utilisé une technique de reconstruction basée sur des patchs pour générer des prédictions sur l'ensemble du volume. Tout d'abord, le volume a été ré-échantillonné pour obtenir une résolution voxel isotrope de 0,4 mm. Ensuite, des patchs de $48 \times 48 \times 48$ voxels ont été extraits, couvrant la totalité du volume. Afin de réduire les artefacts causés par les opérations de convolution aux bords des petits patchs, nous avons utilisé un chevauchement de 8 voxels entre les patchs voisins. Seule la partie centrale de chaque patch, soit $32 \times 32 \times 32$ voxels, a été conservée pour reconstituer le volume final, formant ainsi une partition du volume IRM. Après avoir obtenu les prédictions pour chaque patch, nous les avons assemblés et le volume reconstruit a été ré-échantillonné à sa résolution d'origine. Le temps nécessaire pour prédire chaque image TOF-MRA était d'environ 30 secondes. Le code d'entraînement et d'inférence est disponible publiquement sur l'adresse suivante : `https://gitlab.inria.fr/yassis/DeepAneSeg`.

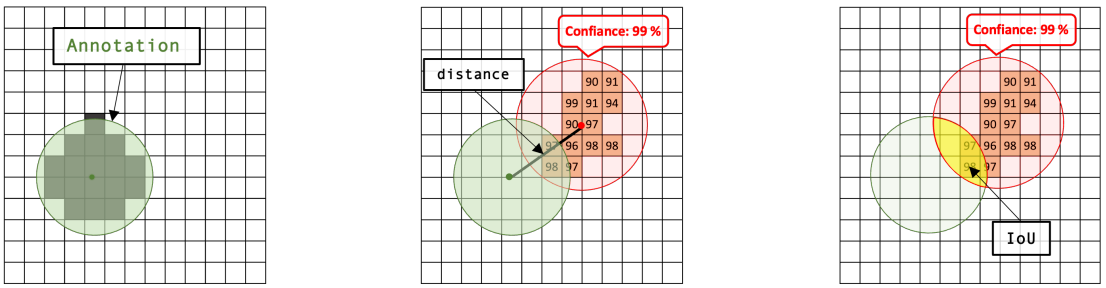
4.3.3 Métriques et méthodologies d'évaluation

Les performances des modèles pendant la phase d'entraînement ont été évaluées en utilisant des métriques de segmentation sur un ensemble de validation. Pour évaluer la concordance entre les valeurs de voxels prédites et la vérité terrain, le coefficient de Dice a été employé. Cependant, en raison de la rareté et de la petite taille des anévrismes, le coefficient de Dice peut se montrer sensible aux petites erreurs de segmentation [309, 192]. Face à cette sensibilité, le coefficient Kappa de Cohen (κ) a été intégré, offrant une évaluation plus robuste en présence de déséquilibre des classes [67].

Les performances sur la tâche de détection dans un ensemble de test ont été évaluées à l'aide de trois principales métriques : la précision moyenne (AP), la sensibilité et les faux positifs par cas (FP/cas). L'AP évalue les performances à différents seuils de classification, tandis que la sensibilité et les FP/cas ont été calculées avec un seuil de classification par défaut de 50%. Pour le calcul de ces métriques, une opération de seuillage à 1% a été appliquée aux images prédites pour éliminer le bruit de segmentation. Les composantes connexes (CC) ont ensuite été extraites et représentées par des sphères avec leur centre au centre de gravité de chaque CC et leur diamètre équivalent à la distance maximale entre les voxels de la CC. Le score de confiance de chaque CC a été déterminé en prenant la valeur maximale parmi ses voxels. Deux approches distinctes ont été utilisées pour calculer les métriques d'évaluation (voir Figure 4.7).

Pour se comparer avec la compétition ADAM, nous avons utilisé des métriques basées sur la distance : un vrai positif (TP) était défini lorsque le centre d'une CC se trouvait dans la vérité terrain, tandis qu'un faux négatif (FN) correspondait à l'absence du centre d'une CC dans la vérité terrain, et un faux positif (FP) se produisait lorsque le centre d'une CC ne correspondait à aucune vérité terrain. Pour un évaluation plus précise, nous avons utilisé des mesures basées sur l'intersection (IoU) entre les sphères de prédictions et de vérité terrain. Contrairement à l'évaluation adopté dans la compétition CADA [137], qui utilisait un seuil IoU de 0%, nous avons choisi un seuil de 10% pour une évaluation plus rigoureuse [26, 192]. Ainsi, une sphère de vérité terrain était considérée comme TP si son score IoU par rapport à une prédiction était de 10% ou plus, sinon elle était classée comme FN. Un FP était attribué à une prédiction sans correspondance avec une vérité terrain (IoU $< 10\%$). En cas de superposition de plusieurs prédictions sur la même vérité terrain, seule la détection avec le score de confiance le plus élevé était conservée, les autres étant considérées comme des FP [221]. Cette approche, sans filtrage basé sur la taille et le volume des CC, conduit à un nombre plus élevé de FP par rapport aux méthodes de la compétition ADAM. De plus, l'utilisation d'un critère IoU de 10% peut entraîner un nombre plus élevé de FP que l'utilisation des métriques ADAM, qui sont basées sur un critère de distance.

Les métriques de détection d'objets ne permettent pas de mesurer les vrais négatifs (TN) et, par conséquent, ne permettent pas le calcul de la spécificité. Pour remédier à cette limitation, une analyse basée sur les patchs a été réalisée. Dans cette analyse, un patch était qualifié de positif s'il contenait au moins un voxel positif, sinon il était considéré comme négatif. Cette approche a permis de construire une matrice de confusion complète pour évaluer la spécificité des modèles.



(a) Annotation. (b) Métriques ADAM. (c) Métriques basées sur IoU.

FIGURE 4.7 – Comparaison des métriques d'évaluation entre la vérité terrain (en vert) et les prédictions (en rouge) définies par des sphères : utilisation (b) des métriques d'ADAM basées sur la distance entre les centres des sphères et (c) des métriques basées sur l'intersection IoU entre les sphères.

4.4 Étude expérimentale et discussion

Cette section est dédiée à l'étude expérimentale visant à évaluer l'efficacité de notre approche pour la tâche de détection des anévrismes. L'évaluation comprend deux aspects principaux : une étude par ablation pour analyser l'impact et la contribution des différentes composantes de notre stratégie de données, ainsi qu'une validation croisée à 5 plis où nous évaluerons la capacité de généralisation de notre approche sur un ensemble de données plus étendu, tout en comparant ses performances par rapport aux méthodes utilisées dans la compétition ADAM.

4.4.1 Étude par ablation

Nous avons mis en place notre méthode en entraînant le réseau 3D UNet [9] avec la fonction de coût BCE, conformément à la stratégie d'échantillonnage et de génération de patchs décrite dans la section 4.2. Cette stratégie consiste à utiliser des petits patchs ($48 \times 48 \times 48$ voxels, avec une taille de voxel de 0,4 mm^3), accompagnés de 50 duplications et de la synthèse de patchs positifs, ainsi que la sélection guidée de 200 patchs négatifs par image TOF-MRA. Dans cette étude d'ablation, nous comparons notre approche, appelée *Modèle0*, avec 8 variantes distinctes :

- *Modèle1* : Étant donné la sensibilité au déséquilibre des classes de la fonction de coût BCE [279], ce modèle est entraîné en utilisant la fonction de coût Focal [176], conçue pour concentrer l'apprentissage sur les classes minoritaires.
- *Modèle2* : Pour évaluer l'impact de notre stratégie de duplication des patchs positifs visant à rééquilibrer les cas positifs, ce modèle réduit le nombre de duplications des patchs positifs à 5.
- *Modèle3* : Afin de déterminer si 50 duplications de chaque patch positif suffisent pour capturer les caractéristiques des anévrismes, ce modèle est entraîné en utilisant 75 duplications.
- *Modèle4* : Afin d'évaluer l'impact de notre stratégie de synthèse de diverses formes d'anévrismes sur les performances, ce modèle maintient 50 duplications pour les patchs positifs, tout en éliminant la synthèse par distorsions.
- *Modèle5* : Examine la proportion entre les patchs positifs et négatifs utilisés dans l'entraînement. Il choisit 100 patchs négatifs par patient, en les répartissant également entre les vaisseaux sanguins et le parenchyme, par contraste avec les 200 patchs utilisés dans *Modèle0*.
- *Modèle6* : Explore une stratégie différente pour la sélection des patchs négatifs en utilisant une approche de sélection aléatoire au lieu de se concentrer spécifiquement sur les vaisseaux sanguins, comme dans notre approche.
- *Modèle7* : Évalue l'influence de l'information contextuelle sur la détection des anévrismes en utilisant des patchs de dimensions augmentées à $96 \times 96 \times 96$ voxels.
- *Modèle8* : Examine l'effet de la résolution spatiale sur la détection des anévrismes en utilisant des patchs de $48 \times 48 \times 48$ voxels avec une taille de voxel de 0,5 mm^3.

En raison de la longue durée nécessaire pour l'entraînement de chaque modèle, estimée à environ 6 jours, l'utilisation de la validation croisée pour une évaluation plus robuste s'est avérée peu pratique. Par conséquent, notre ensemble de données a été divisé en trois sous-ensembles distincts pour cette étude. Le premier est l'ensemble d'entraînement, qui contenait 78 images avec 109 anévrismes, ayant une taille moyenne de 3,95 $\pm$ 2,54 mm. Le deuxième est l'ensemble de validation, composé de 22 images avec 31 anévrismes, dont la taille moyenne était de 3,64 $\pm$ 1,77 mm. Enfin, le troisième est l'ensemble de test, regroupant 11 images contenant un total de 15 anévrismes, avec une taille moyenne de 3,67 $\pm$ 2,28 mm. Tous les modèles ont été formés sur l'ensemble d'entraînement, avec *Modèle0* ayant été formé sur un total de 21 050 patchs, dont 5 450 étaient des patchs positifs. Les performances ont été évaluées régulièrement sur l'ensemble de validation à la fin de chaque époque en utilisant les coefficients Dice et Kappa de Cohen (κ). Après 100 époques d'entraînement, une évaluation finale a été réalisée sur l'ensemble de test en utilisant les métriques de détection d'objets. Les résultats sont résumés dans la table 4.2.

Modèles	*Ensemble de validation*		*Ensemble de test*											
	Segmentation		*Métriques ADAM*			*Métriques IoU*			*Métriques par patch*					
	Dice	κ	**AP**	**Sensibilité**	**FP/cas**	**AP**	**Sensibilité**	**FP/cas**	**TP**	**FP**	**FN**	**TN**	**Sensibilité**	**Spécificité**
Modèle0	33,91	0,66	80,17	97,0	0,45	78,59	80,0	0,72	14	5	2	2194	87,5	99,77
Modèle1	08,95	0,52	76,81	80,3	0,19	75,49	40,0	0,60	12	2	4	2197	75,0	99,91
Modèle2	03,86	-1,2e-8	03,42	0	0	01,70	0	0	0	0	16	2199	0	100
Modèle3	35,14	0,64	81,23	97,0	0,43	78,45	80,0	0,70	14	4	2	2195	87,5	99,81
Modèle4	43,41	0,77	75,51	87,9	1,54	74,53	73,0	1,74	14	11	2	2188	87,5	99,5
Modèle5	22,57	0,58	77,46	83,3	1,0	76,85	66,0	1,31	13	8	3	2191	81,25	99,64
Modèle6	32,95	0,55	76,71	91,2	0,7	75,11	66,0	0,92	13	10	3	2189	81,25	99,55
Modèle7	31,82	0,63	79,01	93,7	0,24	76,53	73,0	0,64	13	3	3	2196	81,25	99,86
Modèle8	20,70	0,51	67,19	78,1	0,14	68,10	53,0	0,57	11	2	5	2197	68,75	99,91

TABLE 4.2 – Comparaison des variantes du modèle : *Modèle0* représente notre méthode proposée. Les métriques Dice et κ ont été évaluées sur l'ensemble de validation, tandis que les métriques d'ADAM, d'IoU, ainsi que par patchs ont été calculées sur l'ensemble de test. La sensibilité, le nombre de faux positifs (FP/cas), et la spécificité sont calculées à un seuil de classification par défaut de 50%.

Notre stratégie de gestion des données a produit les résultats les plus prometteurs. L'incorporation de la fonction de coût Focal [176] dans le *Modèle1* n'a pas induit d'améliorations significatives. En revanche, l'utilisation de la fonction de coût BCE a favorisé une augmentation de la sensibilité, mettant en évidence l'efficacité de notre méthode d'échantillonnage et de génération de données pour contrer le déséquilibre des classes. Bien que le coefficient Dice demeure relativement faible pour le *Modèle1*, en raison des dimensions restreintes des composantes connexes prédites, le coefficient κ a révélé des performances relativement bonnes pour ce modèle. Le *Modèle2* n'a pas convergé, soulignant ainsi la persistance du défi lié au déséquilibre des classes. L'ajout de 75 duplications des patchs positifs dans le *Modèle3* n'a pas entraîné une amélioration significative des performances. Cela suggère que 50 duplications sont adéquates pour représenter la diversité des anévrismes. Quant au *Modèle4*, il a affiché de bonnes performances, mais a été pénalisé par un nombre significatif de faux positifs, attribuables à une insuffisance de diversité dans les formes d'anévrisme durant la phase d'entraînement. L'augmentation de la taille des composantes connexes explique les scores Dice et κ plus élevés. Cependant, l'équilibre entre les classes s'est amélioré avec le *Modèle5*, bien que ses performances n'aient pas atteint le niveau du *Modèle0*. Cette situation a mis en évidence la limitation en patchs négatifs, soulignant leur importance pour une modélisation statistiquement robuste du contexte.

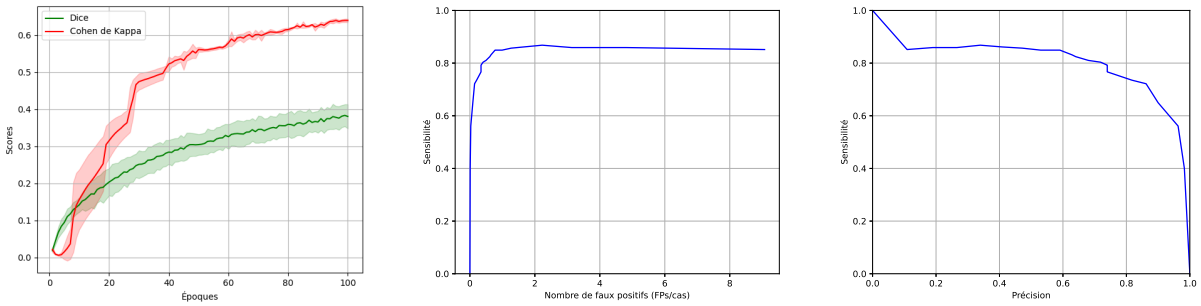
L'utilisation de la sélection aléatoire de patchs négatifs pendant l'entraînement du *Modèle6* a eu un impact négatif sur ses performances, en particulier en ce qui concerne les faux positifs, dont certains étaient localisés en dehors des vaisseaux sanguins. Cette observation suggère que la sélection ciblée de patchs négatifs centrés sur les vaisseaux améliore la capacité du modèle à différencier les vaisseaux sains des vaisseaux présentant des anévrismes. Afin d'évaluer l'influence de l'information contextuelle et la résolution spatiale sur les performances de détection, nous avons également examiné les résultats des *Modèle7* et *Modèle8*. Bien que *Modèle7* a été soumis à un entraînement prolongé de 11 jours et à une utilisation plus substantielle des ressources de calcul pour traiter des patchs de dimension $96 \times 96 \times 96$ voxels (soit $38{,}4^3$ mm^3), il n'a pas montré d'améliorations significatives, à l'exception d'une réduction des faux positifs. Ces résultats démontre que notre stratégie de données a réussi à capturer une diversité d'informations anatomiques, même avec des patchs de taille réduite. En outre, le *Modèle8* souligne l'importance de l'utilisation de patchs de haute résolution pour détecter le signal rare des anévrismes. Bien que les patchs de $48 \times 48 \times 48$ voxels avec une faible résolution spatiale (24^3 mm^3 par patch) ont permis de réduire les faux positifs en intégrant davantage d'informations contextuelles, ils ont présenté une sensibilité de détection plus faible.

La différence entre les résultats obtenus par les métriques d'ADAM et celles basées sur l'IoU provient principalement de la méthode d'évaluation des détections. Les métriques ADAM évaluent la distance entre les centres des composantes connexes prédites et ceux de la vérité terrain. En revanche, les métriques basées sur l'IoU, avec un seuil de 10%, mesurent l'intersection entre les CC prédites et la vérité terrain, représentées sous forme de sphères. Par conséquent, de multiples petites détections localisées au niveau de la sphère de vérité terrain entraînent des vrais positifs lors de l'utilisation des métriques ADAM, tandis que les métriques basées sur l'IoU peuvent générer des faux positifs, et uniquement un seul vrai positif si IoU est supérieur ou égal à 10%. Cette situation se traduit par une augmentation du nombre de faux positifs, une réduction simultanée du score de sensibilité et de la précision moyenne lors de l'utilisation des métriques basées sur l'IoU. À noter notre évaluation s'est limitée à un ensemble de données de test composé de 11 images avec 15 anévrismes, ce qui limite la généralisation des résultats. Pour une évaluation plus significative, prenant en compte la variabilité des scores sur différentes instances de données, nous procéderons à une étude de validation croisée dans laquelle nous comparerons nos résultats avec ceux des méthodes participantes à la compétition ADAM.

4.4.2 Évaluation par validation croisée

Afin d'élargir notre évaluation et d'obtenir des mesures de variabilité de performance, nous avons réalisé une validation croisée à 5 plis sur un ensemble de données comprenant 111 images. Cette procédure a divisé les données en 5 partitions distinctes, chacune contenant 22 ou 23 images, présentant des caractéristiques variables en termes de nombre et de diamètre moyen des anévrismes. Plus précisément, les caractéristiques étaient les suivantes : 30 anévrismes avec une taille moyenne de 4,01 $\pm$ 1,87 mm, 33 anévrismes avec une taille moyenne de 4,31 $\pm$ 3,83 mm, 29 anévrismes avec une taille moyenne

de 3,41 ± 1,50 mm, 34 anévrismes avec une taille moyenne de 3,94 ± 1,84 mm, et 29 anévrismes avec une taille moyenne de 3,54 ± 1,83 mm. Les modèles ont été entraînés en utilisant 4 des partitions, tandis que la cinquième servait d'ensemble de test. Cette approche a généré des prédictions pour chaque patient dans chaque ensemble de test, fournissant ainsi des prédictions pour l'ensemble des patients. La variabilité des performances, sous forme d'écart-type, a ensuite été calculée sur les 5 différents ensembles de test. En moyenne, chaque modèle a utilisé environ 19 000 patchs lors du processus d'entraînement, dont 1 200 étaient des patchs positifs. La Figure 4.8a illustre la variabilité des scores Dice et Kappa de Cohen entre ces modèles au cours de la phase d'entraînement.



(a) (b) (c)

FIGURE 4.8 – (a) Évolution des coefficients Dice et Kappa pour les 5 modèles lors de la validation croisée. (b, c) Courbes FROC et Sensibilité-Précision de notre méthode utilisant les métriques ADAM.

Étant donné que les organisateurs de la compétition ADAM n'acceptent plus de nouvelles soumissions, notre capacité à effectuer une comparaison directe avec les méthodes des participants à cette compétition s'est trouvée limitée. De plus, l'absence d'implémentations publiques ou de descriptions détaillées des méthodes employées dans cette compétition a entravé la possibilité de mener une comparaison directe sur le même ensemble de données. Par conséquent, nous avons adopté deux stratégies de comparaison distinctes. La première stratégie se fonde sur l'évaluation des performances en utilisant les scores de sensibilité et de faux positifs dérivés de la courbe FROC de notre méthode. Bien que cette évaluation soit effectuée sur deux ensembles de données différents, la base de données du CHRU de Nancy présente des similitudes à celle de la compétition, notamment en termes de nombre d'images et de taille des anévrismes. La seconde stratégie vise à comparer notre méthode avec le framework auto-configurable nnUNet [135], qui a été utilisé par les équipes junma [186] et uw2 [197]. Ce framework s'appuie sur une variante du réseau 3D UNet, utilisant des patchs de grande dimension mesurant $192 \times 128 \times 96$ voxels, une approche d'apprentissage ensembliste et des étapes de post-traitement. Pour cette comparaison, les composantes connexes ont été extraites des images prédites et des détections sous forme de sphères ont été générées suivant notre méthodologie (voir section 4.3.3). Table 4.3 synthétise les résultats de ces expérimentations, incluant à la fois les métriques ADAM et celles basées sur l'indice IoU.

Selon les métriques de la compétition ADAM, notre méthode a montré une sensibilité de 82,08 ± 6,58% et un taux de 0,61 ± 0,28 FP/cas. Les variations de sensibilité et du nombre de faux positifs à divers seuils de classification sont présentées dans la courbe FROC (voir Figure 4.8b). Cela nous permet de comparer la sensibilité de notre méthode avec celles des méthodes concurrentes à un taux de FP/cas équivalent. En comparaison avec la méthode abc [326], notre méthode a atteint une sensibilité de 80% avec un taux de 0,40 FP/cas. Par rapport à la méthode xlim [250], notre méthode a montré une sensibilité de 83,71% avec un taux de 4,03 FP/cas. L'aire sous la courbe AUC-FROC a été évaluée à 85,24 ± 7,04%, ce qui reflète sa capacité à maintenir des performances élevées sur une gamme variée de seuils de classification. En choisissant le seuil de classification optimal, situé au point le plus proche du coin supérieur gauche de la courbe FROC, notre méthode a obtenu une sensibilité de 72,10%, avec un taux de 0,14 FP/cas. Pour mieux visualiser la relation entre la sensibilité et la précision à différents seuils de classification, nous avons également créé la courbe Sensibilité-Précision (voir Figure 4.8c). Cette courbe met en évidence la capacité de notre modèle à maintenir un équilibre entre la sensibilité et la précision à différents seuils de classification, obtenant ainsi un score AUC de 80,40 ± 6,71%.

Méthodes	*Métriques ADAM*			*Métriques IoU*		
	AP (%)	**Sensibilité (%)**	**FP/cas**	**AP (%)**	**Sensibilité (%)**	**FP/cas**
abc [326]		68	0,40			
xlim [250]		70	4,03			
unil-chuv [80]		68	2,50			
joker [321]		63	0,16			
nnUNet [135]		63	1,08			
nnUNet [135]		61	0,18			
Kubiac [76]		60	0,36			
Notre méthode	72,49 ± 6,11	82,08 ± 6,58	0,61 ± 0,28	62,10 ± 5,80	69,76 ± 5,41	0,63 ± 0,27
nnUNet [135]	74,50 ± 5,20	83,71 ± 9,27	0,40 ± 0,31	65,81 ± 4,40	72,10 ± 4,13	0,42 ± 0,10

TABLE 4.3 – Comparaison des méthodes de détection basées sur la segmentation : (En haut) Résultats officiels des 7 meilleures méthodes d'ADAM, évaluées sur les données privées d'ADAM. (En bas) Notre méthode est comparée à nnUNet en utilisant l'ensemble de données du CHRU. La variabilité des performances, mesurée par l'écart-type, est calculée sur les différents plis de test.

Concernant les métriques basées sur l'indice IoU entre les prédictions et la vérité terrain exprimée sous forme de sphères, notre méthode a montré une sensibilité de 69,76 ± 5,41% avec un taux de 0,63 ± 0,27 FP/cas en utilisant un seuil de classification par défaut de 50%. En comparaison, le framework nnUNet [135] a montré une sensibilité de 72,10% avec un taux de 0,42 FP/cas. Pour une évaluation globale des performances à travers différentes mesures de classification, nous avons calculé la précision moyenne (AP). Notre méthode a obtenu un score de 62,10 ± 5,80%, tandis que nnUNet a atteint un score de 65,81 ± 4,40%.

4.4.3 Synthèse et analyse des résultats

Dans cette étude, nous avons évalué l'impact de notre stratégie de données sur l'apprentissage d'un réseau UNet standard utilisant la fonction de coût BCE et une optimisation simple. Les hyperparamètres, tels que le taux d'apprentissage et le nombre d'époques, ont été fixés pour mettre en évidence l'influence de notre stratégie de données sur les performances du réseau. Deux aspects importants de notre méthodologie sont à noter. Tout d'abord, nous avons opté pour une annotation simple mais approximative des images, ce qui a permis une annotation rapide d'un grand nombre d'images. De plus, notre approche repose sur l'utilisation de petits patchs d'images non intersectants, considérés comme indépendants les uns des autres. Cette approche s'est avérée efficace même avec un ensemble de données relativement restreint de 111 images TOF-MRA. En moyenne, plus de 21 000 échantillons de patchs ont été utilisés dans le processus d'apprentissage, comme indiqué dans l'étude d'ablation. De plus, le modèle résultant présente une faible empreinte mémoire, le distinguant des méthodes concurrentes telles que nnUNet [135]. Cette dernière implique l'utilisation de l'apprentissage ensembliste et des opérations de post-traitement, ce qui nécessite environ 10 jours pour l'entraînement de chaque modèle, en contraste avec notre méthode qui demande environ 6 jours.

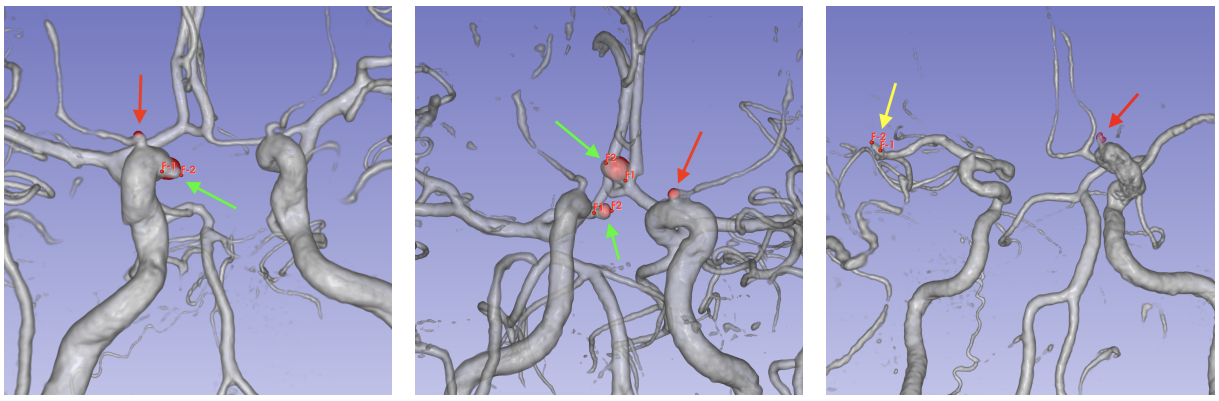
Stratégie de gestion de données

Notre stratégie d'échantillonnage de données se compose de deux phases distinctes. La première phase implique un échantillonnage guidé, où les patchs négatifs (sans anévrismes) sont extraits, avec la moitié centrée sur les vaisseaux sanguins et l'autre moitié en dehors de ces zones. L'utilisation de 200 patchs s'est avérée plus efficace que 100 pour capturer les caractéristiques statistiques du fond (*Modèle0* vs *Modèle5*). De plus, la sélection guidée autour des vaisseaux sanguins a amélioré la capacité à différencier les anévrismes des vaisseaux sains, réduisant ainsi le taux de faux positifs (*Modèle0* vs *Modèle6*). Dans la deuxième phase, nous avons utilisé une synthèse de données qui consiste à dupliquer les patchs positifs pour remédier au déséquilibre et à la rareté des anévrismes dans les images TOF-MRA (*Modèle0* vs *Modèle2*). L'utilisation de 50 duplications de patchs positifs s'est avérée suffisante pour représenter la diversité des anévrismes dans les données (*Modèle0* vs *Modèle3*). De plus, des distorsions aléatoires non rigides ont été introduites pour générer diverses formes d'anévrismes, améliorant ainsi la performance de détection et réduisant les faux positifs (*Modèle0* vs *Modèle4*). L'adoption de patchs de plus grande

taille ($96 \times 96 \times 96$ voxels) a permis d'intégrer davantage d'informations contextuelles et de réduire les faux positifs. Néanmoins, nos résultats indiquent que les petits patchs, en conjonction avec notre méthode de sélection et de synthèse, ont produit des performances compétitives tout en réclamant moins de ressources de calcul (*Modèle0* vs *Modèle7*). Pour améliorer la précision de la détection des anévrismes, qui se caractérisent par leur petite taille et leur signal rare, il est nécessaire d'utiliser des patchs de haute résolution ou en tant proches de la résolution nominale des images (*Modèle0* vs *Modèle8*).

Analyse et visualisation des résultats

Selon les métriques de la compétition ADAM, notre méthode proposée a démontré une sensibilité de 82,08 ± 6,58%, avec un taux de 0,61 ± 0,28 FP/cas. En comparaison avec le framework nnUNet, notre méthode présente des performances légèrement inférieures, bien que la différence ne soit pas statistiquement significative. Pour une évaluation plus précise face aux méthodes d'ADAM, un test sur le même jeu de données est nécessaire. Néanmoins, en se basant sur l'analyse de la courbe FROC à un seuil identique de faux positifs, notre méthode se révèle plus performante. Malgré les performances prometteuses, le nombre de faux positifs reste élevé pour une application en pratique clinique. Les faux positifs les plus difficiles sont situés là où une petite artère, proche de la limite de résolution, se branche sur une grande artère. Ceux-ci sont confondus avec de petits anévrismes (voir Figure 4.9). Notre modèle montre des performances moins bonnes pour la détection de petits anévrismes. Sur les 155 anévrismes, 34 n'ont pas été détectés (FN), dont 18 avaient un diamètre inférieur à 2 mm, et 10 avaient un diamètre inférieur à 3 mm. La sensibilité de notre modèle est de 53% pour les anévrismes inférieurs à 2 mm, mais elle atteint 89% pour les anévrismes de plus grande taille. La détection de petits anévrismes en imagerie par résonance magnétique est une problématique documentée [140].



(a) (b) (c)

FIGURE 4.9 – Visualisation des résultats de notre méthode : les CC prédites en rouge, anévrismes détectés sont indiqués en flèche verte. (a, b) La ramification de petites artères peut être confondue avec un anévrisme (flèches rouges). (c) Petit anévrisme non détecté mesurant 2,74 mm (flèche jaune).

Métriques et méthodologies d'évaluation

Les réseaux de segmentation sont couramment utilisés pour la détection des anévrismes en raison de leur simplicité d'implémentation [139]. Cependant, pour transformer les images prédites en détections, ces réseaux nécessitent des étapes de post-traitement, telles que le filtrage des composantes connexes (CC) en fonction de leur taille, volume et/ou localisation. Contrairement à la plupart des méthodes d'ADAM, notre approche évite ces étapes de post-traitement pour refléter sa performance réelle, même si cela a entraîné une augmentation des faux positifs. L'utilisation de réseaux de segmentation peut également influencer l'évaluation des problèmes de détection d'objets, en particulier pour les petits objets tels que les anévrismes, qui sont constitués de seulement quelques voxels. De légères variations dans les images prédites, dues à la nature non déterministe des algorithmes et/ou au seuillage appliqué, peuvent avoir un impact sur la taille des CC et, par conséquent, sur les métriques d'évaluation. Dans notre approche d'évaluation, nous avons choisi un seuil de classification très bas pour réduire la variabilité liée au seuillage

des images prédites, malgré le fait que cette approche produit plus de faux positifs. De plus, nous avons préféré les métriques d'évaluation basées sur l'IoU avec un seuil de 10%, plutôt que celles basées sur la distance, comme c'est le cas pour ADAM. À noter que ces métriques basées sur l'IoU peuvent également conduire à des faux positifs lorsque l'indice IoU est inférieur à 10%. En outre, nous avons utilisé la métrique de précision moyenne (AP) pour évaluer les performances à différents seuils de classification. Enfin, afin de quantifier la variabilité des performances, nous avons adopté une approche d'évaluation par validation croisée.

4.5 Données de référence et intégration de l'a-priori médical

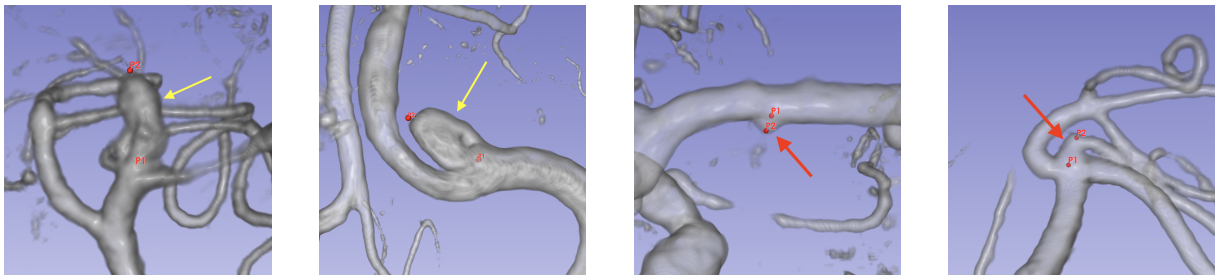
Notre méthode d'échantillonnage et de génération de données a montré des résultats prometteurs dans la détection d'anévrismes. Elle peut s'adapter à des architectures de réseaux plus complexes et à des fonctions de coût sophistiquées. Cependant, notre analyse a révélé des pistes d'amélioration que nous aborderons dans cette section. Lors de l'analyse visuelle des prédictions générées par nos modèles, nous avons identifié plusieurs anévrismes non annotés, ce qui nous a incités à réévaluer la qualité des annotations pour garantir des processus d'entraînement et d'évaluation plus précis. De plus, le temps nécessaire à l'entraînement, d'environ 6 jours par modèle, constitue également un défi. Par conséquent, nous avons recherché des stratégies pour réduire ce temps, ce qui nous facilite l'exploration de réseaux de neurones plus avancés. Dans ce contexte, pour réduire le nombre de faux positifs et améliorer la sensibilité dans la détection d'anévrismes de petite taille, nous avons envisagé l'intégration de connaissances médicales dans le processus d'apprentissage de ces réseaux.

4.5.1 Consolidation des annotations et reproductibilité

Amélioration des annotations et expansion de données

L'évaluation visuelle des prédictions générées par les modèles de notre méthode a révélé des limitations concernant la fiabilité des annotations existantes. En effet, plusieurs anévrismes détectés par les modèles n'avaient pas été initialement annotés, soulignant ainsi la nécessité de mettre à jour ces annotations afin d'accroître la fiabilité des données de référence. Dans cette optique, un radiologue expérimenté avec 10 années d'expérience a entrepris une révision des images de notre base de données en se basant sur les résultats des modèles et les annotations initiales. Cette démarche a permis d'identifier et d'annoter 18 nouveaux cas d'anévrismes. Ces nouveaux anévrismes avaient une taille moyenne de 2,58 $\pm$ 1,65 mm (allant de 0,96 mm à 7,09 mm, avec une taille médiane de 2,22 mm) et étaient répartis sur 15 patients distincts. Deux exemples de ces anévrismes sont illustrés dans les Figures 4.10a et 4.10b.

D'autre part, deux annotations incorrectes ont été supprimées, car elles ne correspondaient pas à de véritables anévrismes. Ces deux erreurs d'annotation concernaient des régions anatomiques distinctes, à savoir l'origine d'une petite artère (voir Figure 4.10c) et une zone courbée des vaisseaux sanguins



(a) (b) (c) (d)

FIGURE 4.10 – Variabilité des annotations de la base de données CHRU de Nancy : (a, b) Exemples d'anévrismes non annotés, mesurant respectivement 7,09 mm et 6,63 mm. (c, d) Deux annotations incorrectes, en réalité ne sont pas des anévrismes.

(voir Figure 4.10d). Après cette phase de correction des annotations, le nombre total d'anévrismes dans notre base de données est passé de 155 à 171 anévrismes. Ces anévrismes avaient une taille moyenne de 4,00 ± 2,45 mm (allant de 0,96 mm à 19,63 mm, avec une taille médiane de 3,55 mm). Ces ajustements soulignent la complexité de la détection des anévrismes et l'importance des systèmes automatiques pour assister les experts dans leur diagnostic. L'expert de notre étude a montré une sensibilité de 89,47% et un taux de faux positifs de 0,018 FP/cas. Par ailleurs, fin 2021, notre base de données a été enrichie de 21 nouvelles images TOF-MRA, portant le total à 132 images. Cette collection inclut maintenant 206 anévrismes, avec une taille moyenne de 3,97 ± 2,32 mm, une médiane de 3,41 mm, et des tailles allant de 0,96 mm à 19,63 mm. Parmi ces anévrismes, 81 mesurent moins de 3mm et 77 sont entre 3 et 5mm. Ces données et annotations sont utilisées dans le reste des expériences présentées dans ce manuscrit, sous le nom de base de données CHRU de Nancy.

Passage de Keras à PyTorch

Les frameworks Keras [64] et PyTorch [224] sont bien connus pour leur convivialité, leur documentation approfondie et leurs communautés actives, ce qui en fait des choix privilégiés pour le développement de modèles de réseaux de neurones [83]. Keras était largement utilisé par la communauté au début de nos travaux, mais à partir de 2021, PyTorch a gagné en popularité. Dans le domaine de l'imagerie médicale, des outils de pointe tels que MONAI [48], nnUNet [135], nnDetection [27], et TorchIO [228] ont été développés sur la base de PyTorch. En considération de cette tendance et pour faciliter les comparaisons futures avec des méthodes de l'état de l'art, nous avons opté pour une transition de Keras vers PyTorch.

Une fois l'impémentation faite en PyTorch, notre première démarche a été de reproduire les résultats obtenus précédemment avec Keras. Toutefois, des difficultés en matière de reproductibilité se sont présentées, notamment une augmentation du nombre de FP (voir table 4.4). Une investigation a révélé que ces fluctuations de performance découlaient des différences entre Keras et PyTorch, en ce qui concerne leurs algorithmes internes et leurs méthodes d'initialisation des couches neuronales. À titre d'exemple, PyTorch utilise par défaut l'initialisation He [117], tandis que Keras emploie l'initialisation Xavier [103]. De plus, les paramètres d'initialisation varient selon les types de couches neuronales. Des divergences sont également observées dans les paramètres par défaut des couches de normalisation par lots, notamment en ce qui concerne le paramètre de momentum, qui est de 0,99 dans Keras et de 0,1 dans PyTorch. Ces différences peuvent influer sur la convergence vers les mêmes minima locaux après entraînement. En outre, PyTorch fonctionne généralement avec des données en virgule flottante de simple précision (`Float32`), alors que les poids dans le modèle Keras étaient en double précision (`Float64`). Malgré nos efforts pour aligner les paramètres d'initialisation, il n'a pas été possible d'obtenir des résultats identiques à ceux de Keras. Cela s'explique en grande partie par la nature non déterministe des algorithmes d'apprentissage profond [241, 112]. Cependant, les résultats obtenus avec Keras étaient cohérents avec l'intervalle de variation attendue, ce qui nous a conduit à maintenir l'utilisation de PyTorch dans nos travaux.

4.5.2 Accélération de l'entraînement et optimisation des hyper-paramètres

Profilage du code pour la réduction du temps d'entraînement

Pour notre étude, nous avons choisi une configuration simple, avec une architecture de réseau de neurones élémentaire et l'emploi de patchs de petite taille ($48 \times 48 \times 48$ voxels). Le traitement initial des données se faisait en chargeant et en manipulant les informations d'entraînement directement en mémoire vive pour accroître l'efficacité. Néanmoins, la durée nécessaire à l'entraînement restait long. Chaque modèle, entraîné sur une carte NVIDIA RTX 1080Ti, requérait environ 6 jours, ce qui restreignait l'examen de structures CNNs plus avancées. Pour pallier ce défi, des modifications ont été apportées tant dans l'approche d'entraînement que dans la préparation de données, réduisant ainsi le temps nécessaire à environ 4 jours par modèle.

En ce qui concerne le processus d'entraînement, nous avons recours à l'optimisation en précision mixte, qui associe les formats de virgule flottante en demi-précision (`Float16`) et en simple précision (`Float32`). Cette méthode facilite les calculs lors des étapes de propagation avant et arrière des gradients. Pour la production des données, une évaluation du mécanisme de création des patchs d'entraînement a été menée

via le module `cProfile` [3] de Python, mettant en évidence que l'essentiel du temps était alloué à l'ajout de distorsions sur les patchs avec anévrismes. Cela était particulièrement vrai pour l'opération d'interpolation d'image exécutée par `scipy.ndimages.map_coordinates`. L'option "mode" de cette fonction, définissant la méthode d'interpolation pour les nouveaux voxels, s'est avérée cruciale. Pour optimiser cette étape, nous avons préféré le mode "constant" au mode "nearest", sans recourir au pré-filtrage. Ce choix a diminué la précision des champs de distorsion, augmentant ainsi l'efficacité de l'interpolation et préservant la qualité des données produites. De plus, une optimisation du code a été réalisée en parallélisant les opérations grâce à la bibliothèque `JobLib` lorsque cela était réalisable.

Optimisation des hyper-paramètres

Pour améliorer les performances de notre méthode, nous avons mené une série d'expérimentations axés sur l'optimisation des hyperparamètres. Cela a inclus l'exploration de différentes techniques d'augmentation des données, de fonctions de coût, et de stratégies d'ajustement du taux d'apprentissage. Toutefois, le temps requis pour entraîner chaque modèle, environ 4 jours, rendait l'emploi de la validation croisée pour une évaluation précise impraticable. En réponse, nous avons divisé les données du CHRU de Nancy en deux ensembles : un pour l'entraînement, comprenant 92 images et 138 anévrismes de taille moyenne 4,11 ± 2,48 mm (allant de 1,00 à 19,63 mm, dont 52 de moins de 3 mm et 51 entre 3 et 5 mm) ; l'autre pour les tests, avec 40 images et 68 anévrismes de taille moyenne 3,67 ± 1,93 mm (étendue de 0,96 à 10,52 mm, incluant 13 de moins de 3 mm et 26 entre 3 et 5 mm). Les expérimentations décrites dans la suite de cette section ont été réalisées sur l'ensemble d'entraînement, tandis que l'évaluation finale a été effectuée sur l'ensemble de test. L'absence d'un ensemble de validation s'explique par deux raisons : notre méthode d'ajustement du taux d'apprentissage n'en dépend pas, et nos tests avec un nombre élevé d'époques (100 contre 200 époques) n'ont pas montré de signes de sur-apprentissage, ne nécessitant pas un arrêt anticipé de l'entraînement.

Nos premières expériences ont étendu l'augmentation des données au-delà des rotations et translations habituelles, en intégrant l'ajout de bruit gaussien (moyenne de 0, écart-type de 0,15) et l'application d'effets miroir sur les trois axes anatomiques (axial, sagittal, coronal). Une autre piste explorée fut l'emploi de la normalisation par le z-score, contrastant avec la méthode min-max traditionnelle. Concernant les fonctions de coût, nous avons testé diverses configurations, incluant la combinaison de l'entropie croisée binaire (BCE) et du coefficient Dice - une pratique répandue en segmentation d'images [188]. L'introduction du score Kappa de Cohen comme fonction de coût, aux côtés de la BCE, a également été évaluée pour sa capacité à capturer le signal rare des anévrismes.

Pour l'ajustement du taux d'apprentissage, au lieu d'un taux constant de 10^{-4}, nous avons expérimenté avec différentes stratégies visant à améliorer la convergence et la stabilité. Ces stratégies comprenaient l'approche par paliers, réduisant progressivement le taux après chaque série de 20 époques jusqu'à un minimum de 10^{-4}, et une méthode d'échauffement suivi d'une décroissance polynomiale, à la manière du modèle DeepLab [55]. Cette dernière augmentait le taux progressivement durant les trois premières époques jusqu'à 10^{-2}, pour ensuite le réduire selon $10^{-2} \times (1 - (iter/iter_{max})^{0.9})$, avec $iter$ et $iter_{max}$ désignant l'itération courante et l'itération finale, respectivement. Les détails et résultats de ces approches sont synthétisés dans la table 4.4.

L'introduction de bruit gaussien dans les données d'entraînement n'a pas significativement amélioré les résultats, probablement en raison des niveaux de bruit déjà faible dans notre ensemble de données provenant d'une machine IRM 3T. En revanche, l'effet miroir appliqué aux images le long du plan sagittal a eu un impact positif, vraisemblablement grâce à la symétrie anatomique intrinsèque du cerveau et des vaisseaux sanguins dans ce plan. Concernant la normalisation, bien que les méthodes z-score et min-max ont montré des résultats similaires, nous avons choisi la normalisation z-score en raison de sa moindre sensibilité aux valeurs extrêmes [234]. En ce qui concerne les fonctions de coût, l'introduction du coefficient Kappa de Cohen a légèrement amélioré les performances, mais la combinaison de la BCE et du coefficient Dice s'est avérée plus efficace pour la segmentation, équilibrant à la fois la sensibilité et le taux de FP/cas. Enfin, pour les politiques de mise à jour du taux d'apprentissage, aucune différence significative n'a été observée au niveau des résultats de l'évaluation finale. Cependant, la politique d'échauffement suivi

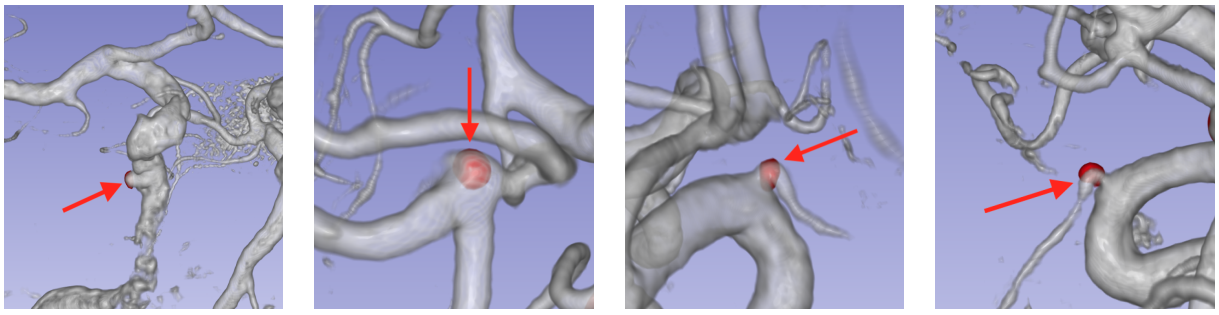
3. `https://docs.python.org/3/library/profile.html`

	Modèles	AP (%)	Sensibilité (%)	FP/cas
Keras	Modèle de base	59,82	70,58	0,61
Pytorch	Modèle de base	58,94	67,64	0,66
	Bruit gaussien	54,70	64,70	0,79
	Effet miroir − axial	57,10	66,17	0,71
	* Effet miroir − sagittal	62,09	70,58	0,61
	Effet miroir − coronal	58,42	66,17	0,67
	* Normalisation z-score	60,03	69,11	0,63
	* BCE + Dice	65,04	73,52	0,58
	Kappa	58,71	61,76	1,04
	BCE + Kappa	61,52	67,64	0,84
	Plateau	59,41	69.11	0,63
	* Échauffement + Polynomial	61,10	67.64	0,60
	Modèle final	69,45	73,52	0,51

TABLE 4.4 – Optimisation des hyperparamètres sur un ensemble de test (40 images, 68 anévrismes). Les métriques sont basées sur un seuil IoU de 10%. * indique les éléments intégrés dans le modèle final.

d'une réduction polynomiale offre une stabilité en début d'entraînement, favorisant ainsi une amélioration constante des performances tout au long du processus d'apprentissage [108].

Suite à ces résultats, nous avons adopté une configuration finale pour notre modèle. Cette configuration comprend l'effet miroir autour du plan sagittal pour l'augmentation des données et la normalisation z-score. Nous avons choisi d'utiliser la combinaison de la BCE et du Dice comme fonction de coût, avec une stratégie d'ajustement du taux d'apprentissage basée sur l'échauffement et une diminution polynomiale [55]. Le modèle final a démontré un score de précision moyenne (AP) de 69,45%. À un seuil de classification par défaut de 50%, une sensibilité de 73,52% associée à un taux de 0,51 FP/cas cas ont été observés. Cependant, malgré ces améliorations, le modèle final présente encore des limites, notamment pour la détection des anévrismes de petite taille (< 2 mm), avec une sensibilité de 53,84% et un taux relativement élevé de faux positifs (voir exemples en Figure 4.11). Dans cette perspective, nous avons exploré des stratégies pour surmonter ces limitations, notamment l'intégration de connaissances médicales a priori dans les réseaux de neurones.



(a) (b) (c) (d)

FIGURE 4.11 – Exemples de faux positifs dans différentes localisations : (a) irrégularité de l'artère carotide, et (b, c, d) origine de petites artères.

4.5.3 Intégration de l'a-priori médical pour améliorer la détection

Dans la pratique clinique, les radiologues se basent sur leur expertise pour repérer les anévrismes dans les images médicales, qui se caractérisent par leur abondance d'informations visuelles. Cette expertise intègre divers aspects, notamment la morphologie et la localisation des anévrismes au sein du système vasculaire cérébral, avec une attention particulière portée au polygone de Willis [208]. Dans cette perspec-

tive, nous examinons plusieurs approches visant à améliorer la détection des petits anévrismes et à réduire les faux positifs. L'objectif est d'incorporer cette expertise humaine et ces connaissances préalables dans le processus d'apprentissage des réseaux CNNs.

Amélioration de la détection des petits anévrismes

La détection d'anévrismes de petite taille par le réseau UNet présente plusieurs défis. L'augmentation du champ réceptif au fil des couches, en particulier pendant le sous-échantillonnage, entraîne une réduction progressive du signal anévrismal [130]. Par exemple, un anévrisme de 2 mm de diamètre est représenté par une zone d'environ 5 voxels de diamètre. Cette taille est réduite à 0,6 voxel à partir du quatrième niveau du réseau UNet, après trois couches de sous-échantillonnage. De plus, les petits anévrismes ont un faible contraste par rapport aux tissus environnants, ce qui rend difficile l'extraction de leurs caractéristiques de manière fiable. De surcroît, la rareté des données d'entraînement contenant de tels anévrismes, avec seulement 28 anévrismes de moins de 2 mm de diamètre sur un total de 206, limite la capacité d'apprentissage de leurs caractéristiques et la généralisation du modèle.

Pour améliorer la détection de ces petits anévrismes, nous avons explorées les 5 approches suivantes :

1. **Augmentation des données par zoom :** Cette technique augmente aléatoirement la résolution spatiale des patchs d'entraînement pour les petits anévrismes de moins de 2mm, en utilisant un facteur de zoom de 1 à 2 fois. L'objectif est d'améliorer la visibilité et la détection de ces anévrismes, souvent difficiles à identifier du fait de leur petite taille et de leur faible signal sur les images.

2. **Échantillonnage basé sur la taille :** Cette approche vise à pallier le déficit de petits anévrismes dans les données. Pour mieux représenter les anévrismes de toutes tailles, une méthode d'échantillonnage basée sur la taille a été adoptée : les patchs avec de petits anévrismes (< 2mm) ont été dupliqués 75 fois, contre 25 fois pour les autres. L'objectif est d'augmenter la présence des petits anévrismes dans l'ensemble de données d'entraînement.

3. **Réduction de la profondeur du réseau UNet :** Bien que les réseaux de neurones profonds soient efficaces pour identifier des caractéristiques complexes, ils peuvent négliger des détails fins cruciaux pour repérer de petits objets [240, 154, 209]. Pour résoudre ce problème, nous avons testé l'effet de la réduction de la profondeur du réseau UNet de quatre à trois niveaux. Cette ajustement cherche à équilibrer l'extraction de caractéristiques complexes tout en conservant les détails fins nécessaires à la détection de petits anévrismes.

4. **Supervision profonde du réseau UNet :** Pour améliorer la précision de la segmentation des anévrismes à différents niveaux de résolution, le réseau UNet a été renforcé par une supervision profonde (voir Figure 4.12). Cette technique introduit plusieurs couches de prédiction opérant à des résolutions distinctes. Elles contribuent significativement à la reconstruction et à la formulation de la segmentation finale, en tirant parti des informations provenant de diverses échelles.

5. **Mécanisme d'auto-attention :** Inspirée par le succès des architectures de type Transformers [84], cette approche intègre un mécanisme d'auto-attention à huit têtes au sein du réseau U-Net. Son objectif est d'analyser les interactions spatiales entre les éléments des cartes de caractéristiques. Conformément au modèle présenté dans [229], ce mécanisme d'attention est positionné entre les phases d'encodage et de décodage du réseau. Ce placement vise à exploiter la capacité d'encodage du mécanisme pour améliorer les caractéristiques extraites par le réseau. De plus, il fonctionne sur des cartes de caractéristiques de dimension $512 \times 6 \times 6 \times 6$, tenant compte de la capacité de calcul importante requise par ce type de mécanisme.

Les différentes approches établies pour améliorer la détection des petits anévrismes ont été évaluées en utilisant les mêmes ensembles d'entraînement et de test que ceux mentionnés précédemment (voir section 4.5.2). L'entraînement a duré 100 époques, utilisant la combinaison de Dice et BCE comme fonction de coût. Les résultats sont synthétisés dans la table 4.5 (en haut).

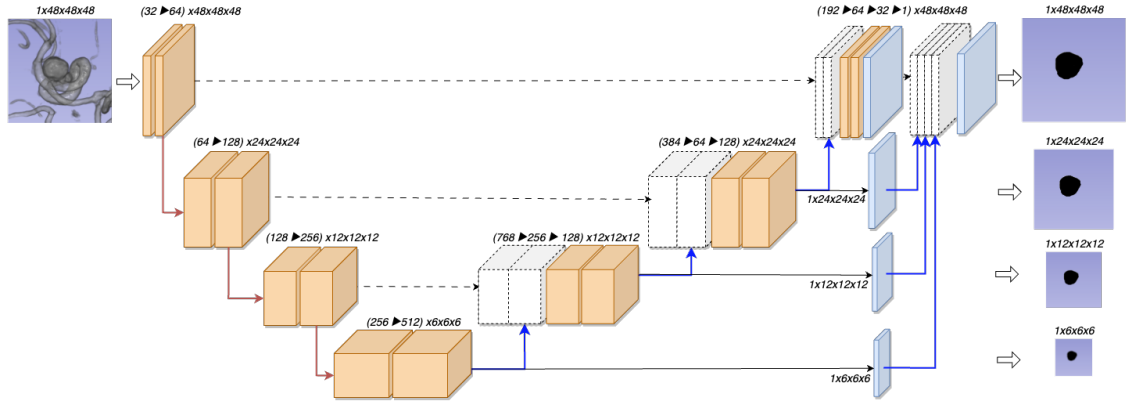


FIGURE 4.12 – Réseau UNet avec une supervision profonde : Combinaison des coûts pondérés, calculés par la formule $\sum_{i=1}^{4} \frac{1}{2^{i-1}} \times L_i$ à chaque résolution i. La vérité terrain est dérivée par interpolation.

L'utilisation du zoom comme technique d'augmentation des données n'a pas entraîné d'amélioration significative dans la détection d'anévrismes, ce qui pourrait résulter du besoin d'appliquer une augmentation lors du test (Test-Time Augmentation, TTA), impliquant l'utilisation de différentes résolutions pour une même image de test et une agrégation des résultats par vote majoritaire. Par ailleurs, l'accentuation des petits anévrismes dans l'ensemble d'entraînement a amélioré la détection, faisant passer la sensibilité de 53,84 à 69,23%, bien que cette technique ait été moins efficace pour les anévrismes plus grands. La réduction de la profondeur du réseau UNet à trois niveaux a diminué la performance générale, avec une précision moyenne tombée à 53,72%, indiquant que l'identification des caractéristiques discriminantes des anévrismes bénéficie d'un réseau plus profond. L'intégration d'une supervision profonde dans le décodeur a favorisé une meilleure reconstruction et segmentation, augmentant ainsi la sensibilité. Cependant, cette amélioration était modérée pour les petits anévrismes. L'adoption d'un mécanisme d'auto-attention multi-têtes a profité à la détection des anévrismes de taille moyenne par une meilleure gestion des relations spatiales éloignées lors de l'extraction des caractéristiques, mais sans avancée notable pour les petits anévrismes. De plus, cette approche a allongé le temps d'entraînement de 16 heures supplémentaires par rapport au modèle UNet standard.

	Paramètres (millions)	**AP (%)**	**Sensibilité (%)**				**FP/cas**
			Totale	< 2mm	$2-5$mm	> 5mm	
Modèle de référence	19,07	69,45	73,52	53,84	73,80	92,30	0,51
Zoom	19,07	69,10	68,27	61,53	73,80	92,30	0,49
Sur-échantillonnage	19,07	70,40	73,52	69,23	76,17	69,23	0,55
Profondeur réduite	4,58	53,72	67,64	46,15	71,41	76,92	0,31
Supervision profonde	19,07	72,29	75,00	61,53	73,81	92,30	0,48
Auto-attention	19,20	73,10	76,47	53,84	78,57	92,30	0,43
Annotations raffinés	19,07	71,17	75,00	46,15	78,57	92,30	0,47
Deux résolutions	20,78	72,82	76,47	69,23	73,80	92,30	0,40
Apprentissage par transfert	19,07	67,74	69,11	61,53	66.66	84,61	0,55
UNet à deux sorties	31,11	64,51	61,76	38,46	61.90	100	0,30
Double UNet	38,14	74,60	79,41	53,18	85,71	100	0,34

TABLE 4.5 – Stratégies d'amélioration de la détection des petits anévrismes (en haut) et de réduction des faux positifs (en bas). Métriques d'évaluation basées sur un seuil d'IoU de 10%, appliquées à un ensemble de test composée de 68 anévrismes (13 anévrismes $<$ 2mm, 42 anévrismes entre 2 et 5mm).

Réduction des faux positifs

Notre modèle de référence présente un taux de 0,51 faux positifs par cas. Bien qu'il soit concurrentiel par rapport aux méthodes de l'état de l'art, il reste significativement plus élevé que celui d'un expert, ce qui

pourrait restreindre son application en contexte clinique. Nous avons constaté que ces erreurs se produisent principalement au niveau des origines des petites artères, nécessitant ainsi une analyse particulière de la part des radiologues pour différencier ces cas. Afin de réduire le nombre de faux positifs, nous avons exploré deux approches principales. La première repose sur l'analyse morphologique des anévrismes afin d'apprendre leurs caractéristiques distinctes par rapport aux structures vasculaires normales. La seconde approche évalue la similarité entre les prédictions et les vaisseaux sanguins, en intégrant davantage de contexte anatomique afin d'améliorer la détection des anévrismes.

Les diverses stratégies adoptées dans cette perspective sont les suivantes :

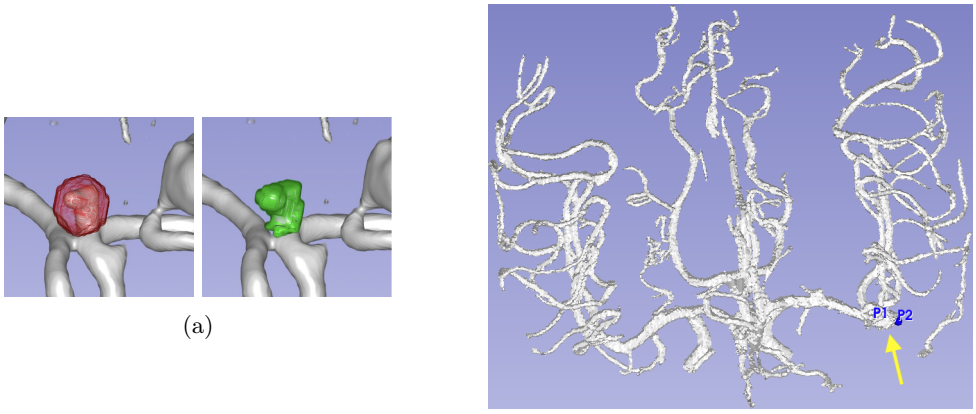
1. **Raffinement des annotations sphériques :** Pour améliorer la précision des annotations sphériques des anévrismes, nous avons adopté une méthode automatisée. L'objectif était de mieux refléter la morphologie réelle des anévrismes, un aspect qui pourrait être intéressant pour améliorer l'efficacité de l'apprentissage des CNNs [155]. Notre méthode commence par un seuillage d'Otsu [219] pour analyser les intensités dans un volume de $48 \times 48 \times 48$ voxels centré sur l'anévrisme. Cette analyse est suivie d'une opération de dilatation pour raffiner les annotations (voir un exemple dans la Figure 4.13a). Cette dilatation est surtout avantageuse pour les anévrismes de grande taille, caractérisés par des niveaux d'intensité variés.

2. **UNet avec deux résolutions en entrée :** L'analyse par ablation a révélé qu'une augmentation des informations contextuelles dans le patch entraîne une réduction des faux positifs, tandis qu'une résolution plus élevée améliore la sensibilité de la détection. Inspirés par le modèle DeepMedic [153], nous avons mis en œuvre un réseau UNet doté de deux entrées pour les branches encodeurs : une fonctionnant à une résolution normale et l'autre à basse résolution (voir Figure 4.14). Les caractéristiques extraites de chaque branche sont fusionnées par concaténation, offrant ainsi une double perspective qui intègre une vision globale des structures anatomiques et des détails fins [173].

3. **Apprentissage par transfert :** Pour guider le modèle dans la compréhension des interactions entre anévrismes et structures vasculaires, nous avons d'abord concentré l'entraînement sur la segmentation des vaisseaux sanguins. La segmentation manuelle des vaisseaux, effectuée à l'aide du logiciel Slicer [92], a servi de données de référence (voir Figure 4.13b). Sur cette base, le réseau UNet a subi un pré-entraînement de 50 époques. Ensuite, utilisant la technique de transfert d'apprentissage, nous avons adapté le réseau pour qu'il se concentre pendant 50 époques additionnelles sur la segmentation des anévrismes, tirant parti des connaissances acquises lors du pré-entraînement.

4. **Segmentation simultanée :** Nous avons également évalué une technique d'entraînement du réseau UNet pour la segmentation simultanée des anévrismes et des vaisseaux. Cette technique a impliqué l'ajout de deux couches de convolution parallèles à la sortie du décodeur, produisant ainsi deux images distinctes : une pour la segmentation des vaisseaux et une autre pour les anévrismes. Pour éliminer les anévrismes situés en dehors des vaisseaux, nous avons réalisé un produit entre ces deux images lors de l'entraînement. La fonction de coût a été modifiée pour se concentrer sur la segmentation des anévrismes, avec une pondération de $2/3$, contre $1/3$ pour les vaisseaux.

5. **Double UNet avec deux résolutions :** Cette approche utilise deux réseaux UNet distincts, ciblant séparément la segmentation des vaisseaux sanguins et des anévrismes (voir Figure 4.15). Le premier réseau, travaillant à une résolution réduite, se focalise sur les vaisseaux sanguins pour capturer un contexte plus large et identifier les structures vasculaires étendues. Parallèlement, le deuxième réseau, opérant à une résolution normale, est dédié à la segmentation des anévrismes. Les cartes de caractéristiques identifiées par ces deux réseaux sont combinées à chaque niveau du décodeur, grâce à un mécanisme d'attention, afin de renforcer la détection des anévrismes et de réduire les faux positifs. La fonction de coût attribue une pondération de $2/3$ pour la segmentation des anévrismes et de $1/3$ pour celle des vaisseaux sanguins.



(a)

(b)

FIGURE 4.13 – (a) Comparaison des annotations sphériques (rouge) et des annotations raffinées (vert). (b) Segmentation manuelle des vaisseaux sanguins sur une image présentant un anévrisme (flèche).

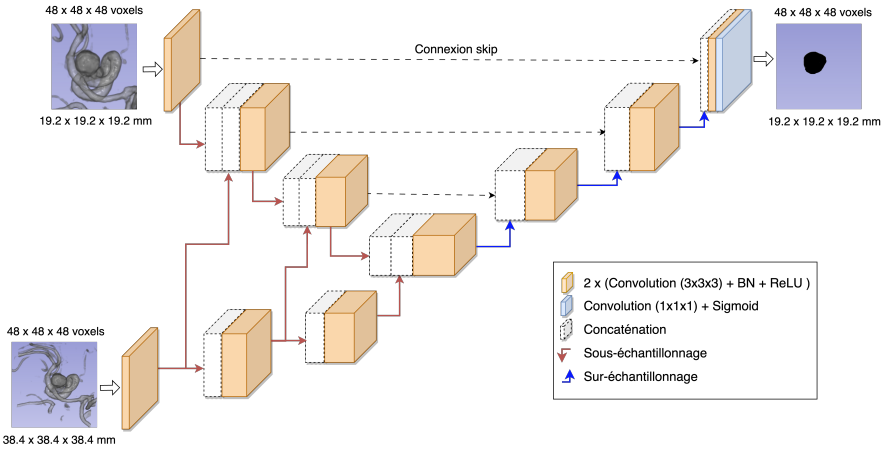


FIGURE 4.14 – Réseau UNet avec deux résolutions en entrée : Fusion des cartes de caractéristiques à chaque niveau de l'encodeur via une concaténation.

Les résultats des stratégies mentionnées sont synthétisés dans la table 4.5 (en bas). Elles ont contribué à la diminution des faux positifs, bien que la sensibilité de plusieurs modèles ait baissé. L'emploi d'annotations plus précises pour l'entraînement a amélioré la précision des segmentations, les alignant davantage sur les contours réels des anévrismes, contrairement aux formes sphériques simplifiées produites par le modèle initial. Cependant, ces améliorations n'ont pas significativement influencé la capacité de détection, évaluée par l'indice IoU entre les sphères : dans de nombreux cas, les composantes connexes (CC) se révélaient plus petites que les sphères annotées. Avec l'application de deux résolutions différentes au réseau UNet, une baisse des faux positifs (de 0,51 à 0,40) et une amélioration marginale de la sensibilité ont été constatées. Néanmoins, l'adaptation de connaissances transférées pour la segmentation des vaisseaux n'a pas notablement renforcé la détection d'anévrismes, le score Dice atteignant une moyenne de 74,15% après 50 époques. Un modèle UNet à double sortie, ciblant séparément les vaisseaux et les anévrismes, a limité les faux positifs en dehors des zones vasculaires tout en réduisant la sensibilité aux anévrismes (de 69,45% à 61,76%). Par contre, le modèle Double UNet, traitant les vaisseaux et anévrismes simultanément, a vu sa sensibilité augmenter à 79,41%. Malgré cet effet positif, les contraintes de mémoire et la durée d'entraînement prolongée, environ 8 jours, limitent son application pratique.

À noter que pour une évaluation précise de ces différentes stratégies, il est essentiel d'avoir des données d'évaluation supplémentaires et d'utiliser un schéma de validation croisée. Ces méthodes permettent une

analyse plus approfondie et fiable des performances des modèles. Cependant, dans notre cas, nous sommes confrontés à des ressources limitées, tant en termes de puissance de calcul que de disponibilité des données.

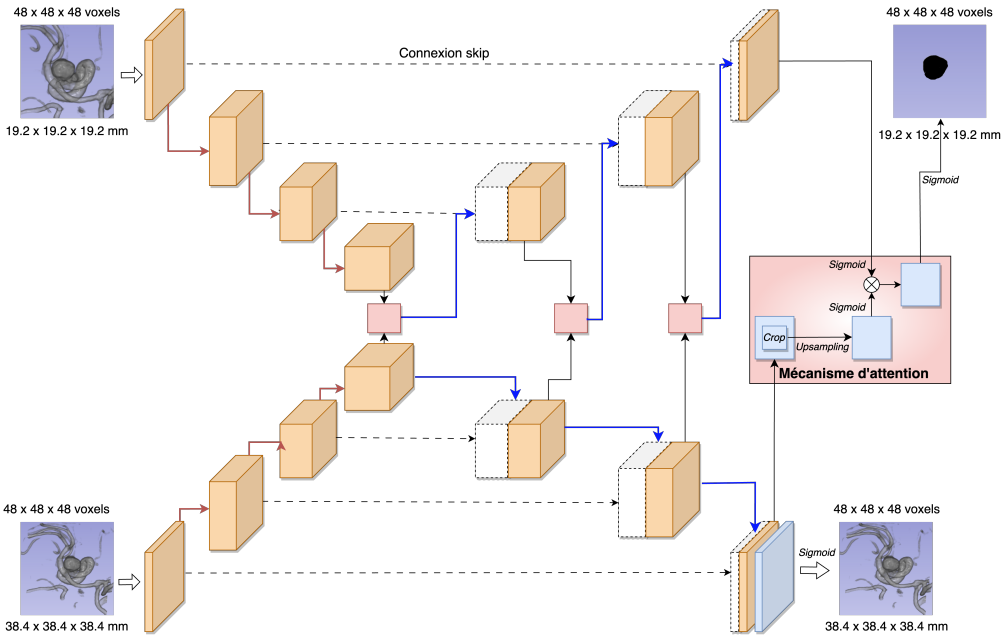


FIGURE 4.15 – Réseau Double UNet traitant deux résolutions : Fusion des caractéristiques des vaisseaux sanguins (en bas) et des anévrismes (en haut) dans le décodeur avec un mécanisme d'attention.

Conclusion

Ce chapitre a présenté notre stratégie de gestion des données pour la détection d'anévrismes intracrâniens. Elle s'appuie sur l'utilisation de petits patchs, des techniques d'échantillonnage et de synthèse de données pour surmonter les ensembles de données, le déséquilibre des classes et la rareté des cas d'anévrismes. L'évaluation de notre approche a été réalisée en utilisant le réseau de segmentation classique 3D UNet et comparée aux méthodes issues de la compétition ADAM [285]. Toutefois, les métriques de cette compétition, axées sur la distance, se sont révélées inadaptées aux approches de détection par des réseaux de segmentation. En revanche, nous avons évalué notre approche en utilisant le critère IoU, atteignant une sensibilité de 69,76% avec 0,63 FP/cas. En comparaison, le radiologue a obtenu une sensibilité de 89,47% avec 0,018 FP/cas. Le défi des méthodes automatisées est de trouver le juste équilibre entre sensibilité et taux de faux positifs. Les petits patchs ont tendance à montrer une bonne sensibilité mais génèrent davantage de faux positifs, tandis que les patchs plus grands réduisent les faux positifs mais avec une sensibilité moindre.

Notre étude a mis en évidence deux défis principaux : la détection difficile des anévrismes de petite taille et un taux élevé de faux positifs. Pour adresser ces enjeux, diverses approches ont été testées, y compris l'adoption de mécanismes d'attention et l'intégration de connaissances médicales préalables. Néanmoins, la capacité de ces méthodes à détecter de petits objets, comme les anévrismes, s'avère restreinte lorsqu'on se repose sur la segmentation. Cette limitation est souvent liée à la qualité des images reconstruites et aux méthodes de post-traitement utilisées, entraînant des variations marquées dans les résultats de détection. Une des hypothèses que nous formulons tient à la complexité intrinsèque du problème, exacerbée par la taille réduite des bases de données disponibles et par les défis liés à la détermination de l'appartenance d'un voxel à la poche anévrismale, même pour un expert. Le chapitre suivant introduira une approche de détection d'anévrismes s'appuyant sur un réseau dédié à la détection d'objets. Le principe est d'aborder la question sous une forme moins ambitieuse, mais sans perdre de pertinence clinique, et voir si cette formulation plus contrainte mène une détection plus performante.

Chapitre 5

Détection d'anévrismes intracrâniens : une perspective de détection d'objet

Sommaire

5.1 Limitations des travaux existants de détection d'anévrismes 102
5.2 Collection et préparation de données 104
5.2.1 Ensembles de données et annotation 104
5.2.2 Échantillonnage et génération de données 105
5.3 Détection des anévrismes avec des représentations sphériques 106
5.3.1 Réseau de détection d'objets et fonction de coût 106
5.3.2 Détails d'implementation 108
5.3.3 Métriques d'évaluation et méthodes de référence 109
5.4 Étude expérimentale et résultats 110
5.4.1 Étude par ablation 111
5.4.2 Évaluation par validation croisée 112
5.4.3 Évaluation par experts et consolidation des annotations 114
5.4.4 Influence des petits patchs sur des données limitées 116
5.5 Discussion et analyse qualitative 117
5.5.1 Architectures de réseau et déséquilibre de classes 117
5.5.2 Représentations sphériques et analyse des résultats 118

Dans ce chapitre, nous exposons notre deuxième contribution à la détection des anévrismes intracrâniens à partir d'images TOF-MRA. La méthode proposée se distingue par l'adoption d'une approche de détection d'objets sans ancres, fondée sur des représentations sphériques. La section 5.1 commence par une synthèse de l'état de l'art en détection d'objets dans l'imagerie 3D, mettant en évidence les limites des méthodes existantes. Cette revue met en lumière les limites inhérentes aux méthodes de détection existantes, notamment celles qui font usage d'ancres et de boîtes englobantes lorsqu'elles sont appliquées à la détection des anévrismes. La section 5.2 décrit le processus de collecte et de préparation des données pour notre étude. Elle détaille les ensembles de données utilisés, l'annotation mise en place, ainsi que des techniques d'échantillonnage et de génération de données destinées à alimenter notre réseau de neurones. La section 5.3 présente de notre réseau CNN, conçu spécifiquement pour la détection des anévrismes en utilisant des sphères. On y aborde l'architecture du réseau, la fonction de coût employée pour traiter le déséquilibre des classes, ainsi que les métriques d'évaluation déployées. La section 5.4 présente les résultats de nos expérimentations, incluant une étude par ablation, une validation croisée, et une évaluation plus approfondie par des experts. Ces différentes approches permettent d'évaluer à la fois l'efficacité de notre méthode et la fiabilité des annotations utilisées. Enfin, la section 5.5 offre une analyse qualitative des résultats, comparant notre méthode aux méthodes existantes.

5.1 Limitations des travaux existants de détection d'anévrismes

Dans le domaine de la détection des anévrismes intracrâniens, deux approches principales se distinguent : l'utilisation de réseaux de segmentation d'images et de réseaux de détection d'objets. Les réseaux de segmentation d'images sont privilégiés en grande partie en raison de leur facilité d'implémentation [139]. Toutefois, cette approche requiert des étapes de post-traitement pour extraire les détections, notamment leurs centres et dimensions, à partir des images prédites. La précision des détections extraites est étroitement liée à la qualité du résultat de segmentation et aux étapes de post-traitement employées, et des métriques adaptées sont nécessaires pour une évaluation plus précise. En contraste, l'adoption des réseaux de détection d'objets pour localiser les anévrismes dans les images est moins courante [81]. Ces réseaux se concentrent sur la localisation des anévrismes à l'aide de représentations spécifiques, souvent par des boîtes englobantes. De plus, ils se révèlent particulièrement adaptés au contexte clinique, où l'accent est mis sur la détection des anévrismes plutôt que sur leur segmentation. Par ailleurs, comme nous l'avons vu dans le chapitre 2, l'évaluation qualitative des anévrismes (taille, volume, etc.) se fait généralement à partir d'images d'angiographie 3DRA, utilisée en pré-intervention et offrant une résolution supérieure par rapport aux images TOF-MRA. Cette section examine les avantages et les limites des réseaux de détection fréquemment utilisés, en particulier dans le domaine de l'imagerie médicale.

Réseaux de détections d'objets

La moins fréquente utilisation des réseaux de détection d'objets par rapport aux réseaux de segmentation s'explique en grande partie par leur complexité intrinsèque. Ces systèmes nécessitent des ajustements spécifiques, notamment l'utilisation de plusieurs fonctions de coût à la fois pour évaluer la localisation et le score de classification des objets. De plus, ils requièrent des configurations d'ancres personnalisées pour chaque problématique et ensemble de données. Ces configurations comprennent des paramètres tels que le nombre, la taille et la localisation des ancres, afin de compenser le déséquilibre des classes couramment rencontré dans les tâches de détection [139, 176]. Parmi les architectures basées sur des ancres, RetinaNet [176] et YOLO [238] sont devenus des références en matière de détection d'objets en 2D. RetinaNet est reconnu pour sa capacité à détecter des objets à différentes résolutions grâce à son architecture de réseau de pyramides de caractéristiques (FPN). Cependant, cette capacité de détection se traduit souvent par une vitesse de traitement plus lente, ce qui peut représenter un inconvénient dans des contextes à ressources limitées, notamment en imagerie 3D. En revanche, le réseau YOLO se distingue par un meilleur équilibre entre vitesse et précision, grâce à une architecture plus simple. De plus, les dernières versions du réseau YOLO intègrent des stratégies d'optimisation dédiées, comprenant des techniques du type "Bag-of-Freebies" et "Bag-of-Specials", dans le but d'améliorer ses performances [300, 37]. Bien que YOLO puisse être légèrement moins précis que RetinaNet dans certains cas, il demeure un choix préféré pour les applications nécessitant une détection en temps réel. En revanche, l'utilisation de RetinaNet en imagerie médicale 3D est déjà établie [341, 205], tandis que YOLO est moins adopté dans ce domaine, nécessitant des ajustements pour le traitement des données 3D.

Approches de détection par ancres

Dans le domaine spécifique de la détection d'anévrismes, seules deux méthodes ont adopté une approche de détection d'objets. La première méthode [26], basée sur le framework nnDetection [27], a obtenu la troisième place dans la compétition ADAM. La seconde méthode [318], introduite en 2021, utilise un réseau encodeur-décodeur initialement conçu pour détecter les nodules pulmonaires sur des images CT [174]. Pour détecter les anévrismes dans les images, ces deux méthodes utilisent des boîtes englobantes définies par leurs coordonnées centrales (C_x, C_y, C_z) et leurs dimensions (largeur, hauteur, profondeur). Elles s'appuient également sur des réseaux de détection basés sur des ancres, subdivisant l'image en cellules, chacune analysée pour détecter les objets et prédire les boîtes englobantes dans les cellules contenant le centre des objets. nnDetection utilise RetinaUNet [139], une variante de RetinaNet adaptée à l'imagerie médicale, avec 27 ancres par cellule. En contraste, la méthode [318] emploie un réseau RPN [239] enrichi de modules d'attention et de convolutions dilatées, avec 3 ancres par cellule. Ces ancres ont pour rôle de sélectionner et d'ajuster les détections en termes de taille, de rapport d'aspect et de localisation, pour les adapter à chaque objet dans l'image. Cependant, malgré leur efficacité, les détecteurs basés sur des ancres requièrent des configurations spécifiques à chaque application, impliquant

la sélection des caractéristiques des ancres et la mise en place de stratégies d'échantillonnage avancées pour résoudre le problème de déséquilibre entre les ancres positives et négatives [323]. La configuration manuelle de ces paramètres, souvent laborieuse et dépendante de l'expertise humaine, peut limiter les performances globales de ces détecteurs [82]. De plus, ils rencontrent des difficultés dans la détection d'objets de formes irrégulières ou complexes [341]. Leur efficacité diminue pour les petits objets, un défi exacerbé dans notre cas par la grande variabilité de tailles des anévrismes, qui peut aller de moins de 1 à plus de 30 mm. Face à ces défis, des méthodes de détection sans ancres gagnent en popularité, offrent plus de flexibilité et sont adaptées aux particularités de l'imagerie médicale [317, 271].

Approches de détection sans ancres

Les méthodes de détection d'objets sans utilisation d'ancres se démarquent des approches classiques qui reposent sur des ancres pré-configurées. Elles se focalisent sur des points clés pour la détection des objets, comme leurs centres [324, 85] ou leurs extrémités [168, 335]. Dans ces méthodes, une ou plusieurs cellules de l'image sont mobilisées pour détecter chaque objet en estimant les positions des centres et les dimensions des régions occupées par les objets grâce au calcul des décalages par rapport à ces points clés. Par exemple, FCOS [284] génère une boîte englobante pour chaque pixel de l'objet, en mesurant les distances entre le point central et les bords de la boîte, et en incluant un score de centrage pour affiner les prédictions en tenant compte de la proximité des pixels par rapport au centre de l'objet. D'un autre côté, CenterNet [85] considère chaque objet comme étant le centre de sa propre boîte englobante, produisant ainsi trois éléments essentiels : une carte thermique centrée sur l'objet, la taille de l'objet et un décalage local. Ces diverses approches améliorent la précision de la localisation des objets sans recourir à des ancres pré-définies.

Dans le domaine de l'imagerie médicale en 3D, l'utilisation de détecteurs sans ancrages demeure relativement peu courante. En 2020, le réseau CPM-Net [271] a adopté une stratégie de correspondance des points centraux pour la détection de nodules pulmonaires sur des images CT. Cette approche repose sur l'utilisation des 7 cellules les plus proches du centre de chaque objet pour le détecter tout en calculant la taille de la boîte englobante et le décalage du centre par rapport à chaque cellule. CPM-Net utilise une architecture encodeur-décodeur pour générer des prédictions à différents niveaux de résolution. Afin de mieux capturer l'information spatiale et le contexte en 3D, cette architecture intègre des cartes de coordonnées spatiales à plusieurs niveaux, accompagnées de modules d'attention Squeeze-and-Excitation [127]. Pour améliorer la détection des objets rares, CPM-Net utilise une technique d'entraînement appelée "Hard Negative Mining" [267], en plus de la fonction de coût Focal [176]. Cependant, ces détecteurs sans ancres dépendent fortement de la sélection des points clés au niveau des objets et peuvent rencontrer des difficultés lors de la détection d'objets présentant des formes irrégulières ou de petite taille [271].

Représentations des détections

La plupart des travaux de recherche en détection d'objets utilisent des boîtes englobantes pour représenter les objets à détecter. Cependant, pour des objets biomédicaux de forme sphérique tels que les anévrismes, les boîtes englobantes peuvent ne pas être la meilleure représentation [213, 185]. En revanche, les représentations sphériques s'avèrent particulièrement pertinentes pour les anévrismes. La taille de l'anévrisme, un facteur crucial dans l'évaluation du risque de rupture [34], peut également être exprimée par le diamètre de la sphère. Certaines études ont exploré des alternatives, telles que l'utilisation de cercles [213] et d'ellipses [53], pour représenter des objets dans des images 2D. En 2021, le modèle SCPM-Net [185], une variation du modèle CPM-Net [271], a introduit l'utilisation de représentations sphériques pour la détection de nodules pulmonaires en imagerie CT. Il convient de noter que dès le début de nos travaux de thèse en novembre 2020, nous avons adopté une annotation par sphères pour les anévrismes dans notre base de données.

Le modèle SCPM-Net est basé sur la même architecture du modèle CPM-Net, mais intègre une fonction de coût spécifique pour optimiser l'intersection sur l'union (IoU) entre les sphères pendant l'entraînement. La fonction de coût totale inclut plusieurs composantes et hyper-paramètres, dont une version simplifiée de cette fonction est présentée dans l'équation 5.1.

$$\text{Coût total} = \text{Confiance} + \beta \cdot (\text{Rayon} + \text{Décalage} + 2 \times \text{IoU}_{Sphere}) \tag{5.1}$$

Pour optimiser les scores de confiance générés par le modèle, une variante de la fonction Focal est appliquée à toutes les cellules (α=0,375 et γ=2). La localisation des objets, à savoir leurs centres et leurs rayons, est optimisée uniquement pour les cellules positives (β=1). L'optimisation du rayon est réalisée en utilisant la perte L1, tandis que l'ajustement du décalage entre chaque cellule positive et le centre de l'objet utilise l'erreur quadratique moyenne (MSE). L'optimisation de l'indice de similarité IoU entre les sphères s'appuie sur une formule qui intègre divers paramètres, notamment l'aire d'intersection, la distance entre les centres et l'angle d'intersection. Une limitation de cette approche réside dans sa dépendance à l'égard de multiples hyper-paramètres, lesquels ont été ajustés à la suite de plusieurs expérimentations. Cette caractéristique est analogue à la limitation des méthodes de détection basées sur des ancres. Une autre observation importante est que, en outre, elle a montré un taux de faux positifs élevé (7 FP/cas associés à une sensibilité de 89,2%) lorsqu'elle a été appliquée à la détection de nodules pulmonaires.

Contributions

Dans ce chapitre, nous introduisons une approche de détection d'anévrismes en une étape, évitant l'usage d'ancres. Notre méthode, nommée DeepAneDet, utilise des représentations sphériques pour les anévrismes, ce qui simplifie l'annotation et en augmente l'efficacité. Pour compenser le manque de données d'entraînement, nous avons adapté la stratégie de gestion des données décrite dans le chapitre précédent, recourant à de petits patchs et à des méthodes d'échantillonnage et de synthèse. En plus de notre approche de gestion de données, nous avons également abordé le déséquilibre des classes au niveau de la fonction de coût, de manière simple et sans nécessiter d'ajustements complexes. L'efficacité de DeepAneDet a été validée par une étude d'ablation sur un ensemble de données privées, suivie d'une évaluation par validation croisée sur un large ensemble de données publiques, annotées avec soin par deux experts, assurant la fiabilité des références. Nos résultats illustrent la compétitivité de notre méthode face aux approches de pointe et mettent en lumière les bénéfices des représentations sphériques sur les boîtes englobantes pour la détection des anévrismes.

5.2 Collection et préparation de données

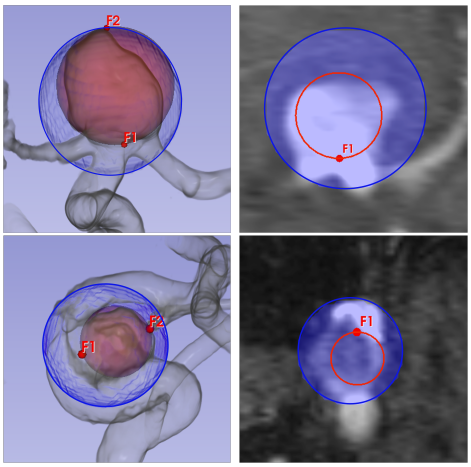
5.2.1 Ensembles de données et annotation

Nous avons utilisé deux ensembles de données TOF-MRA. Le premier ensemble provient du CHRU de Nancy, le même que celui présenté dans le chapitre précédent. Pour rappel, cet ensemble se compose de 132 images, acquises entre 2015 et 2021, avec 75 femmes et 57 hommes. Les critères d'inclusion ont été restreints aux anévrismes sacculaires ayant une taille inférieure à 20 mm, tout en excluant les images de suivi, les anévrismes fusiformes, les anévrismes rompus, et les anévrismes ayant déjà subi un traitement. Ces images présentent des dimensions de $512 \times 512 \times 254$ voxels, avec une résolution voxel de $0,47 \times 0,47 \times 0,4$ mm^3. L'ensemble des 132 images renferme un total de 132 anévrismes, avec une taille moyenne de 3,97 $\pm$ 2,32 mm, une taille maximale de 19,63 mm, et une taille minimale de 0,96 mm. Le nombre d'anévrismes par image varie de 1 anévrisme dans 84 images à 5 anévrismes dans 4 images.

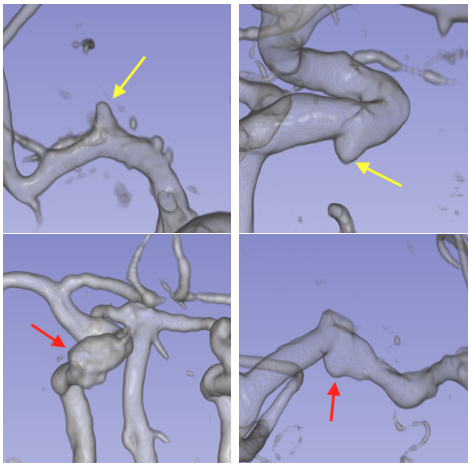
À la fin de l'année 2022, la base de données du Centre Hospitalier Universitaire Vaudois (CHUV) de Lausanne [80] a été rendue publique et a été intégrée dans notre étude. Cette base de données se compose de 296 images TOF-MRA provenant de 284 patients. Ces images se répartissent en 127 images de sujets sains (sans anévrismes) et 168 présentant des anévrismes. Cependant, pour nos analyses, nous avons exclu les images de suivi ainsi que celles contenant des anévrismes fusiformes ou ayant une taille supérieure à 20 mm, réduisant ainsi notre ensemble de données à 270 images. Ces images ont été recueillies entre 2010 et 2015 au moyen de scanners de 1,5T et 3T, et elles présentent diverses résolutions.

La base de données CHUV a été annotée par un radiologue ayant deux années d'expérience en neuro-imagerie, puis vérifiée par un neuroradiologue possédant plus de 15 ans d'expérience. L'annotation a été réalisée en utilisant à la fois des annotations voxel dans 37 images, ainsi que des annotations approximatives sous forme de sphères, similaires à notre approche d'annotation (voir section 4.2.1). Cependant, il convient de noter que notre méthode diffère en termes de précision, car nous définissons les sphères en se basant sur deux points clés : le centre du collet et le dôme, contrairement aux sphères de plus grande taille utilisées dans la base de données du CHUV. Ces dernières ont tendance à inclure des vaisseaux sanguins

sains dans les annotations, ce qui peut être observé dans la Figure 5.1a. Afin de garantir la précision et la fiabilité des annotations, le même expert du CHRU de Nancy a effectué une vérification et une re-annotation des images de la base de données CHUV de Lausanne suivant notre approche d'annotation. Au cours de ce processus, 7 anévrismes supplémentaires ont été identifiés et annotés, tandis que 5 annotations ont été retirées en raison de leur correspondance avec de simples dysplasies ou des irrégularités de la surface vasculaire (voir des exemples en Figure 5.1b). Par conséquent, les 270 images de cette base de données contient désormais un total de 164 anévrismes, avec une taille moyenne de 3,74 ± 2,17 mm, une taille maximale de 18,52 mm, et une taille minimale de 1,23 mm. Le nombre d'anévrismes par image varie de 0 anévrisme dans 130 images à 3 anévrismes dans 3 images. Dans le reste de ce manuscrit, nous désignons ces annotations comme les "annotations initiales". Les caractéristiques des anévrismes dans les deux ensembles de données sont résumées dans la table 5.1.



(a) Comparaison des annotations.



(b) Exemples d'annotations incorrectes.

FIGURE 5.1 – Comparaison des données de la base CHUV [80] : (a) Rendu 3D et plans axiaux montrant notre annotation (rouge) par rapport à l'annotation initiale (bleu) incluant les vaisseaux sains. (b) Exemple d'anévrismes non annotés (3,06 mm et 3,22 mm) en haut, et annotations incorrectes (dysplasie simple ou irrégularité sur athérome à gauche, infundibulum à droite) en bas.

	CHRU Nancy	**CHUV Lausanne [80]**
# Images (sans anévrismes)	132 (0)	270 (130)
# Anévrismes	206	164
< 3mm	81	67
3–5mm	77	72
> 5mm	48	25
# Anévrismes par image	1 – 4	0 – 3
Taille moyenne (mm)	3,97 ± 2,32	3,74 ± 2,17
Taille médiane (mm)	3,41	3,23
Taille min–max (mm)	0,96 – 19,63	1,23 – 18,52

TABLE 5.1 – Comparaison des anévrismes entre les ensembles de données CHRU et CHUV [80].

5.2.2 Échantillonnage et génération de données

L'efficacité d'un détecteur d'objets repose étroitement sur sa capacité à modéliser de manière à la fois les éléments d'intérêt, à savoir les anévrismes, et l'arrière-plan, qui comprend les vaisseaux sanguins et les autres tissus cérébraux. Comme il a été précédemment évoqué dans le chapitre antérieur, l'utilisation de patchs de grande taille favorise cette modélisation en fournissant un contexte anatomique plus vaste,

ce qui tend généralement à réduire le nombre de faux positifs. Toutefois, cette approche présente des défis liés à la charge computationnelle nécessaire pour traiter ces patchs de grande taille. De plus, elle nécessite l'intégration de données provenant de sujets sains afin d'assurer une modélisation fiable des régions sans anévrismes, étant donné la difficulté d'extraire de grands patchs sans inclure des anévrismes. En revanche, notre méthode se base sur l'utilisation de patchs de petite taille, ce qui permet l'extraction de multiples patchs négatifs à partir de chaque image, éliminant ainsi la nécessité a priori d'utiliser des données provenant de patients sains.

Au lieu d'utiliser des patchs de $48 \times 48 \times 48$ voxels, nous avons opté pour l'utilisation de patchs de dimensions plus grandes, à savoir $96 \times 96 \times 96$ voxels. Cette décision découle de la capacité démontrée des patchs de plus grande dimension à établir un équilibre entre la sensibilité et le taux de faux positifs. De plus, cette modification a été motivée par l'accès à une carte graphique plus puissante (NVIDIA RTX A6000), ce qui nous permettra de réduire les temps d'entraînement des modèles. En conséquence, notre approche se fonde sur l'utilisation de patchs isotropes mesurant $96 \times 96 \times 96$ voxels, avec une résolution spatiale de 0,4 mm^3 par voxel, qui se rapproche de la résolution nominale. Ces patchs, ayant une taille d'environ $38,4 \times 38,4 \times 38,4$ mm^3, sont extraits sans chevauchement, en veillant à maintenir une distance minimale de 40 mm entre les centres des patchs.

Afin de compenser le déséquilibre entre les classes, nous avons mis en place une stratégie de gestion des données en accord avec les principes énoncés dans le chapitre précédent. La première stratégie consiste à générer 50 duplications de chaque patch positif, en introduisant des distorsions aléatoires pour varier les formes des anévrismes. Cette méthode inclut le déplacement aléatoire de chaque point de contrôle dans une grille de dimensions $3 \times 3 \times 3$ autour du patch, avec un maximum de 3 mm dans toutes les directions, en utilisant une interpolation spline cubique pour créer le champ de distorsion. Dans le but de mieux distinguer les vaisseaux sanguins sains des anévrismes, notre deuxième stratégie sélectionne un total de 40 patchs négatifs non-intersectants par image, dont 30 sont centrés sur des vaisseaux sanguins.

À chaque époque d'entraînement, nous avons sélectionné l'ensemble de patchs d'entraînement de la manière suivante : il comprenait tous les patchs positifs (avec 50 duplications pour chaque anévrisme), auxquels nous avons ajouté une sélection de patchs négatifs représentant 15% de l'ensemble total de patchs d'entraînement. Ces patchs négatifs ont été choisis de manière aléatoire et équilibrée parmi les 40 patchs initiaux extraits de chaque image de patient. Afin d'accroître la diversité des patchs d'entraînement, nous avons appliqué différentes techniques d'augmentation de données : rotations (de 0 à 180°), translations (de 0 à 10 mm) dans les trois dimensions, effets de miroir le long du plan sagittal. Pour les patchs positifs, nous avons veillé à ce qu'ils englobent complètement l'anévrisme à l'intérieur du patch augmenté, sinon l'augmentation serait réappliquée.

5.3 Détection des anévrismes avec des représentations sphériques

Dans cette section, nous décrivons DeepAneDet, notre méthode de détection d'anévrismes intracrâniens. Cette méthode repose sur un processus de détection direct, évitant l'utilisation d'ancres. Nous avons choisi une architecture de réseau CNN pour sa simplicité, en privilégiant des représentations sphériques qui correspondent mieux à la forme des anévrismes, améliorant ainsi leur localisation. Ensuite, nous détaillons la conception de notre fonction de coût, spécialement adaptée pour gérer le déséquilibre des classes dans les données. Enfin, nous présentons les détails d'implémentation de notre méthode, ainsi que les métriques et les méthodes de référence employées pour évaluer son efficacité dans la tâche de détection des anévrismes.

5.3.1 Réseau de détection d'objets et fonction de coût

Architecture du réseau de neurones

Notre approche repose sur un réseau de détection d'objets en une seule étape. Son objectif consiste à localiser les anévrismes en utilisant des représentations sphériques. Le patch d'entrée 3D, ayant des dimensions de $96 \times 96 \times 96$ voxels, est subdivisé en une grille de $12 \times 12 \times 12$, soit un total de 1728 cellules de $8 \times 8 \times 8$ voxels. Chaque cellule est responsable de la détection d'un anévrisme en fonction de la présence ou de l'absence du centre de l'anévrisme à l'intérieur de la cellule. Chaque détection est représentée par

une sphère dont le centre correspond aux coordonnées du centre de l'anévrisme (C_x, C_y, C_z), avec un rayon (r) équivalent à la moitié de la taille de l'anévrisme.

Comme illustré dans la Figure 5.2, le processus de détection commence par la transformation du patch d'entrée en cartes de caractéristiques à l'aide d'un encodeur. Cet encodeur est composé de blocs convolutifs résiduels et de sous-échantillonnage. L'introduction de connexions résiduelles au sein de chaque bloc de l'encodeur facilite l'entraînement en résolvant le problème de la disparition du gradient [118]. Chaque bloc de l'encodeur comprend une couche de convolution, une normalisation par lots (BN) et une fonction d'activation Leaky ReLU. Ensuite, un module de détection est utilisé pour diviser ces cartes en une grille de 1728 cellules, en utilisant deux ensembles de blocs convolutifs et des couches de convolution en parallèle. La première convolution génère un score de confiance, indiquant la présence potentielle d'un centre d'anévrisme pour chaque cellule. Pour les cellules contenant un anévrisme, la deuxième convolution calcule les coordonnées (C_x, C_y, C_z) et le rayon (r) du centre de l'anévrisme, qui sont normalisés par rapport à la taille du patch et de la cellule, respectivement. Les scores de confiance et les coordonnées sont obtenus grâce à une fonction d'activation sigmoïde, tandis que le rayon est transformé en valeurs positives à l'aide d'une fonction exponentielle.

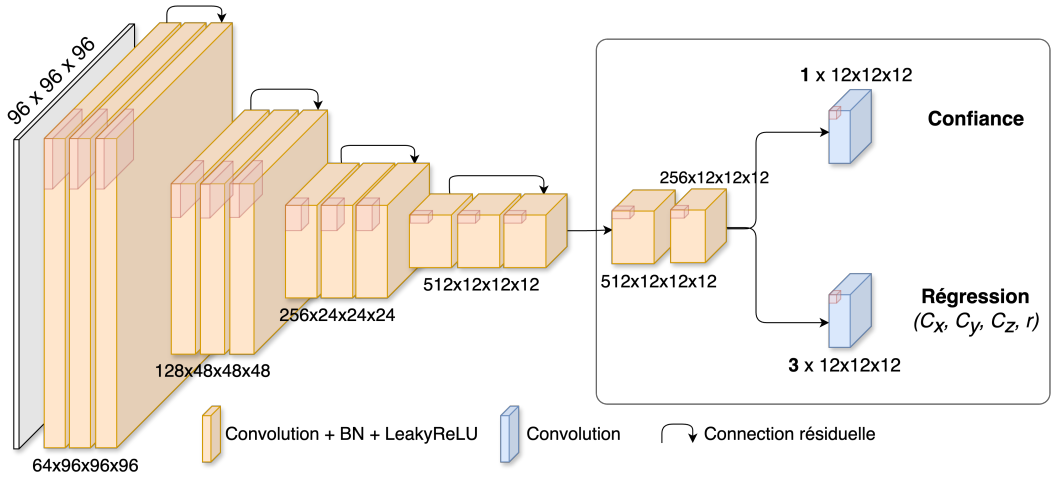


FIGURE 5.2 – Architecture proposée pour la détection des anévrismes utilisant des sphères (C_x, C_y, C_z, r) et des scores de confiance (c). Les cartes de caractéristiques 2D sont représentées à des fins de visualisation ; leurs tailles réelles en 3D sont affichées.

Problème de déséquilibre de classes et fonction de coût

La structure de notre réseau CNN repose sur une approche en grille, ce qui crée un important déséquilibre entre les cellules négatives (N), qui ne contiennent pas d'anévrismes, et les cellules positives (P), qui contiennent les centres des anévrismes. Pour illustrer ce déséquilibre, prenons l'exemple d'un patch d'entrée qui contient un seul anévrisme : dans ce cas, seulement une cellule sur 1728 est responsable de la détection de cet anévrisme, ce qui représente environ 0,057% du nombre total de cellules dans le patch. Afin de corriger cette disparité, nous avons utilisé une fonction de coût pondérée, inspirée du réseau YOLO (version 3) [238], dont les poids ont été déterminés grâce à une recherche exhaustive menée sur un ensemble de 20 patients d'entraînement et 10 patients de test issus de la base de données du CHRU de Nancy. La fonction de coût proposée intègre deux composantes : d'une part, elle intègre les scores de confiance associés aux cellules positives ou négatives, et d'autre part, elle englobe les paramètres de régression des sphères, applicables exclusivement aux cellules positives (voir Équation 5.2). Ces paramètres de régression englobent à la fois la localisation (C_x, C_y, C_z) et l'estimation de la taille (r) des sphères.

$$\text{Coût total} = \text{Classification} + \text{Régression} \tag{5.2}$$

Afin d'optimiser les scores de confiance associées à chaque détection, nous appliquons la fonction de coût d'entropie croisée binaire (BCE) pour les valeurs de confiance des cellules positives (P) et des

cellules négatives (N). La BCE quantifie l'erreur entre les probabilités prédites et les annotations binaires, mesurant ainsi la divergence entre les distributions de probabilité et pénalisant les prédictions incorrectes. Toutefois, cette fonction est très sensible au problème du déséquilibre de classes [279]. Afin de remédier au grand déséquilibre entre les cellules positives et négatives, une pondération est appliquée au terme relatif aux cellules négatives. Cette pondération est définie comme étant égale à la moitié du nombre de cellules positives ($\#P$) présentes dans le lot d'entraînement (voir Équation 5.3).

$$\text{Classification (P,N)} = \text{BCE}(P) + 0.5 \times \#P \times \text{BCE}(N) \tag{5.3}$$

Pour évaluer la précision de la localisation (coordonnées C_x, C_y, C_z) et de l'estimation de la taille (rayon r) des détections dans les cellules positives, nous utilisons la métrique d'erreur quadratique moyenne (MSE). La MSE, qui calcule la moyenne des carrés des différences entre les valeurs prédites et les valeurs réelles, est particulièrement adaptée aux problèmes de régression. Elle permet de maintenir la stabilité en cas de petites erreurs tout en imposant des pénalités significatives en présence de valeurs aberrantes. Cette caractéristique est essentielle pour améliorer la précision de la localisation et de l'estimation de la taille des petits objets [185]. Nous utilisons un poids de 5 pour équilibrer les contributions des deux termes - de confiance et de régression - dans le coût total (voir Équation 5.4).

$$\text{Régression (P)} = 5 \times \sum_{i=1}^{\#P} \text{MSE}(x_i, y_i, z_i, r_i) \tag{5.4}$$

5.3.2 Détails d'implementation

Notre méthode, DeepAneDet, a été implémentée en utilisant le langage de programmation Python (version 3.10.0) avec le framework PyTorch (version 1.10.0). Chaque image TOF-MRA a été normalisée en utilisant la méthode de normalisation z-score, avec une moyenne de 0 et un écart type de 1.

Phase d'entraînement

Concernant la phase d'entraînement, nous avons adopté une approche simple sans une optimisation intensive des hyper-paramètres. Nous avons utilisé l'optimiseur SGD avec un momentum de 0,98 pour le processus d'entraînement du modèle. Les paramètres choisis comprennent 200 époques d'entraînement et une taille de lot de 32 patchs. De plus, nous avons appliqué un planificateur de taux d'apprentissage personnalisé, commençant par une phase d'échauffement au cours de chaque itération des trois premières époques afin d'assurer une optimisation stable et d'éviter les perturbations initiales [109]. Par la suite, le taux d'apprentissage est resté constant à 10^{-2} jusqu'à la 50e époque, puis nous avons utilisé un planificateur de type polynomial [55] pour réduire progressivement le taux d'apprentissage jusqu'à atteindre 10^{-4} à la 200e époque. Les valeurs de ces hyper-paramètres ont été déterminées en utilisant un sous-ensemble de 30 patients issus de la base de données du CHRU de Nancy. Le modèle final, composé d'environ 28 millions de paramètres, nécessite environ 38 Go de mémoire GPU et a été exécuté sur une carte graphique NVIDIA RTX A6000 dotée de 48 Go de mémoire. Le processus d'entraînement de chaque modèle de notre méthode pour 200 époques dure environ 24 heures.

Phase de prédiction

Pendant la phase de prédiction, une approche de reconstruction de patchs a été employée pour générer des prédictions à partir du volume global, ce processus se déroulant en trois étapes majeures :

1. Le volume original a été subdivisé en patchs de dimensions $96 \times 96 \times 96$ voxels, avec une résolution de voxel de 0,4 mm^3 et un chevauchement de 32 voxels entre les patchs adjacents. Cette subdivision a été réalisée dans le but de réduire les artefacts de bords qui peuvent résulter des opérations de convolution sur des patchs de petite taille.
2. Les prédictions ont été générées à partir de chaque patch, puis reconstruites à la résolution d'origine. Pour qu'une prédiction soit considérée valide, le centre prédit (C_x, C_y, C_z) devait se situer dans la partie centrale de $64 \times 64 \times 64$ voxels du patch (voir une illustration en Figure 5.3).

3. Une étape de suppression non maximale (NMS) a été appliquée, avec un seuil d'intersection sur union (IoU) fixé à 10%. L'objectif de cette étape était d'éliminer les prédictions qui se chevauchent, en ne conservant que celle ayant le score de confiance le plus élevé.
4. Afin de réduire le taux de faux positifs, une technique d'augmentation du temps de test (TTA) a été mise en œuvre en effectuant des réflexions miroir des images le long des axes axial, sagittal et coronal. Les prédictions présentes dans l'image originale et toutes les images augmentées ont été conservées, en utilisant toujours avec un seuil IoU de 10%.

La prédiction de chaque volume global prend environ 30 secondes. Par ailleurs, notre code, les annotations et les divisions de validation croisée utilisées dans cette étude sont accessibles au public à l'adresse `https://gitlab.inria.fr/yassis/DeepAneDet`.

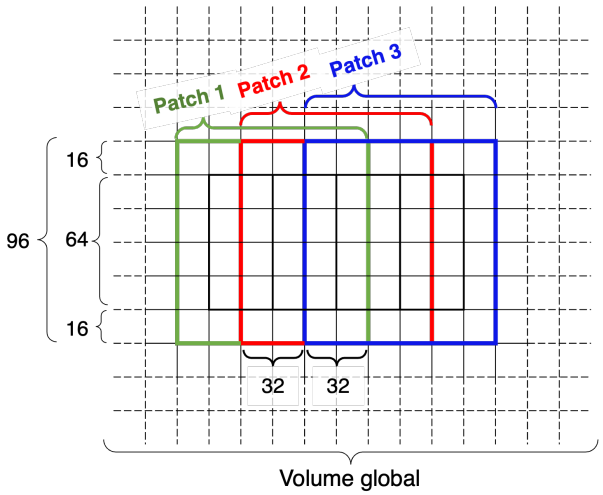


FIGURE 5.3 – Illustration de l'approche de prédiction par patchs sur le volume global d'une image TOF-MRA : division du volume en patchs de $96 \times 96 \times 96$ voxels avec un chevauchement de 32 voxels. Les prédictions sont extraites à partir de la partie centrale ($64 \times 64 \times 64$ voxels) de chaque patch.

5.3.3 Métriques d'évaluation et méthodes de référence

Métriques d'évaluation

Les méthodes de détection d'objets nécessitent l'utilisation de mesures d'évaluation qui tiennent compte à la fois de la précision de localisation des objets détectés et de la confiance accordée à chaque détection. Cependant, comme cela a été démontré dans le chapitre précédent, la plupart des méthodes de détection d'anévrismes existantes utilisent souvent des métriques peu adaptées à la tâche de détection d'objets, notamment par manque d'une estimation fiable du score de confiance associée à chaque détection. Pour une évaluation plus précise, nous avons opté pour l'utilisation de métriques basées sur un seuil d'intersection sur union (IoU) de 10% entre les prédictions et les vérités terrain (GT), où les anévrismes sont représentés sous forme de sphères [27, 192]. Cette représentation assure une évaluation plus précise des performances en matière de détection d'anévrismes, qui sont généralement de forme sphérique [137]. De plus, nous avons choisi un seuil IoU de 10% pour cette évaluation, contrairement à plusieurs travaux antérieurs qui utilisaient un seuil de 0%, notamment la compétition CADA [137].

Une sphère GT avec un IoU supérieur à 10% par rapport à une prédiction est classée comme un vrai positif (TP), sinon elle est considérée comme un faux négatif (FN). De plus, toute sphère prédite avec un IoU inférieur à 10% est identifiée comme un faux positif (FP). En cas de superposition de multiples détections sur la même GT, seule la détection avec le score de confiance le plus élevé est conservée, les autres étant considérées comme des faux positifs [221]. Les métriques d'évaluation que nous utilisons se concentrent principalement sur la métrique de la précision moyenne (AP), ce qui permet une analyse globale de la relation entre la sensibilité et la précision à différents seuils de classification. Nous examinons

la variabilité de la sensibilité et du nombre de faux positifs à travers divers seuils de classification en utilisant la courbe FROC. En complément, nous calculons le score de sensibilité et le nombre de faux positifs par cas (FP/cas) en utilisant un seuil de classification par défaut de 50%.

Méthodes de référence

Dans notre étude, nous avons évalué notre méthode de détection d'objets en la comparant à trois références dans le domaine de l'imagerie médicale 3D : nnUNet [135], nnDetection [27] et SCPM-Net [185]. nnUNet et nnDetection sont des frameworks auto-configurables qui sélectionnent automatiquement divers paramètres, allant des données d'entraînement au modèle d'apprentissage et à l'approche d'optimisation. Cette comparaison nous a permis d'évaluer l'efficacité de notre méthode en la mettant en perspective avec ces techniques qui ont déjà démontré leur performance dans de nombreux problèmes de détection.

nnUNet Le framework nnUNet est un système qui utilise des patchs contextuels de grande taille, sélectionnés automatiquement en fonction des données d'entraînement. Comme décrit dans le chapitre précédent, nnUNet est spécifiquement conçu pour la segmentation d'images médicales en 3D et repose sur une variante du réseau UNet [65]. Pour le comparer avec notre méthode, qui repose sur des représentations sphériques, nous avons adapté une sphère à chaque composante connexe (CC) de l'image prédite. Le diamètre de la sphère correspond à la distance maximale entre deux voxels à l'intérieur de la CC, et le score de confiance est déterminé comme étant la valeur maximale parmi ses voxels.

nnDetection De manière similaire à nnUNet, nnDetection exploite des patchs contextuels de grande taille, choisis de manière automatique en fonction des données d'entraînement. Basée sur le réseau RetinaUNet [139], cette approche de prédiction multi-échelle permet la détection d'objets à l'aide de boîtes englobantes définies par leurs coordonnées (C_x, C_y, C_z) et leurs dimensions (largeur, hauteur, profondeur). nnDetection est sur une approche de détection par ancres, utilisant 27 ancres pour chaque cellule. Pour le comparer avec notre méthode, nous avons converti ces boîtes englobantes en sphères. Cette transformation consistait à utiliser la diagonale de la boîte comme diamètre pour la sphère, et le centre de la sphère correspondait au centre de la boîte. Le score de confiance était identique à celui de la boîte englobante.

SCPM-Net Enfin, SCPM-Net est un modèle de détection d'objets en une seule étape et sans ancres. Il repose sur une architecture encodeur-décodeur et génère des prédictions à deux échelles, ce qui facilite la détection d'objets de tailles variées. SCPM-Net utilise des patchs de dimension $96 \times 96 \times 96$ voxels et détectent les objets en utilisant des sphères (C_x, C_y, C_z, rayon). L'entraînement de SCPM-Net consiste en 170 époques avec l'optimiseur SGD, un momentum de 0,9, une taille de lot de 24, et un taux d'apprentissage personnalisé qui inclut des phases d'échauffement suivies d'une réduction en fonction de la politique de plateau. Pour traiter le déséquilibre des classes, SCPM-Net met en œuvre une stratégie de correspondance de points centraux, sélectionnant les sept cellules les plus proches pour la détection de chaque objet. De plus, il utilise plusieurs techniques, telles que le Hard Negative Mining, pour renforcer l'apprentissage sur les exemples les plus difficiles. Il est à noter qu'en l'absence de détails sur la méthode de sélection des patchs d'entraînement dans la publication de SCPM-Net, nous avons basé l'entraînement de notre modèle SCPM-Net sur notre propre stratégie de gestion des données (voir section 5.2.2). Nous avons tenté d'incorporer différentes composantes de leur fonction de coût, y compris l'indice de similarité IoU entre les sphères. Cependant, l'intégration de cette fonction n'a pas permis la convergence du modèle. Après plusieurs investigations, nous avons exclu la composante IoU_{Sphere} de la fonction de coût finale (voir Équation 5.1), ce qui a conduit à une convergence réussie du modèle.

5.4 Étude expérimentale et résultats

Cette section présente les expérimentations menées pour évaluer l'efficacité de la méthode de détection d'anévrismes à partir d'images TOF-MRA. L'évaluation s'est déroulée en plusieurs étapes. Tout d'abord, une étude d'ablation a été réalisée sur la base de données du CHRU de Nancy pour évaluer l'impact de chaque composant de la méthode. Ensuite, une validation croisée en 5 plis a été effectuée en utilisant la base de données publique du CHUV de Lausanne [80]. Cette phase a permis d'examiner la robustesse et la capacité de généralisation de la méthode en la comparant aux méthodes de référence. Enfin, deux

radiologues experts ont mené une évaluation approfondie de ces différentes méthodes afin d'évaluer la pertinence des outils de détection automatique dans le contexte clinique.

5.4.1 Étude par ablation

L'étude par ablation a débuté avec l'entraînement de notre réseau de détection sans ancres, en utilisant des représentations sphériques (C_x, C_y, C_z, rayon), comme décrit dans la section 5.3.1. Les patchs d'entraînement ont été générés en suivant les techniques d'échantillonnage et de synthèse de données exposées dans la section 5.2.2. Ceci incluait la création de 50 duplications et de déformations pour chaque patch positif, ainsi qu'une première sélection de 40 patchs négatifs à partir de chaque image. À chaque époque d'entraînement, une sous-sélection des patchs négatifs a été réalisée pour maintenir 15% du nombre total de patchs d'entraînement. Les prédictions finales ont été obtenues en utilisant une méthode de reconstruction par patchs, combinée à l'algorithme NMS et à la technique d'augmentation TTA, qui comprenait l'application d'effets miroir le long des trois axes anatomiques, tel que détaillé dans la section 5.2.2. Dans ce qui suit, le modèle résultant est désigné sous le nom de *Modèle0*.

Pour analyser l'impact de chaque composante de notre méthode de base (*Modèle0*), nous avons développé et comparé les 8 variantes suivantes :

— *Modèle1* : Entraîné sans l'application de déformations aléatoires sur les patchs positifs, ce modèle vise à évaluer l'impact de la synthèse de données sur la détection des anévrismes.

— *Modèle2* : L'entraînement de ce modèle se limite à l'utilisation exclusive de patchs positifs, ce qui permet d'évaluer l'influence des échantillons négatifs dans le processus d'apprentissage pour la discrimination des anévrismes.

— *Modèle3* : Fait usage de tous les patchs négatifs pendant chaque époque d'entraînement, augmentant leur proportion à 34% du total des patchs, par rapport au taux de 15% établi dans notre méthode.

— *Modèle4* : Utilise des annotations sphériques surdimensionnées (33% plus grandes) pour examiner l'influence de la précision des annotations sur les performances de détection, en simulant les annotations fournies avec les données du CHUV de Lausanne (voir Figure 5.1a).

— *Modèle5* : Entraîne notre réseau à détecter les anévrismes en utilisant des boîtes englobantes (C_x, C_y, C_z, largeur, hauteur, profondeur) positionnées sur les centres des sphères. Les dimensions de ces boîtes ont été ajustées pour correspondre aux dimensions des anévrismes dans les trois directions spatiales. Pour effectuer cet ajustement, nous avons développé un plugin du logiciel 3D Slicer [92] spécifique qui permet la visualisation et la modification des dimensions des boîtes englobantes.

— *Modèle6* : De manière similaire à nnDetection [27], ce modèle se base sur une approche de détection par ancres. Au total, 6 ancres prédéfinies sont utilisées par cellule, et leurs diamètres sont déterminés à l'aide de l'algorithme K-means appliqué aux données d'entraînement (avec des tailles de 2,31, 3,74, 5,49, 7,83, 13,76, et 19,63 mm).

— *Modèle7* : Afin d'évaluer l'efficacité de notre fonction de coût en présence de déséquilibre de classes, ce modèle n'adopte pas une approche où une seule cellule est utilisée pour détecter chaque anévrisme. Au lieu de cela, il fait appel aux 10 cellules les plus proches du centre de l'anévrisme. Cette approche est similaire à celle du modèle SCPMNet [185], qui utilise 7 cellules par objet.

— *Modèle8* : Pour évaluer son influence sur les performances de détection, ce modèle élimine la technique d'augmentation en temps de test (TTA).

Comme dans le chapitre précédent (voir section 4.5.2), la base de données du CHRU de Nancy a été divisée en deux ensembles distincts : 92 images pour l'ensemble d'entraînement et 40 pour l'ensemble de test. L'ensemble d'entraînement comptait 138 anévrismes avec une taille moyenne de 4,11 $\pm$ 2,48 mm, tandis que l'ensemble de test en contenait 68, avec une taille moyenne de 3,67 $\pm$ 1,93 mm. Tous les modèles ont été soumis à un entraînement de 200 époques en utilisant les données de l'ensemble d'entraînement. Pour notre méthode, cela a conduit à la création de 8096 patchs d'entraînement, parmi lesquels 1196 étaient des patchs négatifs. Les performances finales ont été évaluées sur l'ensemble de test en utilisant des métriques basées sur un seuil de IoU de 10%, tel que détaillé dans la section 5.3.3. Les résultats de nos expériences sont présentés dans la table 5.2.

	AP (%)	Sensibilité (%)	FP/cas
Modèle0	80,24	77,94	0,40
Modèle1	74,69	72,06	0,35
Modèle2	71,72	70,59	0,57
Modèle3	77,66	77,94	0,60
Modèle4	69,59	70,59	0,37
Modèle5	72,47	75,00	0,72
Modèle6	81,64	70,59	0,37
Modèle7	81,95	79,41	0,64
Modèle8	78,97	80,88	0,61

TABLE 5.2 – Étude d'ablation de notre méthode : Expérimentations menées sur la base de données du CHRU de Nancy. Les métriques d'évaluation sont calculées en utilisant un seuil IoU de 10%.

L'étude de différentes variantes du modèle a révélé l'influence de certains composants sur la performance de détection. Pour *Modèle1*, l'omission de la distorsion aléatoire appliquée aux patchs positifs a engendré un problème de sous-confiance, ce qui s'est traduit par des scores de confiance de détection plus faibles. Cette observation souligne l'importance de synthétiser une diversité de formes d'anévrismes pour améliorer la capacité des modèles à apprendre des caractéristiques liées à leur morphologie variable. À la différence des détecteurs d'objets 2D classiques, qui se concentrent principalement sur les images positives, *Modèle2* a mis en évidence l'utilité des patchs négatifs dans le contexte de la détection d'anévrismes en 3D. Plus précisément, l'inclusion de ces patchs négatifs a contribué à une meilleure modélisation du fond, renforçant ainsi les performances globales de la méthode. *Modèle3* a montré que cette modélisation du fond était associée à une sensibilité accrue. Cependant, un excès de patchs négatifs a eu un impact négatif sur la modélisation des anévrismes, entraînant une augmentation des faux positifs et une réduction globale de la confiance de détection.

L'utilisation de sphères sur-dimensionnées pour entraîner le *Modèle4* a montré un effet négatif sur les performances de détection. En revanche, l'utilisation de sphères proportionnelles aux dimensions réelles des anévrismes a amélioré leur capacité discriminative par rapport aux vaisseaux sains. Le *Modèle5* a confirmé les avantages de la représentation sphérique par rapport aux boîtes englobantes classiques dans la détection des anévrismes, qui sont des objets de forme sphérique. En particulier, les boîtes englobantes ont tendance à générer plus de faux positifs. Le *Modèle6* a révélé que l'utilisation d'ancres n'était pas nécessaire pour la détection des anévrismes avec des représentations sphériques, car le rayon pouvait être efficacement optimisé pendant l'apprentissage. De plus, l'utilisation d'ancres a exacerbé le déséquilibre entre les ancres positives et négatives, entraînant des scores de confiance plus faibles. Notre approche, basée sur une seule cellule par anévrisme, a produit des résultats compétitifs par rapport au *Modèle7*, qui utilisait 10 cellules pour chaque anévrisme. Cependant, le *Modèle7* avait tendance à générer davantage de petites détections près de la localisation de l'anévrisme, conduisant à plus de faux positifs. Enfin, l'élimination de la technique d'augmentation TTA dans le *Modèle8* a conduit à une augmentation des faux positifs. Cette technique a permis de réduire la variabilité du modèle tout en maintenant un taux élevé de sensibilité.

5.4.2 Évaluation par validation croisée

Ensemble de données privées du CHRU de Nancy

Pour évaluer notre méthode (*Modèle0*) sur un ensemble de données plus vaste, nous avons réalisé une validation croisée à 5 plis sur la base de données du CHRU de Nancy. Les données ont été divisées en 5 sous-ensembles, dont 4 ont été utilisés pour l'entraînement (composés de 105 ou 106 images chacun) et un pour les tests. Les anévrismes dans chaque sous-ensemble avaient des diamètres moyens respectifs de 4,15 mm, 3,61 mm, 3,88 mm, 3,85 mm et 4,25 mm, avec des nombres d'anévrismes de 47, 32, 45, 38 et 44. Notre méthode a été comparée à trois méthodes de référence décrites dans la section 5.3.3 : SCPM-Net [185], ainsi que les deux frameworks auto-configurables nnDetection [27] et nnUNet [135]. Les dimensions des patchs d'entraînement diffèrent en fonction des méthodes. Notre méthode et SCPM-Net utilisent des

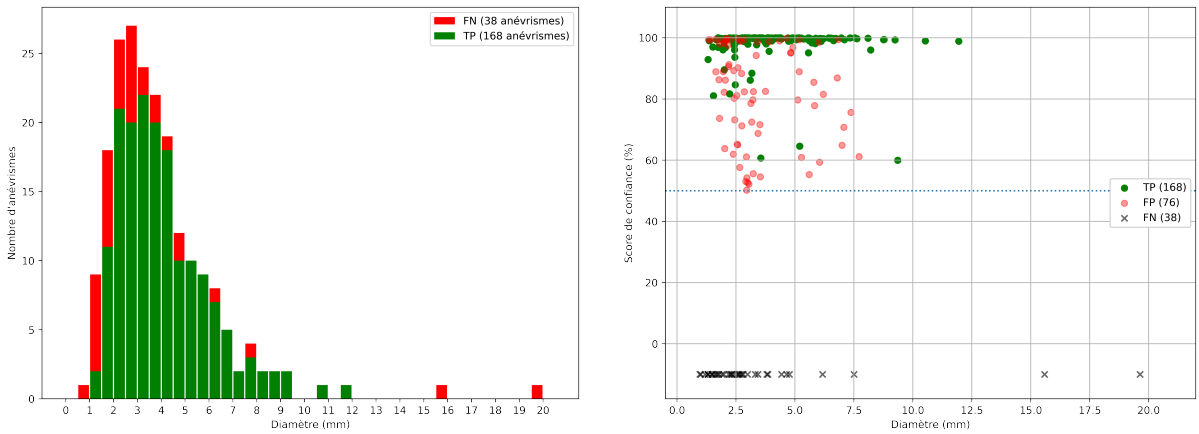
patchs de dimention fixé à $96 \times 96 \times 96$ voxels, tandis que nnDetection et nnUNet ont automatiquement adopté des patchs de taille $160 \times 96 \times 192$ voxels et $92 \times 128 \times 96$ voxels, respectivement. Nous rappelons que le modèle SCPM-Net utilise notre stratégie de génération de données pour l'entraînement. Les résultats, ainsi que les scores de variabilité (écart-type) calculés pour les différents plis de test, sont présentés dans la table 5.3.

	AP (%)	**Sensibilité (%)**	**FP/cas**
DeepAneDet	81,22 ± 5.34	81,55 ± 5,40	0,57 ± 0,06
SCPM-Net [185]	80,14 ± 6,49	82,03 ± 7,50	0,71 ± 0,14
nnDetection [27]	80,88 ± 6.33	77,15 ± 8,84	0,68 ± 0,13
nnUNet [135]	79,45 ± 4,70	76,04 ± 4,19	0,27 ± 0,10

TABLE 5.3 – Validation croisée à 5 plis sur la base du CHRU de Nancy : Comparaison avec des méthodes de référence. DeepAneDet est notre méthode proposée. Les scores de variabilité sont obtenus à partir des différents sous-ensembles de test.

Les expériences révèlent les performances concurrentielles de notre approche, notamment en ce qui concerne la faible variance entre les instances de données. Notre méthode a obtenu un score de précision moyenne (AP) de 81,22%, surpassant nnDetection avec 80,88%, SCPM-Net avec 80,14%, et nnUNet avec 79,45%. À un seuil de classification de 50%, elle a montré une sensibilité de 81,55% avec 0,57 FP/cas, tandis que SCPM-Net a généré davantage de faux positifs (0,71 FP/cas). Il est à noter que, bien que la différence entre les scores de précision moyenne et de sensibilité ne soit pas significative, elle l'est en revanche en ce qui concerne le taux de FP/cas.

La Figure 5.4a présente les anévrismes détectés et les faux négatifs classés en fonction de la taille des anévrismes. Pour les anévrismes de moins de 3mm, notre méthode a démontré une sensibilité de 66,66% (vs 64,19% pour nnDetection, 65,43% pour SCPM-Net, et 51,85% pour nnUNet). Pour les anévrismes de taille comprise entre 3 et 5mm, notre méthode a atteint une sensibilité de 90,90% (vs 93,50% pour nnDetection, 94,80% pour SCPM-Net, et 90,90% pour nnUNet). Pour les grands anévrismes, notre méthode a montré une sensibilité de 91,66% (vs 93,75% pour nnDetection, 89,58% pour SCPM-Net, et 93,75% pour nnUNet). Par ailleurs, la Figure 5.4b illustre les scores de confiance des prédictions de notre méthode en relation avec les diamètres des anévrismes. À noter que seules les détections avec un score de confiance supérieur à 50% sont affichées.



(a) (b)

FIGURE 5.4 – Performance de notre méthode sur la base de données CHRU de Nancy : (a) Sensibilité de détection par taille d'anévrismes. (b) Scores de confiance des détections.

Ensemble de données publiques du CHUV de Lausanne

Afin d'évaluer la capacité de généralisation de notre méthode, nous avons effectué une validation croisée en 5 plis sur la base de données publique du CHUV de Lausanne [80]. Nous rappelons que cet ensemble se compose de 270 images, comprenant un total de 164 anévrismes, provenant de divers équipements médicaux. Cette diversité reflète une représentation des conditions cliniques réelles. Les données ont été réparties aléatoirement en 5 sous-ensembles, chacun contenant un nombre équilibré d'images avec et sans anévrismes, totalisant 54 images par sous-ensemble. Le nombre d'anévrismes varie entre 32, 33, 43, 28 et 28 par pli, avec des tailles moyennes d'anévrisme respectives de 3,56 mm, 3,19 mm, 4,05 mm, 4,21 mm et 3,70 mm. Chaque pli a été utilisé pour l'entraînement de 5 modèles, qui ont utilisé 4 sous-ensembles pour l'entraînement et un pour les tests. En moyenne, chaque modèle de notre méthode a utilisé 7875 patchs d'entraînement, comprenant 1075 patchs négatifs. Notre méthode a été comparée à trois méthodes de référence : SCPM-Net, nnDetection et nnUNet. Il convient de noter que nnDetection et nnUNet ont utilisé des patchs de dimension $64 \times 224 \times 192$ voxels et $64 \times 224 \times 160$ voxels, respectivement. Chaque modèle nécessite des durées d'entraînement distinctes, avec environ 24 heures pour notre méthode, 32 heures pour SCPM-Net, 35 heures pour nnDetection et 24 heures pour nnUNet.

La table 5.4 (en haut) détaille les résultats obtenus par les différentes méthodes sur l'ensemble de la base de données, qui comprend 270 images. Ils incluent également les valeurs d'écart-type pour refléter la variabilité des performances sur les différents plis de la validation croisée. Notre méthode s'est distinguée par une précision moyenne de 75,80%, supérieure aux autres méthodes, et une sensibilité de 84,76% associée avec un taux de faux positifs de 0,57 FP/cas. En comparaison, SCPM-Net et nnDetection ont enregistré des taux de faux positifs plus élevés, respectivement à 0,75 et 0,67 FP/cas. nnUNet, quant à lui, a présenté un taux inférieur de faux positifs, mais sa sensibilité était moindre, s'établissant à 71,95%.

Annotations de test	Méthodes	AP (%)	Sensibilité (%)	FP/case
Initiales	DeepAneDet	75,80 ± 7,11	84,76 ± 5,00	0,57 ± 0,17
	SCPM-Net	72,41 ± 7,89	85,36 ± 6,18	0,75 ± 0,19
	nnDetection	73,68 ± 6.38	84.76 ± 4,72	0,67 ± 0,12
	nnUNet	72,46 ± 4,74	71,95 ± 9,11	0,13 ± 0,06
Révisées	DeepAneDet	78,96 ± 5,24	86,78 ± 3,98	0,53 ± 0,14
	SCPM-Net	76,35 ± 5,91	86,20 ± 5,22	0,71 ± 0,15
	nnDetection	78,01 ± 6,07	86,78 ± 6,16	0.63 ± 0,10
	nnUNet	73,47 ± 4,17	71,26 ± 8,09	0,11 ± 0,04
	DeepAneDet*	69,33 ± 6.64	84,48 ± 3.93	0,45 ± 0,05
	SCPM-Net*	64,87 ± 7,14	83,90 ± 5.47	0.64 ± 0,12
	nnDetection*	46,26 ± 11,80	64,36 ± 28,63	0,34 ± 0,28
	nnUNet*	66.67 ± 4,27	69,54 ± 5,64	0,06 ± 0,02

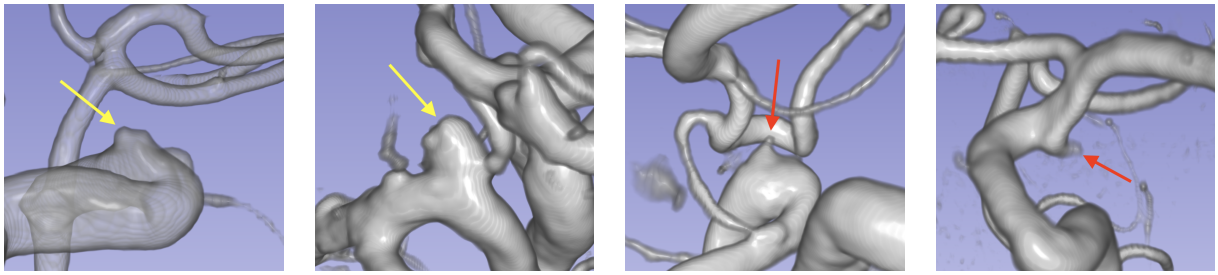
TABLE 5.4 – Validation croisée à 5 plis (moyenne ± écart-type) sur l'ensemble de données CHUV. Chaque méthode a été entraînée avec les annotations initiales et testée avec ces mêmes annotations (en haut), ainsi qu'avec des annotations révisées (au milieu et en bas). * indique les méthodes entraînées uniquement sur des images présentant des anévrismes et testés sur l'ensemble des images, y compris celles sans anévrismes.

5.4.3 Évaluation par experts et consolidation des annotations

Après un examen visuel des détections générées par les différentes méthodes, il est apparu que de nombreux anévrismes non annotés ont été identifiés, soulevant ainsi des préoccupations quant à la qualité des annotations dans l'ensemble de données du CHUV de Lausanne [80]. Pour remédier à cette situation, une nouvelle vérité terrain plus fiable a été établie. Deux experts en neuroradiologie interventionnelle, LL et RA, avec respectivement 14 et 32 années d'expérience, ont examiné les détections de notre méthode et de nnDetection, y compris les cas de faux négatifs (anévrismes non détectés). Le choix de ces deux méthodes a été motivé par leurs performances satisfaisantes, tout en prenant en compte la contrainte de temps pour mener cette étude. Afin de faciliter cette évaluation, un module du logiciel 3D Slicer [92] a été développé et utilisé par les experts (voir Annexe A). Ce module leur a permis d'évaluer chaque détection

de manière individuelle (vrai, faux ou incertain) et d'ajouter de nouvelles annotations en cas de besoin. Cette étude a duré environ 6 mois pour collecter les résultats. Chaque méthode évaluée a donné lieu à deux ensembles distincts d'annotations par expert. L'ensemble des annotations ont été analysées, et les décisions finales ont été prises par consensus majoritaire pour chaque cas, avec 7,6% des annotations faisant l'objet de désaccords. Ces derniers cas ont été résolus à travers une revue collective, aboutissant à un consensus.

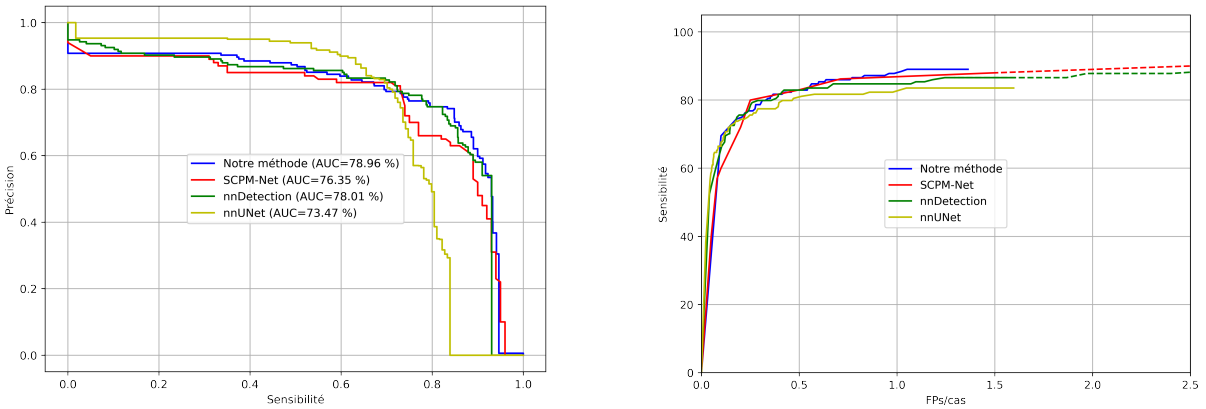
Cette vérification a révélé que 23 faux positifs étaient en réalité des anévrismes qui n'avaient pas été annotés (voir deux exemples dans les Figures 5.5a et 5.5b), et que 13 annotations initiales correspondaient en réalité à de simples irrégularités vasculaires (voir deux exemples dans les Figures 5.5c et 5.5d). À noter que aucun nouvel anévrisme n'a été annoté par les experts. Ainsi, les annotations initiales ont été mises à jour, faisant passer le nombre total d'anévrismes de 164 à 174, avec une taille moyenne de 3,65 $\pm$ 2,15 mm. Parmi ces anévrismes, 79 avaient une taille inférieure à 3 mm, tandis que 71 se situaient entre 3 et 5 mm. Les performances des méthodes ont été réévaluées en utilisant ces annotations révisées, et les résultats sont présentés dans la table 5.4 (au milieu).



(a) (b) (c) (d)

FIGURE 5.5 – Exemples d'erreurs d'annotations dans la base de données CHUV [80] : (a, b) anévrismes non annotés, (c, d) annotations incorrectes (ici, départ de petites artères).

La réévaluation des détections générées par les quatre méthodes en utilisant les annotations révisées a révélé des améliorations significatives dans leurs performances. Plus particulièrement, une réduction systématique de l'écart-type a été observée, mettant en évidence leur aptitude à apprendre les caractéristiques distinctives des anévrismes, même en présence de bruit d'annotation. Ainsi, notre méthode a atteint une précision moyenne de 78,96% et une sensibilité de 86,78%, avec 0,53 FP/cas. Ces résultats témoignent de performances comparables, voire supérieures, à celles des autres méthodes de référence.



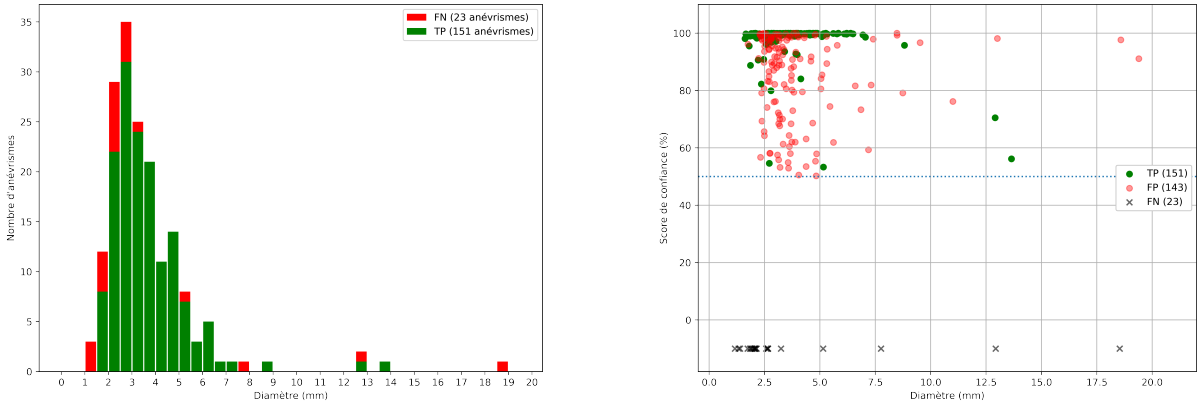
(a) Courbe Précision-Sensibilité. (b) Courbe FROC.

FIGURE 5.6 – Évaluation comparative des performances de quatre techniques sur la base de données CHUV de Lausanne, avec annotations révisées. Notre méthode est DeepAneDet.

Pour illustrer la qualité du compromis entre les vrais positifs (TP) et les faux positifs (FP), des courbes de précision-sensibilité ont été générées pour les quatre méthodes (voir Figure 5.6a). De plus, les courbes FROC ont été tracées en ajustant le seuil de classification (voir Figure 5.6b).

Notre méthode a démontré un bon équilibre entre la sensibilité et le nombre de faux positifs, notamment avec un maximum de 1,31 FP/cas associé à une sensibilité élevée de 91,82%. En revanche, nnUNet a atteint une sensibilité maximale de 83,90% avec un nombre plus élevé de faux positifs (1,56 FP/cas), tandis que nnDetection a obtenu une sensibilité maximale de 94,82%, mais avec un nombre significativement plus élevé de faux positifs (24,42 FP/cas). De même, SCPM-Net a atteint une sensibilité maximale de 93.10%, mais avec un nombre plus élevé de faux positifs (14,07 FP/cas). Ces résultats illustrent la capacité de notre méthode à maintenir une sensibilité élevée tout en limitant le nombre de faux positifs.

Nous avons également procédé à une analyse des performances des méthodes en fonction de la taille des anévrismes (voir Figure 5.7a). Pour les anévrismes de diamètre inférieur à 3 mm, notre méthode a atteint une sensibilité de 77,21%, tandis que SCPM-Net, nnDetection et nnUNet ont obtenu respectivement des sensibilités de 75,94%, 73,42% et 45,57%. Dans la plage de diamètres comprise entre 3 et 5 mm, notre méthode a enregistré une sensibilité de 98,59%, comparée à 97,18%, 98,59% et 92,96% pour SCPM-Net, nnDetection et nnUNet, respectivement. En ce qui concerne les anévrismes de grande taille, notre méthode a démontré une sensibilité de 83,33%, tandis que SCPM-Net, nnDetection et nnUNet ont obtenu des sensibilités de 87,5%, 95,83% et 91,67%, respectivement. Notre méthode surpasse donc les autres sur les anévrismes petits à moyen (< 5 mm). Elle est moins performante sur les gros anévrismes, mais il convient de noter que les radiologues présentent une excellente sensibilité de détection pour ces gros anévrismes. Figure 5.7b présente les scores de confiance associés aux prédictions de notre méthode.



(a) (b)

FIGURE 5.7 – Performance de notre méthode sur la base CHUV de Lausanne en utilisant les annotations révisées : (a) sensibilité de détection par taille d'anévrismes, et (b) score de confiance par détections.

5.4.4 Influence des petits patchs sur des données limitées

L'un des avantages de l'utilisation d'une approche basée sur de petits patchs dans le traitement de données limitées réside dans la possibilité de générer de nombreux patchs négatifs. Cette caractéristique permet une modélisation efficace du fond, même lors de l'utilisation d'images provenant uniquement de patients porteurs d'une pathologie. Pour confirmer cette hypothèse, une étude de validation croisée à 5 plis a été réalisée en utilisant l'ensemble de données du CHUV de Lausanne, composé de 140 images avec anévrismes et 130 sans anévrismes. Les 140 images avec anévrismes ont été reparties de manière aléatoire en 5 sous-ensembles, chacun comprenant 28 images, ce qui a permis l'utilisation de 112 images pour l'entraînement. L'ensemble de test a été formé des 28 images restantes, auxquelles nous avons ajouté 130 patients sains pour l'évaluation finale. Pour chaque pli de cette étude, notre méthode a été entraînée sur un ensemble de 7710 patchs, parmi lesquels 1110 étaient des patchs négatifs. Les annotations initiales ont été utilisées lors de la phase d'entraînement, tandis que les annotations révisées ont été utilisées pour

l'évaluation finale. Les résultats sont présentés dans la table 5.4 (en bas).

Notre méthode proposée a maintenu une performance similaire à celle obtenue avec l'ensemble de données contenant des patients en bonne santé. Une diminution de la précision moyenne a été constatée, passant de 78,96 à 69,33%, tout en étant accompagnée d'une amélioration du taux de faux positifs par cas, de 0,57 à 0,45. Des résultats comparables ont été observés pour la méthode SCPM-Net, dont nous rappelons qu'elle a été adaptée pour utiliser notre stratégie de gestion de données avec des petits patchs de dimensions $96\times96\times96$ voxels. En revanche, les méthodes basées sur des patchs de plus grandes dimensions, telles que nnUNet et nnDetection, ont connu une diminution significative des scores de précision moyenne et de sensibilité. Cette caractéristique de notre stratégie de gestion des données se révèle particulièrement pertinente pour l'exploitation de bases de données cliniques qui sont souvent dépourvues ou, du moins, pauvres en sujets sains.

5.5 Discussion et analyse qualitative

Malgré les avancées récentes prometteuses, la détection automatique des anévrismes intracrâniens non rompus demeure un défi. Les méthodes existantes sont difficiles à évaluer et à comparer en raison du manque de bases de données publiques communes et de métriques d'évaluation appropriées. Notre contribution consiste en un détecteur d'objets 3D en une seule étape et sans ancres, dédié à la détection des anévrismes, incluant leur localisation, leur taille et des scores de confiance. Notre approche repose sur l'utilisation de représentations sphériques, simplifiant l'annotation des données et offrant une évaluation plus précise. De plus, nous avons opté pour de petits patchs non intersectés, considérés comme quasi-indépendants. Ils permettent d'augmenter les échantillons d'entraînement, notamment dans des cas où les ensembles de données peuvent être limités, voire ne contenir aucun patient en bonne santé (sans anévrisme) ou seulement quelques-uns. Nous avons développé et évalué notre méthode sur un ensemble de données privé, puis l'avons testée sur un vaste ensemble de données public.

L'évaluation a utilisé des métriques spécifiques aux tâches de détection d'objets et a inclus une révision de chaque détection par deux experts cliniques. Nos résultats indiquent que notre approche excelle par rapport à nnUNet et affiche des performances comparables, voire supérieures, à nnDetection et SCPM-Net, avec une précision moyenne de 78,96% et une sensibilité de 86,78%, tout en générant 0,53 faux positifs par cas. Néanmoins, ces résultats prometteurs nécessitent une confirmation ultérieure, les écarts-types révélés par la validation croisée n'étant pas suffisants pour établir une différence statistiquement significative. Des évaluations complémentaires sur des ensembles de données plus vastes sont requises.

5.5.1 Architectures de réseau et déséquilibre de classes

Architectures de réseau Le réseau YOLO [238] est couramment privilégié pour la détection d'objets en imagerie naturelle en 2D, en particulier dans les applications qui requièrent un compromis entre la détection en temps réel et la précision [176]. Notre approche repose également sur une architecture CNN simple, avec un encodeur, spécialement adaptée aux données en voxels 3D, ce qui lui permet d'obtenir des performances très compétitives par rapport aux architectures plus complexes de type encodeur-décodeur, telles que RetinaUNet [139], utilisée par le framework nnDetection, ainsi que celle employée par le modèle SCPM-Net [185]. De plus, notre architecture évite l'utilisation de larges patchs contextuels, qui sont exigeants en termes de ressources de calcul et de données. En particulier, au cours de nos expériences de validation croisée en 5 plis, l'apprentissage et l'inférence de chaque modèle ont nécessité environ 35 heures pour nnDetection, 28 heures pour nnUNet, et 24 heures pour notre méthode. In est important de notre que SCPM-Net a été entraîné en utilisant notre stratégie de sélection et d'échantillonnage de données, car leur article ne spécifiait pas leur propre stratégie de sélection des données d'entraînement, et l'utilisation de patchs aléatoires ne permettait pas la convergence de l'apprentissage. Nous avons également démontré que l'emploi d'une approche de détection utilisant des ancres n'est pas indispensable pour la détection des anévrismes par des sphères (*Modèle0* vs *Modèle6*). Les méthodes s'appuyant sur les ancres, comme nnDetection, nécessitent des configurations spécifiques (nombre, dimension, emplacement) des ancres, qui doivent être ajustées en fonction des caractéristiques des objets dans les données. De plus, ces

méthodes demandent des stratégies d'échantillonnage adaptées afin de pallier le déséquilibre de classes, particulièrement marqué entre les ancres positives et négatives [323].

Problème du déséquilibre des classes Pour résoudre le déséquilibre des classes, nous avons adopté une approche en deux volets impliquant la gestion des données et la conception d'une fonction de coût appropriée. La gestion des données visait à préparer les données d'entraînement pour prendre en compte la rareté des anévrismes. Pour les patchs positifs, nous avons utilisé des techniques d'augmentation de données (*Modèle0* vs *Modèle1*) pour diversifier notre ensemble de données en introduisant diverses formes d'anévrismes. En revanche, pour les patchs négatifs, nous avons mis en place des stratégies d'échantillonnage spécifiques (*Modèle0* vs *Modèle2* et *Modèle3*) pour améliorer l'apprentissage des caractéristiques statistiques de l'arrière-plan. Cette approche a renforcé la discrimination entre les anévrismes, les vaisseaux sanguins sains et les tissus cérébraux. Simultanément, nous avons utilisé une fonction de coût visant à atténuer le déséquilibre significatif entre les cellules positives et négatives au sein des patchs. Malgré l'utilisation d'une seule cellule positive par anévrisme (représentant seulement 0,057% lorsque seul un anévrisme était présent dans le patch), notre fonction de coût s'est efficacement concentrée sur cette minorité de cellules, obtenant ainsi des résultats comparables, voire supérieurs, à ceux obtenus par le *Modèle7*, qui utilisait 10 cellules par anévrisme. Par ailleurs, notre fonction de coût est plus simple que celle utilisé dans le modèle SCPM-Net, qui implique une optimisation de l'IoU entre les sphères pendant l'entraînement, et fait également appel à la technique d'optimisation "Hard Negative Mining".

5.5.2 Représentations sphériques et analyse des résultats

Représentations sphériques Nos expériences ont démontré que la représentation sphérique s'avère plus appropriée pour la détection des anévrismes par rapport à la représentation classique des boîtes englobantes (*Modèle5*). En effet, les sphères nécessitent uniquement 4 paramètres (C_x, C_y, C_z, rayon), tandis que les boîtes requièrent 6 paramètres (C_x, C_y, C_z, largeur, hauteur, profondeur). L'optimisation d'un seul paramètre, le rayon, lors de l'apprentissage, s'avère moins complexe que l'apprentissage de la combinaison de la largeur, de la hauteur et de la profondeur des boîtes pour s'ajuster au volume de l'anévrisme. De plus, les sphères présentent l'avantage de bénéficier de l'invariance par rapport à la rotation, comme illustré dans la Figure 5.8, ce qui les rend plus adaptées à l'augmentation des données en utilisant des rotations. Par ailleurs, l'utilisation de sphères bien ajustées pour l'annotation des anévrismes s'est révélée plus efficace que l'adoption de sphères sur-dimensionnées, comme celles utilisées dans l'ensemble de données CHUV de Lausanne (voir Figure 5.1a, *Modèle0* vs *Modèle4*). En outre, la représentation par sphères permet également une évaluation plus précise des performances de détection en utilisant la métrique IoU, étant donné que les anévrismes présentent généralement une forme sphérique [137].

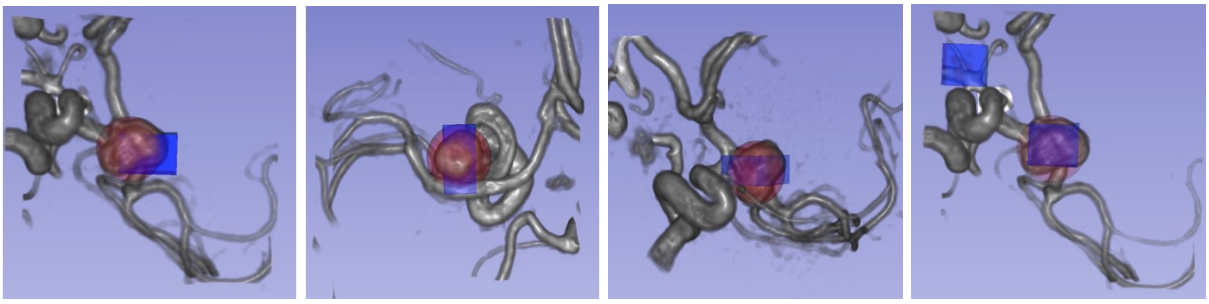


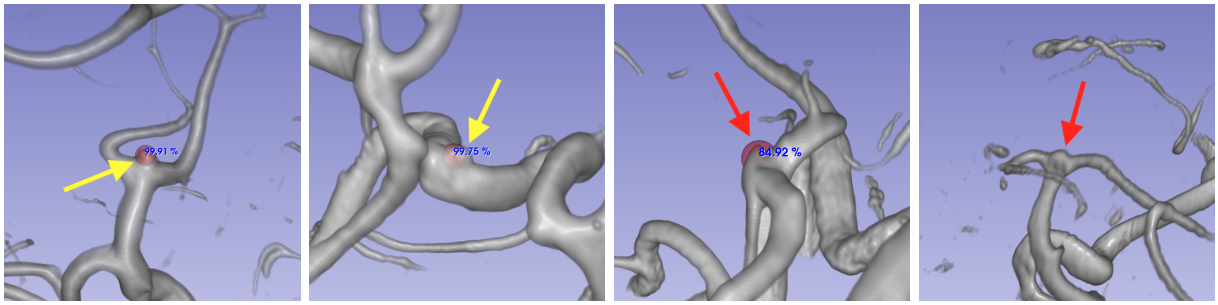
FIGURE 5.8 – Comparaison des détections : notre méthode (sphères en rouge) vs *Modèle5* (boîtes en bleu). Les boîtes montrent une instabilité de taille lors des rotations, entraînant un faux positif (à droite).

Analyse qualitative des résultats Comme illustré dans la Figure 5.7a, notre méthode rencontre encore des difficultés pour détecter les anévrismes de grande taille ($>$ 10 mm). Nous reconnaissons que l'approche basée sur de petits patchs présente des limites dans la détection des anévrismes de grande taille. De plus, notre ensemble de données ne comptait que 4 anévrismes de plus de 10 mm, ce qui constitue un échantillon restreint. Cependant, nous soutenons que la détection des gros anévrismes ne

présente pas de défi clinique majeur, et que les méthodes de détection automatique devraient être plus efficaces pour les petits anévrismes. En effet, les faux positifs les plus difficiles à distinguer sont souvent localisés aux départs des petites artères, proches de la limite de résolution, se confondent avec de petits anévrismes (voir des exemples en Figures 5.9a et 5.9b). Sur les 174 anévrismes de l'ensemble de données du CHUV [80], 14 des 23 anévrismes non détectés (FN) avaient un diamètre inférieur à 2,5 mm (voir un exemple en Figure 5.9d). Certaines de ces erreurs peuvent être attribuées à l'incertitude d'annotation de ces anévrismes, qui sont difficiles à diagnostiquer en utilisant l'imagerie TOF-MRA [140].

Nous avons évalué la précision de la localisation des anévrismes ainsi que de leurs diamètres par rapport aux annotations. Pour la localisation du centre, notre méthode a montré une distance euclidienne moyenne de 0,70 ± 0,51 mm, avec une erreur moyenne absolue (MAE) de 0,40 mm et une erreur quadratique moyenne (MSE) de 0,31 mm. Pour l'estimation des diamètres des anévrismes, notre méthode a obtenu une erreur MAE de 0,66 mm et une erreur MSE de 0,91 mm. Nous rappelons que les IRM de la base de données utilisée présentent des tailles voxels entre 0,4 et 0,7 mm. Par ailleurs, nous avons également évalué les scores IoU entre les anévrismes détectés et les annotations, obtenant une valeur moyenne de 44,32 ± 15,70%, avec une valeur minimale de 16,95% et une valeur maximale de 88,85%.

Dans notre étude, les méthodes ont été entraînées en utilisant les annotations initiales. L'évaluation finale a été effectuée en utilisant les annotations révisées, plus fiables, qui avaient été validées par deux experts cliniques. Cette nouvelle vérité terrain a permis d'évaluer les annotations initiales fournies avec l'ensemble de données [80]. Les résultats révèlent une sensibilité de 82,7% et seulement 0,07 faux positifs par cas. Ces résultats correspondent à ceux rapportés dans la littérature, démontrant que les experts ont une bonne sensibilité et une excellente spécificité. Bien entraînés sur des données bruités, notre méthode surpasse donc les performances des experts en sensibilité (86,78 ± 3,98% pour notre méthode). En revanche, le taux de FP/cas, même si reste cliniquement acceptable (0,53 ± 0,14), reste nettement plus élevé que celui des experts. Améliorer la spécificité des méthodes automatique reste donc un défi qui nécessite des effets supplémentaire de recherche.



(a) (b) (c) (d)

FIGURE 5.9 – Sphères de détection (rouge) et scores de confiance (bleu) : (a, b) exemples de petits anévrismes détectés par notre méthode (taille : 1,90 et 1,65 mm), (c) exemple de faux positif, et (d) anévrisme non détecté de taille 1,91 mm.

Conclusion

Dans ce chapitre, nous avons introduit une méthode de détection d'anévrismes à partir d'images TOF-MRA en une seule étape et utilisant des représentations sphériques. Notre approche se distingue par sa simplicité et son efficacité, s'inspirant du réseau de détection YOLO 2D, sans recourir aux ancres traditionnelles. Nous avons également utilisé une stratégie de données basée sur de petits patchs, ce qui a été également un élément important de notre méthode. Lors de notre évaluation approfondie, menée par deux radiologues experts sur une grande base de données publique, nous avons mis en évidence un problème majeur lié à l'incertitude des annotations dans les bases de données existantes. Ce problème représente un défi non seulement pour l'entraînement des réseaux de neurones avec des données limitées, mais aussi pour leur évaluation fiable. Il souligne également la difficulté de détecter les anévrismes,

même pour des experts du domaine. Cependant, malgré ces bruits d'annotation, notre méthode a montré son potentiel en tant qu'approche prometteuse pour la détection d'anévrismes. Elle offre également des possibilités d'extension à d'autres problèmes liés à la détection d'objets de petite taille.

Lorsque les experts analysent et interprètent les images, ils se concentrent principalement sur les plans de coupes en 2D. Ces plans offrent une meilleure clarté visuelle et facilitent la distinction des anévrismes par rapport aux vaisseaux sanguins. Toutefois, le choix du plan approprié dépend de l'emplacement et de l'orientation de l'anévrisme par rapport à la structure vasculaire. Dans cette perspective, le prochain chapitre vise à étendre cette étude en associant la détection des anévrismes à une estimation de leur orientation, autrement dit, à l'estimation de leur position. Cette extension a pour objectif d'améliorer la compréhension et la précision de notre méthode, offrant ainsi une approche plus complète pour améliorer la détection et l'analyse des anévrismes, dans un contexte clinique.

Chapitre 6

Estimation de la pose des anévrismes intracrâniens dans un espace 3D

Sommaire

6.1 Positionnement de plans de coupes en imagerie médicale 122
6.1.1 Formulation du problème d'estimation de la pose d'objets 123
6.1.2 Approches géométriques pour l'estimation de la pose 124
6.1.3 Approches par apprentissage pour l'estimation de la pose 125
6.2 Collection et préparation de données 127
6.2.1 Ensembles de données et annotation 127
6.2.2 Échantillonnage et génération de données 129
6.3 Estimation de la pose des anévrismes par points de repère 129
6.3.1 Architecture du réseau et fonction de coût 129
6.3.2 Détails d'implémentation 131
6.3.3 Métriques d'évaluation 131
6.4 Étude expérimentale et résultats 132
6.4.1 Étude par ablation 133
6.4.2 Évaluation par validation croisée 134
6.5 Discussion et analyse des résultats 137
6.5.1 Méthode pour l'estimation de pose et la détection d'objets 137
6.5.2 Outil de visualisation et d'analyse des anévrismes 138
6.5.3 Analyse qualitative des résultats 139

Ce chapitre présente et développe notre méthode proposée pour l'estimation de la pose des anévrismes à partir d'images TOF-MRA. Ce travail a été publié et présenté lors de la conférence MICCAI 2023 [16]. La méthode proposée repose sur une approche d'apprentissage basée sur des points de repère, adaptée tant à l'estimation de la pose qu'à la détection d'objets. La section 6.1 examine la question du positionnement des plans de coupes en imagerie médicale. Nous commençons par définir le problème et son importance, puis nous passons en revue à la fois les approches géométriques classiques et les méthodes plus récentes basées sur l'apprentissage. La section 6.2 détaille le premier élément de notre méthode, qui porte sur l'acquisition et la préparation des données. Nous abordons à la fois les ensembles de données, l'approche d'annotation de la pose des anévrismes, ainsi que les approches d'échantillonnage et de génération de données. La section 6.3 expose l'architecture du réseau CNN employée pour l'estimation de la pose des anévrismes, la fonction de coût appropriée pour cette tâche, compte tenu du problème de déséquilibre des classes, ainsi que les métriques d'évaluation. Dans la section 6.4, nous présentons les expérimentations menées pour évaluer l'efficacité de notre méthode, comprenant une étude par ablation et une évaluation par validation croisée. Enfin,la section 6.5 propose une analyse qualitative des résultats et une discussion de ses avantages et limitations. Elle présente également un outil développé pour visualiser et manipuler les plans de coupes adaptés à la pose des anévrismes.

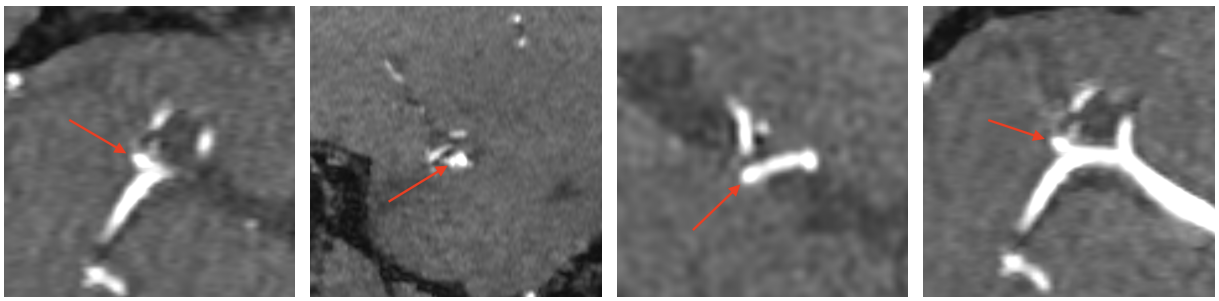
6.1 Positionnement de plans de coupes en imagerie médicale

Analyse des anévrismes à travers les plans de coupes canoniques

La détection et l'analyse des anévrismes intracrâniens constituent deux étapes essentielles du processus de diagnostic médical. La détection a pour objectif de localiser précisément les anévrismes au sein des images, tandis que l'analyse s'attache à évaluer leurs caractéristiques morphologiques et leur niveau de dangerosité [34, 88]. En pratique clinique, ces procédures sont généralement réalisées en utilisant des images en 2D, notamment les plans de coupes canoniques axial, sagittal et coronal (voir Figures 6.1a, 6.1b, et 6.1c). Ces différentes vues offrent une précision supérieure par rapport à la vue en 3D [9]. Chacun de ces plans de coupes permet la visualisation des anévrismes dans des régions spécifiques de l'arbre vasculaire [80]. Le plan axial est privilégié pour la visualisation des sites courants de formation d'anévrismes, la vue sagittale est plus appropriée pour les anévrismes situés dans le tronc basilaire, tandis que la vue coronale convient davantage pour les anévrismes localisés dans des zones telles que les artères cérébrales antérieures ou les segments de l'artère cérébrale moyenne. Mais le plan de coupe idéal pour l'analyse morphologique d'un anévrisme devrait offrir une seule vue transversale pour à la fois la poche anévrismale et le vaisseau porteur. C'est ce qui caractérise la vue de travail utilisée en angiographie 2D DSA pour les traitements endovasculaires : elle dégage la poche anévrismale de toute superposition avec les vaisseaux afférents afin d'éviter tout débordement du matériel d'embolisation dans ces derniers. Le choix du plan de coupe optimal dépend donc de la localisation de l'anévrisme et de son orientation (de l'ostium vers le dôme) par rapport au système vasculaire. Lorsqu'il s'agit de déterminer la localisation et l'orientation d'un objet dans un espace 3D, on parle alors de l'estimation de la pose de l'objet.

Estimation de la pose des anévrismes et plans de reformatés

La visualisation des anévrismes dans des plans de coupes adaptés à leur pose présente plusieurs avantages. Elle facilite d'abord une évaluation plus précise de leurs caractéristiques morphologiques, telles que la forme, la taille, les ratios d'aspect et de taille, ainsi que leur localisation par rapport aux structures vasculaires avoisinantes et leur connexion au vaisseau parent. L'analyse de ces facteurs est essentielle pour évaluer le flux sanguin entrant dans l'anévrisme, influençant directement le risque de rupture [166]. De plus, ces analyses orientent les décisions thérapeutiques, notamment le choix de la taille des spirales pour l'occlusion de l'anévrisme. La localisation et l'orientation des anévrismes contribuent également au recalage des images cérébrales, une étape importante dans l'acquisition et la reconstruction d'images [200]. Par ailleurs, comme évoqué ci-dessus, la pose des anévrismes permet d'obtenir une vue optimale lors de la procédure de traitement. La pose est déterminée en manipulant le volume 3DRA sur écran, jusqu'à trouver une projection adaptée. Elle est ensuite reproduite sur DSA en orientant l'arceau vasculaire. La détermination manuelle du plan de coupe optimal en TOF-MRA en fonction de la pose de l'anévrisme représente un défi pour les radiologues. Cette tâche nécessite de naviguer à travers plusieurs coupes pour identifier la vue la plus appropriée pour chaque cas spécifique. Les plans de coupes canoniques



(a) Axial (b) Sagittal (c) Coronal (d) Reformaté

FIGURE 6.1 – Visualisation d'un anévrisme de 2,69 mm de taille à travers les plans de coupes canoniques (a, b, c) et un plan de coupe reformaté (d) adapté à la direction de l'anévrisme, offrant une meilleure analyse de l'anévrisme et des structures environnantes.

peuvent ne pas convenir à toutes les orientations des anévrismes, ce qui souligne la nécessité d'un plan de coupe reformaté, ajusté à l'orientation particulière de l'anévrisme (voir Figure 6.1d). Par conséquent, une méthode automatisée pour localiser les anévrismes et fournir une visualisation immédiate adéquate serait un outil précieux pour assister les radiologues dans leur pratique clinique. Nous envisageons un parcours dynamique des plans de coupe, en rotation autour de l'axe principal de l'anévrisme, afin de faciliter l'analyse de l'anévrisme et de l'angioarchitecture environnante.

6.1.1 Formulation du problème d'estimation de la pose d'objets

L'estimation de la pose d'un objet implique la détermination précise de sa position (translation) et de son orientation (rotation) par rapport à un système de coordonnées fixe. La position spatiale de l'objet est généralement exprimée en termes de trois coordonnées (x, y, z), représentant par exemple le centre de l'objet. En ce qui concerne l'orientation de l'objet, elle peut être décrite de différentes manières, telles que les angles d'Euler, la matrice de rotation, le vecteur d'Euler ou les quaternions. Chaque méthode a ses avantages et inconvénients spécifiques [261].

Angles d'Euler et matrice de rotation

Les angles d'Euler, notés ϕ, θ, et ψ, permettent de représenter la rotation d'un objet dans un espace tridimensionnel. Ces angles décrivent les rotations autour des axes des coordonnées cartésiennes : X, Y, et Z. La rotation complète d'un objet est exprimée à l'aide d'une matrice de rotation R, qui est le résultat de la multiplication de trois matrices de rotation correspondant à chacun de ces angles. La matrice de rotation globale R est obtenue par la multiplication des matrices de rotation autour des axes Z, Y et X, dans cet ordre spécifique : $R = R_Z(\psi)R_Y(\theta)R_X(\phi)$, avec :

$$R_X(\phi) = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos\phi & -\sin\phi \\ 0 & \sin\phi & \cos\phi \end{bmatrix}, R_Y(\theta) = \begin{bmatrix} \cos\theta & 0 & \sin\theta \\ 0 & 1 & 0 \\ -\sin\theta & 0 & \cos\theta \end{bmatrix}, R_Z(\psi) = \begin{bmatrix} \cos\psi & -\sin\psi & 0 \\ \sin\psi & \cos\psi & 0 \\ 0 & 0 & 1 \end{bmatrix}$$

Les rotations d'un point ou d'un vecteur dans l'espace sont réalisées en effectuant une multiplication entre le vecteur de coordonnées et la matrice de rotation R. Les angles d'Euler sont couramment employés pour représenter les rotations en raison de leur simplicité et de leur capacité à rendre les rotations plus compréhensibles visuellement. Cependant, cette représentation comporte un inconvénient majeur, appelé le phénomène de blocage de Cardan (Gimbal Lock), qui se produit dans certaines orientations et conduit à la perte d'un degré de liberté, rendant impossible la rotation autour d'un axe spécifique [172]. De plus, lorsqu'on utilise les angles d'Euler, l'interpolation entre deux orientations n'est pas intuitive.

Vecteur d'Euler

La représentation axe-angle, également connue sous le nom de vecteur d'Euler, propose une manière concise de représenter les rotations, sous la forme d'un vecteur $\vec{u} = (u_x, u_y, u_z)$. L'angle de la rotation s'en déduit comme la norme du vecteur $\vec{u}$: $\theta = \|\vec{u}\|$, et l'axe de rotation est le vecteur $\vec{u}$ normalisé : $\vec{e} = \vec{u}/\|\vec{u}\|$. La formule de Rodrigues permet de calculer la matrice de rotation R_u associée au vecteur d'Euler $\vec{u}$:

$$R_u = \mathrm{Id} + \sin_\theta \tilde{u} + (1 - \cos_\theta)\tilde{u}^2$$

où Id est la matrice identité et $\tilde{u}$ est la matrice antisymétrique associée au vecteur $\vec{u}$. Comparée aux représentations précédentes, cette paramétrisation a l'avantage d'être concise (trois paramètres) et sans contrainte de normalisation. Elle résout ainsi le problème du blocage de cardan, sans toutefois échapper à une redondance de la représentation, due à la définition non-modulaire de l'angle. Par ailleurs, bien que potentiellement plus intuitive pour représenter une rotation, l'application de la rotation associée à un point de l'espace ou en composition à une autre rotation donne lieu à des expressions mathématiques complexes.

Quaternions

Les quaternions sont une manière robuste de représenter les rotations dans l'espace 3D, et ils présentent des avantages significatifs, notamment en évitant le problème de blocage de Cardan. Un quaternion est défini comme un nombre complexe à quatre dimensions, exprimé généralement sous la forme $q = w + xi + yj + zk$ avec comme contrainte que $w^2 + x^2 + y^2 + z^2 = 1$. Pour une rotation d'angle θ autour d'un axe défini par le vecteur unitaire (x, y, z), le quaternion correspondant s'exprime par :

$$q = \cos(\frac{\theta}{2}) + (xi + yj + zk)\sin(\frac{\theta}{2})$$

La transformation d'un vecteur p au moyen d'une rotation définie par un quaternion q s'effectue via la formule $p' = qpq^{-1}$, où p' représente le vecteur transformé et $q-1$ l'inverse du quaternion q. En outre, pour la combinaison de plusieurs rotations, la multiplication des quaternions respectifs est employée. L'ordre des multiplications est important en raison de la non-commutativité des quaternions.

En pratique, les quaternions sont largement utilisés pour faciliter l'interpolation des rotations, permettant des transitions fluides entre différentes orientations [105]. Cela les rend particulièrement adaptés pour des applications telles que l'animation et la modélisation 3D [75]. De plus, ils minimisent l'accumulation d'erreurs lors de rotations successives, ce qui est essentiel dans les calculs continus dont la stabilité est augmentée. Cependant, les quaternions présentent des défis en termes de compréhension conceptuelle, car leur nature à quatre dimensions et leur structure mathématique les rendent moins intuitifs. De plus, la nécessité de maintenir les quaternions normalisés pour qu'ils restent valides comme représentants de rotation peut rendre leur gestion et optimisation délicates.

6.1.2 Approches géométriques pour l'estimation de la pose

Dans le domaine de la vision par ordinateur, les premières méthodes d'estimation de la pose des objets reposaient sur des approches géométriques [18, 69, 94]. Ces techniques se basent sur la représentation de l'objet dans différentes images afin d'inférer les paramètres de sa position et de son orientation dans l'espace tridimensionnel. Elles nécessitent la superposition d'un modèle géométrique de l'objet sur une image, dans le but d'obtenir la meilleure correspondance possible. Les méthodes existantes ont été largement appliquées à l'analyse d'images naturelles en deux dimensions et de nuages de points, principalement pour des applications liées à l'estimation de la pose des caméras [73].

En imagerie médicale, l'estimation de la pose d'objets est fréquemment abordée sous l'angle de l'alignement de coupes d'images en relation avec des structures anatomiques spécifiques. L'objectif de cette démarche est d'optimiser la compréhension et l'interprétation des images médicales en assurant une correspondance précise entre les différentes coupes et les références anatomiques. Les premières études dans ce domaine se focalisent sur des techniques de recalage d'images [93, 195]. Cette procédure implique l'ajustement des images en fonction d'un repère anatomique préexistant ou d'un modèle préalablement défini. Pour y parvenir, des points de repère ou des caractéristiques communes sont identifiés dans diverses images, puis utilisés pour calculer les ajustements nécessaires en vue d'obtenir un alignement précis et approprié. Les transformations rigides sont couramment employées pour positionner correctement les images, notamment pour des organes tels que le cerveau [28, 294, 57] et le genou [46]. Néanmoins, il est important de noter que la performance des méthodes de recalage peut être impactée par des artefacts dans les images ainsi que par des variations d'apparence ou de position dans les images [333]. Par conséquent, l'utilisation de ces méthodes est principalement restreinte aux organes présentant une rigidité (quasi-)constante, afin de minimiser les erreurs résultant de variations anatomiques excessives.

Dans le contexte spécifique des anévrismes intracrâniens, nous n'avons connaissances que d'une seule méthode développée en 2019 [141]. Elle est utilisée dans un contexte thérapeutique, visant à faciliter l'évaluation du risque de rupture en analysant le flux sanguin entrant dans la poche anévrismale. L'objectif principal de cette méthode est d'automatiser un plan de coupe au niveau du collet des anévrismes, comme illustré dans la Figure 6.2. Cette approche repose sur l'analyse des données de maillage des vaisseaux sanguins, en utilisant des techniques d'analyse de surface pour identifier les centres des anévrismes, les points d'entrée des vaisseaux parents, ainsi que les axes centraux de ces vaisseaux. Par la suite, un plan de coupe est défini orthogonalement à l'axe central de l'anévrisme à partir de ces points de repère. Cependant,

cette méthode a été évaluée sur un ensemble de données limité comprenant 55 anévrismes, ayant une taille médiane de 6,5 mm. Ses performances se sont avérées limitées pour les anévrismes mesurant moins de 7 mm, en raison de l'absence de caractéristiques distinctives de courbure de surface, ce qui complique la précision de la localisation du collet. Cette limitation s'applique également aux anévrismes de forme complexe. De plus, l'efficacité de cette méthode dans l'analyse de données 3D sous forme de voxels n'a pas été explorée. Enfin, la conversion de ces données en maillage peut entraîner une perte de détails, en particulier pour les anévrismes de petite taille [319].

Face aux difficultés associées aux approches géométriques, notamment dans des structures avec d'importantes variations de pose, comme dans le cas des anévrismes, les méthodes basées sur l'apprentissage automatique se montrent plus efficaces [328]. Ces méthodes ont la capacité d'extraire automatiquement les caractéristiques pertinentes des données, réduisant ainsi la nécessité d'extraire et de formuler manuellement ces caractéristiques.

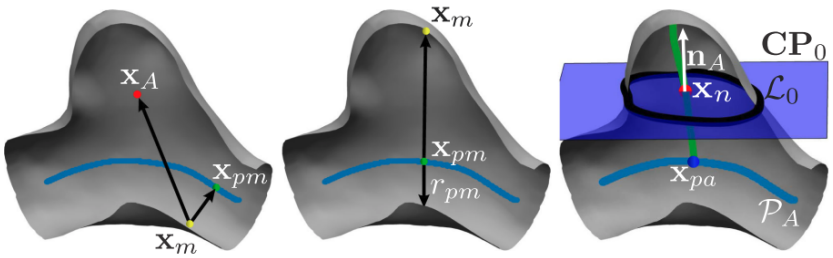


FIGURE 6.2 – Positionnement du plan de coupe au niveau du collet de l'anévrisme : les lignes médianes des vaisseaux et le centre de l'anévrisme sont utilisés pour définir le plan qui sépare l'anévrisme du vaisseau parent.

6.1.3 Approches par apprentissage pour l'estimation de la pose

L'application des techniques d'apprentissage automatique pour l'estimation de la pose implique la régression directe des paramètres de pose d'un objet à partir d'images en utilisant des réseaux CNNs. Des exemples d'architectures telles que PoseNet [194], Pix2Pose [223], et PoseCNN [313] mettent en évidence la capacité de ces méthodes à traiter différentes configurations de pose d'objets. L'entraînement de ces réseaux s'effectue de manière supervisée en utilisant des ensembles de données annotées contenant à la fois des images et leurs paramètres de pose. Ces réseaux apprennent à extraire des caractéristiques visuelles des images, qui sont ensuite utilisées pour estimer la pose des objets, sans recourir explicitement à des modèles géométriques. Par exemple, le réseau PoseNet a été utilisé pour estimer la pose de caméra à partir d'images 2D, où les paramètres de localisation sont exprimés en coordonnées spatiales de la caméra, tandis que ceux d'orientation sont représentés par des quaternions. Il convient de noter que ces réseaux dépendent fortement de la disponibilité de bases de données pour leur apprentissage, un aspect critique dans le domaine médical.

En imagerie médicale, l'utilisation des algorithmes d'apprentissage pour estimer la pose d'organes ou d'anomalies reste un domaine de recherche relativement peu explorée. La majorité des travaux existants se concentrent sur des images de basse résolution, telles que les "scout scans". Ces dernières visent à assurer que l'anatomie ciblée soit intégralement capturée par les images avant de passer à une imagerie de haute résolution. Dans ce cadre, les méthodes existantes vise à faciliter l'alignement de coupes en fonction d'organes spécifiques comme le genou [169, 328, 333], l'épaule [333] et le cerveau [136]. Ces organes sont caractérisés leurs grandes dimensions, leurs formes distinctes, leurs positions précises dans l'image et leurs orientations standardisées. Toutefois, l'annotation de données représente un défi majeur en raison de la difficulté à obtenir les paramètres de pose qui sont en générale définie anatomiquement. De plus, la capacité de généralisation de des méthodes par apprentissage dépend fortement de la disponibilité de bases de données volumineuses.

L'estimation de la pose d'objets tels que les anévrismes intracrâniens constitue un défi majeur en raison de la diversité et de la variabilité de ces anomalies vasculaires. Les défis à relever incluent leur

petite taille, leur nombre fluctuant, leurs formes variées et leurs emplacements diversifiés dans l'arbre vasculaire. De surcroît, pour réussir l'estimation de la pose de ces anomalies, l'emploi d'images 3D de haute résolution est essentiel. Cela exige le développement d'algorithmes avancés capables d'apprendre la localisation et l'orientation de ces anomalies à partir de données restreintes. Une difficulté supplémentaire découle du fait que leur pose est déterminée manuellement et de manière subjective. À notre connaissance, aucune définition anatomique précise de cette pose n'existe.

Les méthodes existantes ont suivi un processus en deux étapes. Dans un premier temps, une composante de localisation est entraînée pour détecter les objets en utilisant des points de repère. Ensuite, une matrice de rotation est déduite de ces points afin d'estimer l'orientation. Cependant, l'annotation et la détection de multiples points de repère pour chaque objet représentent une tâche complexe pouvant être influencée par les artefacts des images [333]. Les approches plus récentes pour la localisation des coupes optent pour l'apprentissage profond et décrivent la localisation de l'objet ainsi que son orientation à travers un ensemble de paramètres décrivant la pose de l'objet dans l'espace. Dans cette approche, les régions d'intérêt (RoIs) sont détectées en première étape, tandis que l'orientation de l'objet est estimée pour chaque RoI en deuxième étape. Par exemple, dans [333], les réseaux Faster R-CNN [239] et V-Net [201] ont été employés pour localiser et segmenter le plan de coupe, suivi d'un algorithme d'analyse en composantes principales (PCA) pour dériver la matrice de rotation à partir de deux vecteurs unitaires (voir Figure 6.3). Il est néanmoins important de noter que la plupart des méthodes existantes ont principalement été appliquées à des images de basse résolution et requièrent une puissance de calcul substantielle pour les images de haute résolution. Dans ce contexte, les approches en une étape, telles que YOLO [238], présentent des performances prometteuses dans la détection d'objets en 2D, offrant une plus grande flexibilité que les approches en deux étapes [176].

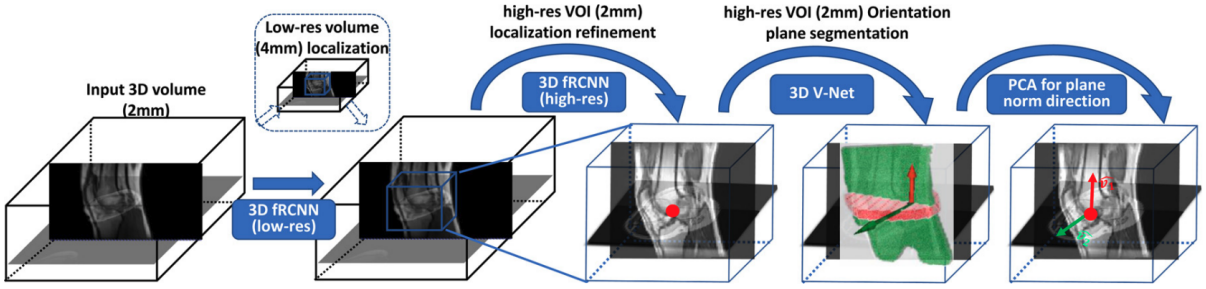


FIGURE 6.3 – Estimation de la pose du genou [333] : Une première étape localise les RoIs avec un réseau Faster R-CNN sur une image à basse résolution, puis un second réseau repère précisément le centre du genou à partir d'une image à haute résolution. Le réseau V-Net segmente le plan d'orientation à partir de chaque RoI, suivi de PCA pour dériver la matrice de rotation.

Contributions

Ce chapitre introduit DeepAnePose, notre méthode pour l'estimation de la pose d'anévrismes à partir d'images TOF-MRA, via un réseau CNN en une seule étape. Pour surmonter les difficultés liées à l'annotation de la pose dans l'imagerie médicale, nous avons utilisé une méthode d'annotation efficace et rapide, identique à celle mentionnée dans les chapitres précédents, qui permet de déterminer la position et l'orientation des anévrismes. Pour pallier le déséquilibre des classes, la stratégie de gestion des données déjà appliquée précédemment a été réutilisée, associant petits patchs, échantillonnage et synthèse. Notre approche s'appuie sur l'estimation de pose à travers des points de repère, évitant les écueils des discontinuités rotationnelles rencontrées avec les angles d'Euler et les quaternions [336]. Cette méthode convient à la fois pour les tâches d'estimation de la pose et de détection des anévrismes. L'architecture de notre réseau neuronal est inspirée de celle utilisée précédemment pour la détection, avec une optimisation de l'orientation des anévrismes par une fonction de coût basée sur la similarité cosinus. Les performances de notre méthode ont été évaluées à l'aide de deux vastes ensembles de données, en utilisant des études par ablation et une validation croisée.

6.2 Collection et préparation de données

6.2.1 Ensembles de données et annotation

Collection de données

Nous avons utilisé deux ensembles de données TOF-MRA identiques à ceux du chapitre précédent, que nous rappelons brièvement. Le premier ensemble de données se compose de 111 examens, recueillis de manière rétrospective au CHRU de Nancy sur une période s'étendant d'avril 2015 à janvier 2020. Ces examens incluaient des données provenant de 56 femmes et 55 hommes. Les critères d'inclusion étaient restreints à la présence d'anévrismes sacculaires, tandis que les anévrismes déjà traités, les anévrismes fusiformes, ceux de grande taille (> 20 mm), ainsi que les examens de suivi étaient exclus. L'imagerie a été effectuée à l'aide d'un appareil IRM de GE Healthcare, fonctionnant à une intensité de champ magnétique de 3T. Cette configuration a permis de générer des images volumétriques de haute résolution, avec des dimensions de $512 \times 512 \times 254$ voxels et une taille de voxel de $0,47 \times 0,47 \times 0,4$ mm^3. L'ensemble de données comprenait des images contenant de un (dans 84 cas sur 132 images) à 5 anévrismes (dans 4 cas), totalisant 206 anévrismes. Le diamètre moyen de ces anévrismes était de 3,97 mm, avec un écart-type de 2,32 mm et une plage allant de 0,96 à 19,63 mm. La majorité de ces anévrismes étaient de petite taille, avec 81 mesurant moins de 3 mm et 77 compris entre 3 et 5 mm.

Le deuxième ensemble de données provient des données publiques du CHUV de Lausanne [80]. Il comprenait 296 images TOF-MRA provenant de 284 patients, réparties entre 127 images sans anévrisme et 168 images contenant des anévrismes. Pour nos analyses, nous avons exclu les images de suivi, les anévrismes fusiformes et ceux de plus de 20 mm, ce qui a réduit notre ensemble de données à 270 images. Ces images ont été collectées entre 2010 et 2015, en utilisant des scanners IRM de 1,5T et 3T. Au total, ces images contenaient 164 anévrismes, présentant des caractéristiques similaires à celles du premier ensemble, notamment un diamètre moyen de 3,74 mm avec un écart-type de 2,17 mm (plage : 1,37 à 13,64 mm). Parmi ces anévrismes, 66 mesuraient moins de 3 mm et 72 se situaient entre 3 et 5 mm. Le nombre d'anévrismes par image variait de 0 (pour 127 images) à 3 (pour 3 images). La Figure 6.4 illustre une comparaison des tailles des anévrismes entre ces deux ensembles de données. Il est important de noter que nous n'avons pas utilisé les annotations révisées du chapitre précédent pour l'entraînement des modèles. Cette décision a été prise dans le but de refléter des conditions réelles, intégrant ainsi les incertitudes et les variations d'annotations. Cependant, lors de l'évaluation des performances, nous avons utilisé à la fois les annotations initiales et les annotations révisées.

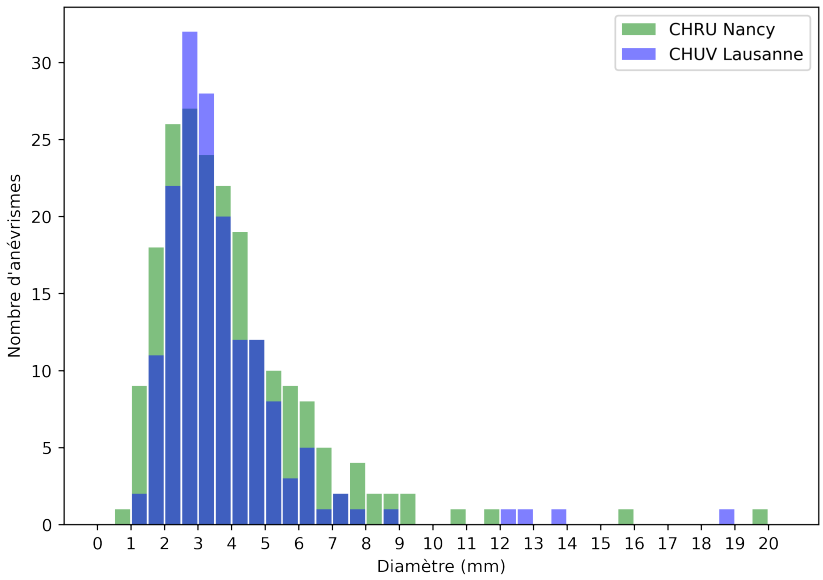
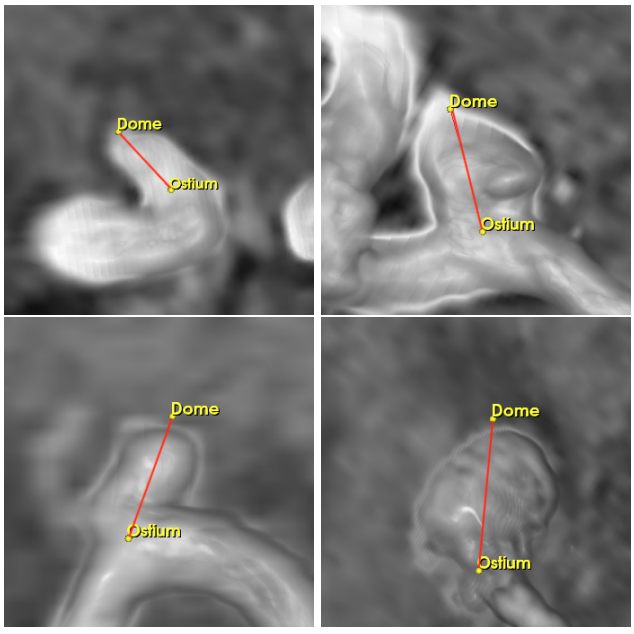


FIGURE 6.4 – Comparaison des tailles des anévrismes dans les ensembles de données CHRU et CHUV.

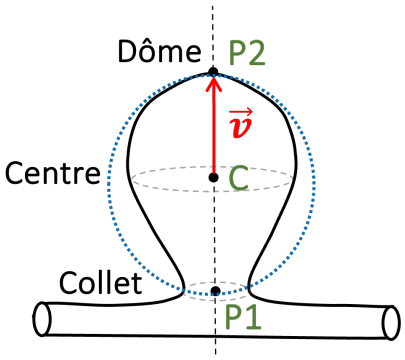
Annotation de données

La plupart des travaux existants concernant la détection et la segmentation des anévrismes reposent sur une annotation voxel par voxel, une approche à la fois fastidieuse et sujette à des variations entre les annotateurs. Cette approche se révèle moins adaptée pour les tâches d'estimation de la pose des anévrismes, où l'orientation des anévrismes ne peut pas être déduite à partir de ces annotations. Par ailleurs, dans la base de données CHUV de Lausanne, une approche d'annotation approximative a été utilisée, basée sur des annotations par sphères. Cependant, comme cela a été démontré dans le chapitre précédent, cette approche ne permet pas une représentation précise des anévrismes par rapport aux vaisseaux parentaux. De plus, les annotations ont été fournies sous forme de masques d'images, ce qui rend impossible la déduction de l'orientation des anévrismes.

En revanche, notre approche d'annotation se distingue par un protocole standardisé, adapté aux tâches de segmentation, de détection et d'estimation de la pose des anévrismes. Chaque anévrisme est annoté au moyen de deux points stratégiques : le centre du collet de l'anévrisme, également désigné sous le nom d'ostium, ainsi qu'un point situé sur son axe principal, au niveau du dôme (voir des exemples dans la Figure 6.5a). Ces points fournissent des informations relatives à la localisation (y compris les coordonnées du centre), aux dimensions (mesurées de l'ostium au dôme), ainsi qu'à l'orientation (par exemple, de l'ostium au dôme) de l'anévrisme. Dans notre travail, nous représentons la pose de chaque anévrisme au moyen des coordonnées du centre de l'anévrisme, notées $C = (P1 + P2)/2$, et du vecteur de l'axe principal allant du centre de l'anévrisme au dôme, exprimé comme $\vec{v} = P2 - C$, tel qu'illustré dans la Figure 6.5b. Pour faciliter davantage le processus d'annotation, nous avons développé un plugin Python spécifique destiné au logiciel 3D Slicer [92]. Ce plugin permet la visualisation et l'ajustement des points d'annotation aussi bien dans la vue 3D que dans les plans de coupes canoniques.



(a)



(b)

FIGURE 6.5 – Annotation de la pose des anévrismes : (a) les points d'annotation (Collet et Dôme, en jaune) et le vecteur de l'axe principal (en rouge). (b) La pose est déduite par le centre $C = (\mathrm{P1} + \mathrm{P2})/2$ et du vecteur d'axe $\vec{v} = \mathrm{P2} - C$. L'anévrisme peut également être représenté sous forme d'une sphère ayant un centre C et un rayon $r = |\vec{v}|$. A noter que les plans de coupes reformatés, contenant cet axe, permet une analyse morphologique et métrologique de l'anévrisme.

6.2.2 Échantillonnage et génération de données

En se basant sur les résultats obtenus dans le chapitre précédent, nous avons adopté la même stratégie de gestion de données utilisée pour la tâche de détection, décrite en détail dans la section 5.2.2. Cette stratégie repose sur l'utilisation de petits patchs de dimensions $96 \times 96 \times 96$ voxels, possédant une taille de voxel isotrope de 0,4 mm^3 par côté, ce qui leur confère une dimension de 38,4 mm de chaque côté du patch. Cette approche se distingue par une meilleure efficacité computationnelle par rapport à celles qui préconisent l'utilisation de patchs de plus grande dimension, telles que les méthodes nnDetection [27] et nnUNet [135]. De plus, elle permet l'extraction de multiples patchs négatifs non intersectants à partir de chaque ensemble de données de patient, ce qui enrichit le jeu de données d'entraînement et assure une représentation fiable des régions sans d'anévrismes.

Pour résoudre le problème de déséquilibre des classes entre les patchs négatifs et positifs, nous avons mis en place des stratégies d'échantillonnage de données appropriées. La première stratégie consiste à augmenter la diversité des formes d'anévrismes en dupliquant chaque patch positif 50 fois à chaque époque d'entraînement, tout en appliquant des distorsions aléatoires. Ces distorsions comprennent le déplacement aléatoire de chaque point de contrôle dans une grille de dimensions $3 \times 3 \times 3$ autour du patch, avec une amplitude maximale de 3 mm dans toutes les directions. Les déformations sont ensuite interpolées à l'aide d'une spline cubique. Pour faciliter la distinction entre les vaisseaux sains et les anévrismes, notre deuxième stratégie pré-sélectionne 40 patchs négatifs non intersectants par image, dont 30 sont centrés sur des vaisseaux sanguins. Ces patchs négatifs sont sélectionnés de manière itérative en choisissant les voxels les plus lumineux comme centres, tout en maintenant une distance de 40 mm entre chaque paire de centres. À chaque époque d'entraînement, nous utilisons l'ensemble complet des patchs positifs, ainsi qu'une sélection aléatoire de patchs négatifs supplémentaires à partir de chaque image, de manière à ce que 15% de l'ensemble d'entraînement soit constitué de patchs négatifs.

Pour augmenter la variabilité des données d'entraînement, nous appliquons des techniques d'augmentation classiques : rotations aléatoires (de 0 à 180°), translations aléatoires (de 0 à 10 mm) dans les trois dimensions, ainsi que des effets miroirs par rapport au plan sagittal. Enfin, pour garantir une représentation précise des anévrismes à l'intérieur des patchs positifs, nous nous assurons qu'ils englobent entièrement l'anévrisme dans le patch augmenté.

6.3 Estimation de la pose des anévrismes par points de repère

Cette section débute par la présentation de l'architecture du réseau neuronal employé. Ensuite, elle aborde la conception de la fonction de coût spécialement élaborée pour remédier au problème du déséquilibre des classes, en utilisant la similarité cosinus pour optimiser l'orientation des anévrismes. Enfin, elle fournit des détails d'implémentation, les métriques d'évaluation, ainsi que les méthodes de référence.

6.3.1 Architecture du réseau et fonction de coût

Architecture du réseau CNN

L'architecture de notre réseau représente une adaptation du réseau utilisé précédemment pour la tâche de détection des anévrismes. Comme illustré dans la Figure 6.6, la structure de réseau repose sur une approche en grille, où le patch d'entrée en 3D, mesurant $96 \times 96 \times 96$ voxels, est subdivisé en $12 \times 12 \times 12 = 1728$ cellules. Chacune de ces cellules, mesurant $8 \times 8 \times 8$ voxels, est chargée de l'estimation de la pose d'un seul anévrisme en se basant sur la présence ou l'absence du centre de l'anévrisme à l'intérieur de la cellule.

Le patch d'entrée est encodé en cartes de caractéristiques à l'aide de blocs de convolution résiduels et d'opérations de sous-échantillonnage. Ensuite, un bloc spécifique est utilisé pour estimer la pose de chaque cellule du patch d'entrée. Cette grille est composée de 1728 cellules et le bloc de pose est constitué de trois blocs de convolution consécutifs, suivis de trois convolutions parallèles : (1) La première convolution génère un score de confiance, variant de 0 à 1, pour indiquer la présence du centre d'un anévrisme dans la cellule. (2) Pour les cellules positives (celles contenant le centre d'un anévrisme), la deuxième convolution est suivie d'une fonction sigmoïde pour prédire les coordonnées normalisées du centre de

l'anévrisme $C = (C_x, C_y, C_z)$ par rapport à la taille de la cellule. (3) La troisième convolution concenrne également les cellules positives. Elle estime la taille et l'orientation de l'anévrisme en calculant le vecteur d'axe $\vec{v} = (v_x, v_y, v_z)$. Le rayon de l'anévrisme est donné par $r = |\vec{v}|$, et son orientation est déterminée par le vecteur $\vec{v}$ normalisé $(\vec{v}/|\vec{v}|)$. De plus, chaque prédiction peut également être représentée sous forme d'une sphère avec un centre C et un rayon r.

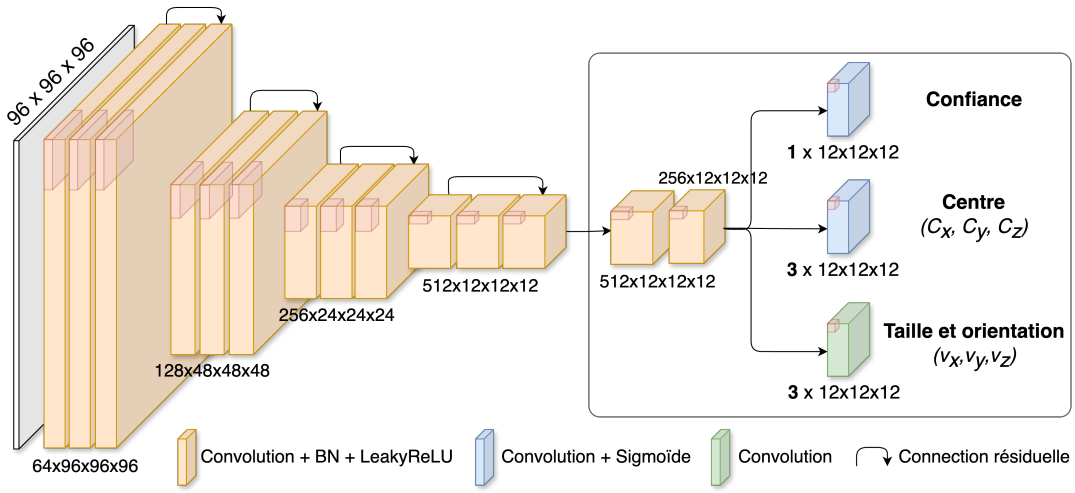


FIGURE 6.6 – Notre architecture d'estimation de la pose de l'anévrisme. Les cartes de caractéristiques en 2D ne sont utilisées qu'à des fins de visualisation ; leurs tailles réelles en 3D sont affichées

Fonction de coût

L'architecture de notre réseau repose sur une approche en grille, engendrant un déséquilibre significatif entre les cellules négatives (N), ne contenant pas d'anévrismes, et les cellules positives (P), renfermant les centres des anévrismes. Afin de remédier à ce déséquilibre, nous avons opté pour l'utilisation d'une fonction de coût pondérée pour l'estimation de la pose des anévrismes, en nous inspirant du réseau de détection présenté dans le chapitre précédent. Cette fonction de coût intègre des termes relatifs à la classification, la régression et l'orientation des anévrismes (voir Équation 6.1). Il convient de noter que les termes de localisation et d'orientation ne sont calculés que pour les cellules positives.

$$\text{Coût} = \text{Classification} + \text{Régression} + \text{Orientation} \tag{6.1}$$

Pour optimiser la classification des cellules contenant le centre des anévrismes, notre méthode utilise la fonction de coût d'entropie croisée binaire (BCE). Cette fonction est appliquée de manière distincte aux cellules positives (BCE_P) et aux cellules négatives (BCE_N). Dans le but de mettre en avant la détection des anévrismes par rapport à l'arrière-plan, nous adaptons le terme lié aux cellules négatives en introduisant une pondération basée sur la moitié du nombre de cellules positives $(\#P)$ présentes dans chaque lot d'entraînement (voir Équation 6.2).

$$\text{Classification (P,N)} = \text{BCE}(P) + 0.5 \times \#P \times \text{BCE}(N) \tag{6.2}$$

Pour évaluer les paramètres associés à la localisation des anévrismes, à savoir les coordonnées du centre (C_x, C_y, C_z) et les composantes du vecteur d'orientation $\vec{v} = (v_x, v_y, v_z)$, nous utilisons la métrique de l'erreur quadratique moyenne (MSE) (voir Équation 6.3). Cette métrique s'avère particulièrement adaptée aux valeurs de petite échelle [185]. Pour évaluer l'orientation des anévrismes, nous utilisons la métrique de similarité cosinus, appliquée aux vecteurs d'orientation $\vec{v}$ des cellules positives (voir Équation 6.4). Un coefficient de pondération de 5 est employé afin d'équilibrer les contributions des termes de confiance (appliqués à l'ensemble des cellules) avec ceux de localisation et d'orientation (appliqués uniquement aux cellules positives).

$$\text{Régression (P)} = 5 \times \sum_{i=1}^{\#P} \text{MSE}(C_x, C_y, C_z, v_x, v_y, v_z) \tag{6.3}$$

$$\text{Orientation (P)} = 5 \times \sum_{i=1}^{\#P} (1 - \text{Similarité Cosinus}(\vec{v_i})) \tag{6.4}$$

6.3.2 Détails d'implémentation

Nous avons implémenté notre méthode en utilisant le langage de programmation Python (3.10.0) et le framework PyTorch (1.10.0). Les images TOF-MRA ont été normalisées à l'aide de la méthode de normalisation en z-score, ayant une moyenne de 0 et un écart-type de 1. Notre modèle de réseau neuronal comporte environ 28 millions de paramètres, occupant 76,31 Mo de mémoire.

Phase d'entraînement

Pour optimiser ces paramètres, nous avons employé l'algorithme de descente de gradient stochastique (SGD) avec un weight decay fixé à 5e-4, un momentum de 0,98, et la fonctionnalité Nesterov activée pour accélérer la convergence. Les hyperparamètres suivants ont été sélectionnés après des tests sur un sous-ensemble de 30 patients de la base de données CHRU de Nancy : 200 époques d'entraînement au total, échantillonnage équilibré des patchs négatifs et positifs, et une taille de lot de 32. Nous avons également mis en place un planificateur de taux d'apprentissage personnalisé, avec une phase d'échauffement pour les trois premières époques, suivi d'un taux d'apprentissage fixe de 10^{-2} jusqu'à la 50ème époque, puis d'un planificateur de type polynomial [55] pour réduire progressivement le taux jusqu'à 10^{-4} à la 200ème époque. L'entraînement du réseau s'est déroulé sur une carte graphique NVIDIA RTX A6000 équipée de 48 Go de mémoire GPU. Chaque itération d'entraînement a requis une allocation de mémoire GPU de 38 Go. Le temps total d'entraînement pour chaque modèle de notre méthode s'est élevé à environ 24 heures.

Phase d'inférence

Pendant la phase de prédiction, nous avons utilisé une approche de reconstruction basée sur des patchs pour estimer la localisation $C = (C_x, C_y, C_z)$ et l'orientation $\vec{v} = (v_x, v_y, v_z)$ des anévrismes à partir de l'ensemble complet de chaque image TOF-MRA. Le processus de prédiction de chaque image TOF-MRA se déroule en quatre étapes principales :

1. Le volume initial a été divisé en patchs de dimensions $96 \times 96 \times 96$ voxels, avec une résolution isotrope de 0,4 mm. Afin de réduire les artefacts causés par les bords des patchs lors des opérations de convolution, une zone de chevauchement de 32 voxels a été maintenue entre les patchs adjacents.
2. Les prédictions ont été générées pour chaque patch, puis redimensionnées pour correspondre à la résolution du volume d'origine. Pour qu'une prédiction soit acceptée, son centre prédit (C) doit se trouver dans la zone centrale du patch, soit une région de $64 \times 64 \times 64$ voxels.
3. De plus, une étape de suppression non maximale (NMS) a été mise en œuvre pour éliminer les prédictions qui se chevauchent, avec un seuil d'intersection sur union (IoU) de 10% : chaque prédiction a été considérée comme une sphère définie par son centre C et son rayon $r = |\vec{v}|$. Seule la prédiction avec le score de confiance le plus élevé est conservée.
4. De manière complémentaire, nous avons appliqué une technique d'augmentation au moment du test (TTA), consistant à appliquer des effets miroir le long des axes axial, sagittal et coronal. Les détections présentes dans toutes les images augmentées ont été conservées, toujours en utilisant un seuil IoU de 10%.

Notre code et les annotations utilisées dans ce chapitre sont disponibles en accès public à l'adresse suivante : `https://gitlab.inria.fr/yassis/DeepAnePose`.

6.3.3 Métriques d'évaluation

Les travaux antérieurs sur le positionnement des plans de coupes se sont focalisées sur des structures anatomiques spécifiques telles que le cerveau, l'épaule et le genou, où la position et le nombre de ces structures sont déjà définis dans les images. Cependant, pour les anévrismes, ces paramètres restent inconnus à l'avance. Il est donc nécessaire de procéder d'abord à une correspondance entre les prédictions et la vérité terrain. Cette étape est accomplie grâce à la tâche de détection d'objets. Suite à cela, l'évaluation des performances de l'estimation de la pose est réalisée pour les anévrismes correctement détectés.

Détection des anévrismes

Pour la détection des anévrismes, chaque anévrisme est représenté sous la forme d'une sphère avec un centre C et un rayon égal à la norme du vecteur d'orientation $\vec{v}$, soit $r = |\vec{v}|$. Les métriques d'évaluation et les méthodes de référence évoquées dans le chapitre précédent sont également appliquées. Pour rappel, l'évaluation se base sur des métriques reliant l'indice d'intersection sur union (IoU) entre les sphères prédites et celles de la vérité terrain (GT), avec un seuil de 10% [27, 192]. Une sphère de la GT obtenant un score IoU supérieur à 10% est considérée comme un vrai positif (TP), tandis qu'une sphère de la GT avec un score IoU maximal inférieur à 10% est classée comme un faux négatif (FN). Les sphères prédites ayant un score IoU inférieur à 10% pour toute sphère de GT sont qualifiées de faux positifs (FP). En cas de multiples détections pour une même sphère de la GT, seule la détection avec le score de confiance le plus élevé est retenue, les autres étant considérées comme des faux positifs [221]. Notre évaluation repose principalement sur la métrique de la précision moyenne (AP), qui évalue la sensibilité et la précision à différents seuils de classification. En outre, nous calculons la sensibilité et le nombre moyen de faux positifs par cas en utilisant un seuil de classification par défaut de 50%.

Afin d'évaluer notre méthode, nous l'avons comparée à plusieurs travaux, y compris notre propre méthode étudiée dans le chapitre précédent (DeepAneDet). Cette dernière consiste à détecter les anévrismes en utilisant des sphères définies par leur centre et leur rayon. En revanche, l'approche présentée dans ce chapitre se distingue de DeepAneDet par la manière dont elle estime le rayon. Dans cette nouvelle méthode, le rayon est remplacé par l'estimation du vecteur d'orientation $\vec{v}$.

En plus de comparer notre nouvelle approche à la méthode DeepAneDet, nous l'avons également comparée à trois méthodes de référence bien établies en détection d'objets en imagerie médicale : nnDetection [27], SCPM-Net [185], et nnUNet [135]. Comme auparavant, pour garantir une comparaison équitable, nous avons converti les résultats de nnDetection et de nnUNet en représentations sphériques. Pour nnDetection, cela a été accompli en transformant les boîtes englobantes 3D en sphères, en utilisant la diagonale de la boîte comme diamètre de la sphère. Pour nnUNet, nous avons ajusté une sphère à chaque composante connectée de l'image prédite : le diamètre de la sphère a été calculé comme la distance maximale entre chaque paire de voxels, et le score de confiance a été déterminé comme la valeur maximale prédite du voxel.

Estimation de la pose des anévrismes

À notre connaissance, il n'existe pas de méthodes de référence permettant de comparer efficacement la tâche d'estimation de la pose des anévrismes, qui implique de multiples objets dans une même image et à diverses localisations. Pour évaluer notre méthode, nous nous sommes concentrés sur les anévrismes correctement détectés, que nous avons qualifiés de vrais positifs (TP). L'évaluation repose sur la métrique de sensibilité de détection, calculée avec un seuil de classification par défaut fixé à 50%. Deux aspects principaux ont été pris en compte dans cette évaluation : la précision de la localisation du centre de l'anévrisme C, et la précision de son orientation, représentée par le vecteur $\vec{v}$.

Pour évaluer la précision de la localisation des anévrismes, nous avons employé la métrique de distance euclidienne. Cette métrique évalue l'erreur en millimètres entre le centre prédit de l'anévrisme (C) et le centre réel défini dans la vérité terrain (GT). Une faible valeur d'erreur indique une performance élevée en termes de localisation. En ce qui concerne l'erreur d'orientation, elle est mesurée à l'aide de la similarité cosinus, permettant d'évaluer la différence angulaire en degrés entre le vecteur d'orientation prédit de l'anévrisme ($\vec{v}$) et le vecteur correspondant dans la GT. Cette mesure peut varier dans une plage allant de 0 à 180 degrés, où une valeur plus faible indique une meilleure précision dans l'estimation de l'orientation de l'anévrisme.

6.4 Étude expérimentale et résultats

Cette section présente les expérimentations réalisées pour évaluer l'efficacité de notre méthode, tant pour l'estimation de la pose que pour la détection des anévrismes. Notre étude expérimentale se divise en deux phases distinctes. La première phase consiste en une étude par ablation, visant à analyser l'impact de chaque composant de notre méthode et à justifier la configuration adoptée. Cette analyse a été effectuée en utilisant notre base de données privée du CHRU de Nancy. Ensuite, une évaluation par validation

croisée a été menée, incluant les données publiques du CHUV de Lausanne [80]. Cette phase a pour but d'évaluer la capacité de généralisation de notre méthode sur un ensemble de données plus large.

6.4.1 Étude par ablation

L'étude par ablation a débuté par l'entraînement de notre réseau neuronal spécialement conçu pour estimer la pose des anévrismes. La fonction de coût, telle que décrite dans la section 6.3.1, a été employée pour optimiser les scores de confiance, la localisation des centres ($C = (P2 + P1)/2$), ainsi que les vecteurs d'orientation ($\vec{v} = P2 - C$) des anévrismes. Les patchs d'entraînement ont été générés en suivant les stratégies d'échantillonnage et de génération de données exposées dans la section 6.2.2.

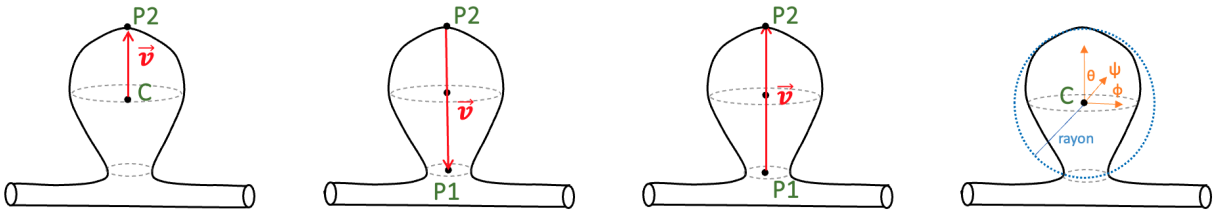
Pour évaluer l'efficacité de notre modèle de base, nommé *Modèle0*, nous l'avons comparé à 4 variantes distinctes de modèles. Chaque variante a été conçue pour explorer différentes configurations dans l'estimation de la pose des anévrismes, en tenant compte de la position des points de repère ainsi que des représentations décrivant l'orientation des anévrismes.

— *Modèle1* : Dans ce modèle, les points $P1$ (ostium) et $P2$ (dôme) des anévrismes sont prédits au lieu du centre C et $P2$ utilisé dans le *Modèle0* (voir Figure 6.7b). Chaque cellule du CNN est chargée de détecter un anévrisme si le centre de l'ostium $P1$ se trouve dans cette cellule. Le vecteur $\vec{v}$ est défini comme $P2 - P1$ pour la similarité cosinus. Ce modèle se focalise davantage sur la détection de l'ostium.

— *Modèle2* : Contrairement au *Modèle1*, dans ce modèle, le dôme $P2$ est au centre de la cellule à partir de laquelle l'ostium $P1$ est estimé (voir Figure 6.7c). Par conséquent, le vecteur $\vec{v}$ d'orientation est défini du dôme vers l'ostium ($P1 - P2$). Ce modèle se concentre davantage sur la détection du dôme.

— *Modèle3* : Pour évaluer l'impact de la similarité cosinus dans notre fonction de coût, ce modèle utilise la fonction de coût excluant le terme de la similarité cosinus. L'estimation du vecteur $\vec{v}$ se fait uniquement à travers le terme MSE (voir Équation 6.3).

— *Modèle4* : Contrairement à une approche basée sur les points de repère, ce modèle estime l'orientation des anévrismes en utilisant les trois angles d'Euler normalisés, allant de 0 à 1. Il effectue simultanément la prédiction du centre de chaque anévrisme (C) ainsi que des trois angles correspondants. Cependant, dans ce cas, l'utilisation de l'algorithme NMS (qui est basée sur les sphères) pour filtrer les détections multiples devient impossible. Pour résoudre ce problème, nous avons ajouté une autre convolution en parallèle à la fin de notre CNN pour estimer une valeur de rayon associée à chaque prédiction. Cette valeur de rayon est également optimisée en utilisant la MSE (voir Figure 6.7d).



(a) *Modèle0* (b) *Modèle1* (c) *Modèle2* (d) *Modèle4*

FIGURE 6.7 – Étude par ablation utilisant diverses configurations pour représenter la pose de l'anévrisme.

La base de données du CHRU de Nancy a été utilisée dans cette étude par ablation, avec la même distribution que celle employée dans le chapitre précédent. Pour rappel, l'ensemble d'entraînement comprenait 92 images contenant 138 anévrismes, avec une taille moyenne de 4,11 ± 2,48 mm. L'ensemble de test se composait de 40 images contenant 68 anévrismes, avec une taille moyenne de 3,67 ± 1,93 mm. Tous les modèles ont été entraînés avec les données de l'ensemble d'entraînement. Après 200 époques d'entraînement, leurs performances ont été évaluées sur l'ensemble de test en utilisant les métriques de détection et d'estimation de pose définies dans la section 6.3.3. Les résultats de cette étude sont récapitulés dans la table 6.1.

Modèles	*Détection d'objet*			*Estimation de la pose*			
	AP (%)	Sensibilité (%)	FP/cas	Erreur de centre (mm)		Erreur d'orientation (°)	
				Moyenne	Min − Max	Moyenne	Min − Max
Modèle0	79,13	77,94	0,27	0,57 ± 0,34	0,07 − 1,69	15,26 ± 12,92	1,75 − 63,04
Modèle1	70,25	75,00	0,32	0,58 ± 0,37	0,07 − 1,86	19,32 ± 12,69	2,44 − 72,77
Modèle2	73,78	69,12	0,30	0,59 ± 0,35	0,11 − 1,76	16,29 ± 13,7	1,99 − 80,03
Modèle3	75,66	76,47	0,30	0,58 ± 0,36	0,05 − 1,80	25,75 ± 15,45	0,98 − 66,53
Modèle4	74,11	75,00	0,32	0,61 ± 0,45	0,09 − 1,75	22,14 ± 17,94	4,82 − 89,41

TABLE 6.1 – Étude par ablation sur la tâche d'estimation de la pose des anévrismes en utilisant la base de données du CHRU de Nancy. Les performances pour la tâche de détection ont été ajoutées.

En comparant différents modèles pour les tâches de détection et d'estimation de la pose des anévrismes, nous pouvons formuler plusieurs observations. Les approches adoptées par *Modèle1* et *Modèle2* reposent sur l'identification de points de repère respectivement situés à l'ostium et au dôme des anévrismes. En revanche, notre *Modèle0* se concentre plus sur le centre de l'anévrisme. Notre modèle a présenté une performance de détection supérieure, avec une précision moyenne de 79,13%, une sensibilité de 77,94%, et un taux de 0,27 FP/cas. De plus, il a démontré des erreurs minimales dans la localisation du centre de l'anévrisme, avec une moyenne de 0,57 ± 0,34 mm, ainsi que dans l'estimation de l'orientation, avec une moyenne de 15,26 ± 12,92 degrés. Cependant, il est important de noter que les variations dans les définitions des points de repère par *Modèle1* et *Modèle2* n'ont pas conduit à des améliorations significatives en termes de performance globale. Cette observation suggère que les caractéristiques associées au point central de l'anévrisme, utilisées dans *Modèle0*, sont plus faciles à capturer et à apprendre en comparaison avec les points situés au niveau du dôme et de l'ostium, qui peuvent présenter des variations dans les annotations.

Le *Modèle3*, en excluant l'optimisation directe de l'orientation pendant l'entraînement, a permis d'évaluer l'impact de la similarité cosinus dans la fonction de coût. Les résultats indiquent que l'intégration de la similarité cosinus dans le *Modèle0* améliore l'estimation de l'orientation par rapport au *Modèle3*, qui n'utilise pas cette métrique. De plus, une légère amélioration a été observée en ce qui concerne les scores de la précision moyenne et la sensibilité, ainsi qu'une réduction du nombre de FP/cas. En revanche, le *Modèle4*, qui utilise trois angles d'Euler pour représenter l'orientation des anévrismes en plus de prédire leur centre, montre une baisse de performance en termes d'orientation et une instabilité pendant l'entraînement. Cela indique que cette approche n'est pas aussi efficace pour cette application spécifique. De plus, l'utilisation de cette approche a nécessité des adaptations, à savoir l'ajout d'une couche convolution pour estimer le rayon, pour la rendre adaptée à la tâche de détection des anévrismes. Afin d'obtenir une évaluation plus significative des performances de notre approche proposée avec des mesures de variabilité sur les différentes instances de données, une étude par validation croisée a été réalisée.

6.4.2 Évaluation par validation croisée

Dans le but d'une évaluation plus approfondie, nous avons effectué une validation croisée à 5 plis sur deux bases de données distinctes : CHRU de Nancy et CHUV de Lausanne [80]. La base de données du CHRU de Nancy contient 132 images comprenant 206 anévrismes, tandis que la base du CHUV de Lausanne comprend 270 images avec un total de 164 anévrismes. La répartition des données dans chaque sous-ensemble de la validation croisée reste identique à celle utilisée dans le chapitre précédent. Plus précisément, chaque sous-ensemble de la base du CHRU de Nancy se compose de 25 ou 26 images, tandis que chaque sous-ensemble de la base du CHUV de Lausanne comprend 54 images. Les sous-ensembles du CHUV de Lausanne ont été conçus pour maintenir un équilibre entre les images présentant des anévrismes et celles n'en ayant pas. En ce qui concerne le nombre d'anévrismes et leurs tailles moyennes par sous-ensemble : Pour la base du CHRU, les sous-ensembles comprenaient respectivement 47, 32, 45, 38 et 44 anévrismes, avec des tailles moyennes de 4,15 mm, 3,61 mm, 3,88 mm, 3,85 mm et 4,25 mm. Pour la base du CHUV, les sous-ensembles comptaient respectivement 32, 33, 43, 28 et 28 anévrismes, avec des tailles moyennes de 3,56 mm, 3,19 mm, 4,05 mm, 4,21 mm et 3,70 mm.

Pour chaque base de données, nous avons entraîné 5 modèles en utilisant 4 sous-ensembles pour l'apprentissage et un pour les tests. Chaque modèle de notre méthode a utilisé environ 9 655 patchs

d'entraînement pour la base du CHRU et 7 595 pour celle du CHUV. Nous rappelons qu'en ce qui concerne la base de données du CHUV, les modèles ont été entraînés en utilisant les annotations initiales, tandis que l'évaluation a été effectuée en utilisant à la fois les annotations initiales et les annotations révisées. Nous avons calculé les variations de performance, représentées par l'écart-type entre les ensembles de test, pour chaque pli de test.

Détection des anévrismes

L'utilisation de la même configuration de validation croisée que dans le chapitre précédent nous permet d'inclure notre nouvelle méthode, DeepAnePose, dans la comparaison avec les trois méthodes de référence : SCPM-Net [185], nnDetection [27], nnUNet [135], ainsi que DeepAneDet, la méthode de détection que nous avons proposée et étudiée dans le chapitre précédent. Les résultats de cette comparaison sont présentés dans la table 6.2.

Données	**Méthodes**	**AP (%)**	**Sensibilité (%)**	**FP/case**
CHRU Nancy	DeepAnePose	82,48 ± 6,66	83,01 ± 6,30	0,34 ± 0,11
	DeepAneDet	75,80 ± 7,11	84,76 ± 5,00	0,57 ± 0,17
	SCPM-Net	80,14 ± 6,49	82,03 ± 7,50	0,71 ± 0,14
	nnDetection	80,88 ± 6.33	77,15 ± 8,84	0,68 ± 0,13
	nnUNet	79,45 ± 4,70	76,04 ± 4,19	0,27 ± 0,10
CHUV Lausanne	DeepAnePose	76,60 ± 5,24	82,93 ± 5,92	0,44 ± 0,04
	DeepAneDet	75,80 ± 7,11	84,76 ± 5,00	0,57 ± 0,17
	SCPM-Net	72,41 ± 7,89	85,36 ± 6,18	0,75 ± 0,19
	nnDetection	73,68 ± 6,38	84,76 ± 4,72	0,67 ± 0,12
	nnUNet	72,46 ± 4.74	71,95 ± 9,11	0,13 ± 0,06
	DeepAnePose*	78,51 ± 3,61	83,91 ± 4,77	0,40 ± 0,04
	DeepAneDet*	78,96 ± 5,24	86,78 ± 3,98	0,53 ± 0,14
	SCPM-Net*	76,35 ± 5,91	86,20 ± 5,22	0,71 ± 0,15
	nnDetection*	78,01 ± 6,07	86,78 ± 6,16	0,63 ± 0,10
	nnUNet*	73,47 ± 4,17	71,26 ± 8,09	0,11 ± 0,04

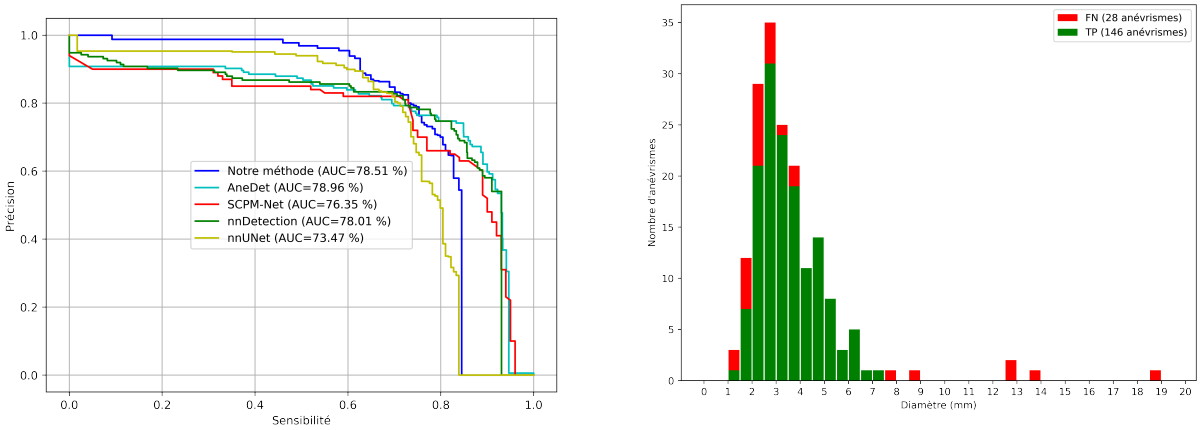
TABLE 6.2 – Résultats (moyenne ± écart-type) de la détection des anévrismes sur les ensembles de données CHRU et CHUV. * indique l'utilisation des annotations révisées de la base CHUV pour l'évaluation. La méthode DeepAneDet fait référence à notre méthode de détection décrite dans le chapitre précédent.

Notre méthode a démontré des performances compétitives par rapport à ces méthodes de référence, notamment DeepAneDet, SCPM-Net, et nnDetection, qui sont spécifiquement conçues pour la tâche de détection d'objets. Sur l'ensemble de données du CHRU, notre méthode a atteint un score de précision moyenne (AP) de 82,48 ± 6,66%, illustrant un équilibre satisfaisant entre la sensibilité et la précision de détection. Lorsqu'un seuil de classification par défaut de 50% a été appliqué, elle a montré une sensibilité de 83,01 ± 6,30%, tout en maintenant un très faible taux de faux positifs à 0,34 ± 0,11 FP/cas.

Concernant l'ensemble de données public du CHUV de Lausanne et en utilisant les annotations initiales, notre méthode a obtenu un score AP de 76,60 ± 5,24%. Elle a ainsi surpassé les performances de SCPM-Net (72,41%), nnDetection (73,68%), et nnUNet (72,46%), tout en présentant des résultats compétitifs par rapport à la méthode DeepAneDet (78,96%). De plus, notre méthode a démontré un meilleur équilibre entre la sensibilité, avec un score de 82,93 ± 5,92%, certes inférieur à celui des méthodes de référence, mais avec un très bas nombre de faux positifs à 0,44 ± 0,04 FP/cas.

Pour une évaluation plus précise, nous avons utilisé les annotations révisées par des experts, telles que détaillées dans l'étude réalisée dans le chapitre précédent (voir section 5.4.3). Il convient de noter que les modèles DeepAneDet et nnDetection sont les mêmes que ceux présentés dans le chapitre précédent, où leurs prédictions avaient été évaluées par des experts, contrairement à notre méthode, dont les prédictions n'avaient pas été évaluées. Notre méthode a obtenu le meilleur score AP de 78,51 ± 3,61% par rapport aux méthodes de référence. Une amélioration au niveau de la sensibilité et du taux de faux positifs a été constatée : la sensibilité est passée de 82,93 à 83,91%, et le taux de faux positifs a diminué de 0,44 à 0,40 FP/cas. Les courbes de Sensibilité-Précision pour les différentes méthodes sont illustrées dans la

Figure 6.8a. Par ailleurs, la sensibilité de notre méthode pour la détection d'anévrismes en fonction de leur taille est représentée dans la Figure 6.8b. Pour les anévrismes de moins de 3 mm, notre sensibilité était de 75,95%, tandis que pour les anévrismes de taille comprise entre 3 et 5 mm, elle a atteint 95,77%, et pour les anévrismes de plus grande taille (> 5 mm), elle s'est établie à 75,0%. Enfin, il est à noter que les écarts-types sont systématiquement plus faibles avec notre méthode, suggérant une meilleure robustesse.



(a) Courbes Précision-Sensibilité.

(b) Détections de notre méthode par taille d'anévrismes.

FIGURE 6.8 – Performances des méthodes dans la détection des anévrismes sur l'ensemble de données du CHUV de Lausanne, en utilisant les annotations révisées. Ici, notre méthode est DeepAnePose.

Estimation de la pose d'anévrismes

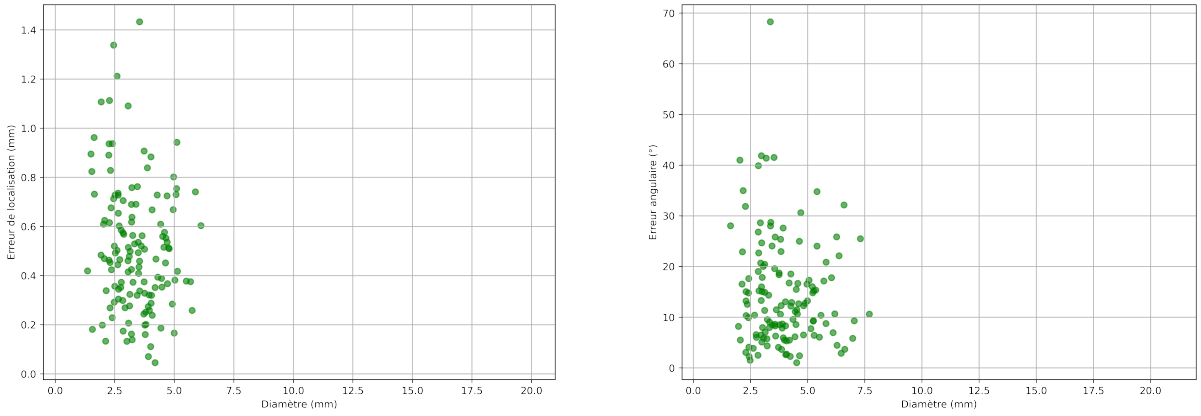
Dans la tâche d'estimation de la pose des anévrismes, il est nécessaire d'évaluer à la fois la précision de la position du centre et de l'orientation. Cependant, aucune méthode à notre connaissance ne résout simultanément ces deux problèmes, ce qui limite la possibilité d'établir une référence pour cette évaluation. Néanmoins, notre méthode de détection, appelée DeepAneDet, que nous avons proposée dans le chapitre précédent, ainsi que la méthode SCPM-Net, fournissent une position pour le centre. Par conséquent, notre méthode a été partiellement validée, pour sa précision de localisation du centre des anévrismes, en la comparant aux méthodes DeepAneDet et SCPM-Net. Nous avons mesuré la distance euclidienne par rapport à la vérité terrain pour évaluer cette précision (en mm). L'évaluation de la qualité de l'estimation de l'orientation a été réalisée en mesurant l'erreur angulaire par rapport à la vérité terrain (en degrés). Les résultats de cette évaluation sur les deux bases de données sont présentés dans la table 6.3. Il convient de noter que nous avons utilisé les mêmes modèles que ceux employés pour la détection des anévrismes.

Données	**Méthodes**	*Erreur de localisation (mm)*			*Erreur d'orientation (°)*		
		Moyenne	**Médiane**	**Min-Max**	**Moyenne**	**Médiane**	**Min-Max**
CHRU	DeepAnePose	0,54 ± 0,32	0,49	0,05 – 1,74	15,26 ± 10,92	11,91	0,21 – 68,35
	DeepAneDet	0,59 ± 0,43	0,49	0,06 – 3,25			
	SCPM-Net	0,57 ± 0,40	0,55	0,11 – 1,97			
CHUV	DeepAnePose	0,51 ± 0,26	0,48	0,05 – 1,43	14,58 ± 10,53	12,27	1,05 – 68,30
	DeepAneDet	0,71 ± 0,53	0,56	0,12 – 2,41			
	SCPM-Net	0,61 ± 0,35	0,61	0,11 – 2,19			
	DeepAnePose*	0,49 ± 0,22	0,45	0,05 – 1,40	13,73 ± 9,88	11,36	1,05 – 68,30
	DeepAneDet*	0,70 ± 0,51	0,56	0,11 – 2,41			
	SCPM-Net*	0,65 ± 0,34	0,58	0,11 – 2,19			

TABLE 6.3 – Évaluation de la précision de l'estimation de la pose de notre méthode sur les ensembles de données CHRU et CHUV. Les modèles évalués en utilisant les annotations révisées sont indiqués par un "*". La méthode DeepAneDet désigne notre méthode de détection décrite dans le chapitre précédent.

Pour la base de données du CHRU de Nancy, notre méthode a présenté une erreur moyenne de localisation du centre de l'anévrisme de 0,54 ± 0,32 mm, avec une médiane de 0,49 mm. En comparaison, DeepAneDet et SCPM-Net ont montré une erreur moyenne légèrement supérieure, respectivement de 0,59 ± 0,43 mm pour DeepAneDet et 0,57 ± 0,40 mm pour SCPM-Net. En ce qui concerne l'orientation de l'anévrisme, notre méthode a enregistré une erreur moyenne de 15,26 ± 10,92°, avec une médiane de 11,91°, et une plage allant de 0,21° à 68,35°. En ce qui concerne l'ensemble de données CHUV, en utilisant les annotations révisées, notre méthode a présenté une erreur moyenne de localisation du centre de l'anévrisme de 0,49 ± 0,22 mm, avec une médiane de 0,45 mm. En comparaison, notre méthode a montré des performances plus importantes par rapport à DeepAneDet et SCPM-Net, qui ont affiché respectivement une moyenne de 0,58 ± 0,40 mm pour DeepAneDet et 0,60 ± 0,34 mm pour SCPM-Net. En ce qui concerne l'orientation de l'anévrisme, notre méthode a enregistré une erreur moyenne de 13,73 ± 9,88°, avec une médiane de 11,36° et une plage allant de 1,05° à 68,30°.

La performance de notre méthode en termes d'erreur de localisation et d'orientation des anévrismes varie selon leur taille, comme illustré dans la Figure 6.9. Pour les petits anévrismes (moins de 3 mm), l'erreur de localisation affiche une médiane de 0,54 mm (moyenne : 0,58 mm, min : 0,13 mm, max : 1,34 mm), tandis que l'erreur angulaire présente une médiane de 14,79° (moyenne : 16,31°, min : 1,52°, max : 41,86°). Concernant les anévrismes de taille moyenne (entre 3 et 5 mm), l'erreur de localisation a une médiane de 0,45 mm (moyenne : 0,46 mm, min : 0,05 mm, max : 1,43 mm) et l'erreur angulaire une médiane de 11,49° (moyenne : 14,03°, min : 1,05°, max : 68.3°). Pour les anévrismes plus grands, l'erreur de localisation montre une médiane de 0,51 mm (moyenne : 0,56 mm, min : 0,26 mm, max : 0,94 mm), et l'erreur angulaire a une médiane de 10,69° (moyenne : 14,19°, min : 2,9°, max : 34,81°).



(a) Erreur de localisation du centre.

(b) Erreur angulaire d'orientation.

FIGURE 6.9 – Les performances de notre méthode d'estimation de la pose des anévrismes sur l'ensemble de données CHUV de Lausanne ont été évaluées. Ces performances ont été calculées sur les anévrismes correctement détectés et reposent sur des annotations révisées.

6.5 Discussion et analyse des résultats

6.5.1 Méthode pour l'estimation de pose et la détection d'objets

L'évaluation de notre méthode a été menée en utilisant deux ensembles de données distincts, tout en veillant à ne pas biaiser notre étude. La base privée du CHRU de Nancy a été employée pour l'étude par ablation, tandis que l'évaluation finale a été réalisée en utilisant la base publique plus large du CHUV de Lausanne. Cette approche permet également d'offrir la possibilité aux futures méthodes de se comparer avec notre méthode de manière équitable.

Pour évaluer la performance de notre méthode en matière de détection d'anévrismes, nous l'avons comparée à trois méthodes de référence, à savoir SCPM-Net [185], nnDetection [27] et nnUNet [135],

ainsi qu'à notre propre méthode de détection, DeepAneDet, présentée dans le chapitre précédent. En utilisant les annotations initiales de la base de données du CHUV de Lausanne, notre méthode a obtenu un score de précision moyen de 76,60 ± 5,24%, accompagné d'une sensibilité de 82,93 ± 5,92% et d'un taux de 0,44 ± 0,04 FP/cas. Même lorsque nous avons utilisé les annotations révisées, notre méthode a maintenu des performances compétitives, atteignant un taux de précision de 78,51 ± 3,61%. De plus, elle a maintenu un bon équilibre entre la sensibilité et le taux de faux positifs par cas. Cependant, il est important de noter que notre méthode présente une limitation en ce qui concerne la détection des anévrismes de grande taille (supérieurs à 7 mm), comme illustré dans la Figure 6.8b. Il est à noter que notre ensemble de données ne contient qu'un nombre très limité de ces grands anévrismes. De plus, cette catégorie d'anévrismes peut être facilement identifiée par les experts.

Pour la tâche d'estimation de la pose, notre méthode fournit une estimation plus précise de la localisation du centre en comparaison des méthodes de références, DeepAneDet (notre méthode du chapitre précédent) et SCPM-Net. Il est important de noter que l'implémentation de la méthode SCPM-Net utilisée bénéficie déjà de la stratégie d'échantillonnage que nous avons proposée dans cette thèse pour converger. De plus, l'erreur moyenne (0,49 mm) et la médiane (0,45 mm) sont très proches de la taille du voxel des données utilisées (0,4 à 0,7 mm pour la base du CHUV de Lausanne). En revanche, l'estimation de la pose semble moins précise, ce qui n'est pas surprenant en comparaison des travaux récents en calcul de pose en vision par ordinateur classique. Étant donné l'absence de méthodes de référence, nous avons sollicité l'avis des médecins partenaires de ce projet pour évaluer visuellement la pose proposée. À cet effet, nous avons développé un plugin pour le logiciel Slicer 3D [92].

6.5.2 Outil de visualisation et d'analyse des anévrismes

Le plugin développé a été conçu pour faciliter la visualisation des anévrismes en utilisant des plans de coupes reformatés adaptés à leurs orientations (voir Figure 6.10). L'outil intégré prend en charge les prédictions générées par les modèles d'apprentissage et propose des fonctionnalités pour manipuler les détections et les plans de coupes reformatés. Pour chaque anévrisme détecté, le plugin offre la possibilité de visualiser non seulement l'anévrisme lui-même, mais aussi un plan de coupe reformaté contenant son axe estimé, dans une vue 3D. Il permet aux utilisateurs de personnaliser divers aspects de l'image, tels que la couleur, le seuillage et la visibilité. De plus, il offre la possibilité d'ajuster le plan de coupe en termes de visibilité, de zoom, ainsi que de rotations autour de l'axe estimé (axe X) et

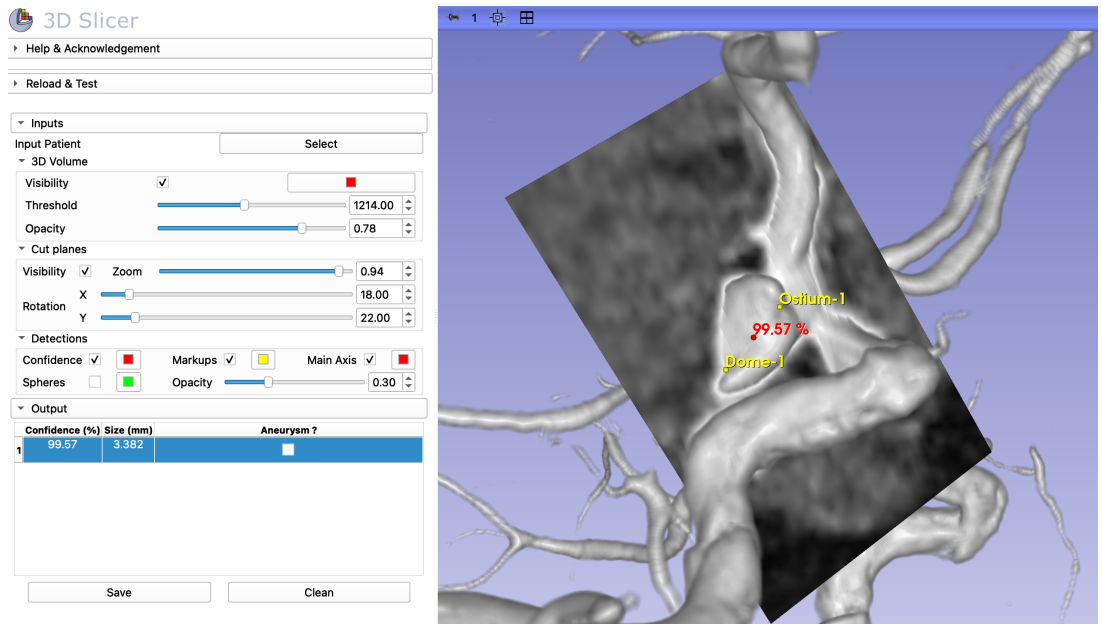
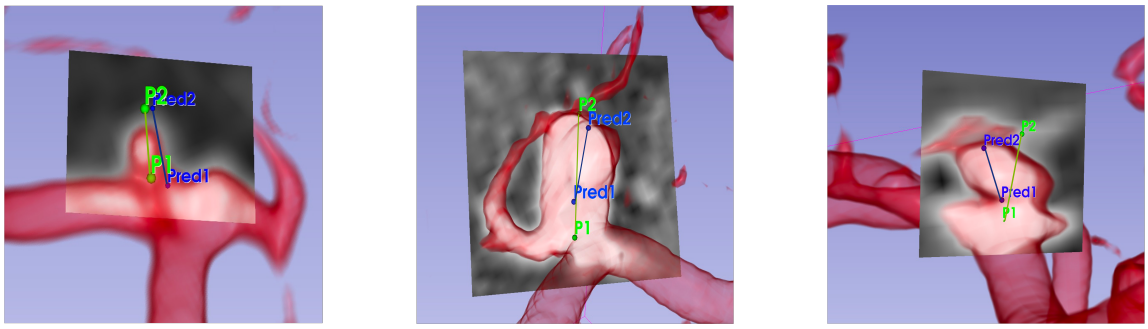


FIGURE 6.10 – Extension du logiciel 3D Slicer développé pour la visualisation de la pose des anévrismes.

autour de la normale au plan (axe Y). Le plugin permet également d'afficher des informations concernant les détections, telles que le score de confiance, les points situés au niveau du collet de l'anévrisme, sa taille et la sphère de détection. Ces fonctionnalités multiples visent à simplifier l'analyse et la visualisation des caractéristiques morphologiques des anévrismes, ainsi que leur localisation par rapport aux structures vasculaires avoisinantes. Ce plugin est accessible publiquement à l'adresse suivante : `https://gitlab.inria.fr/yassis/slicer-reformat-aneurysm`.

6.5.3 Analyse qualitative des résultats

Les exemples illustrés dans la Figure 6.11 montrent que la pose calculée par notre méthode est suffisamment précise pour une utilisation clinique. La Figure 6.11a présente le cas d'un petit anévrisme. La pose a été estimée avec une erreur d'orientation de 8,20° et une erreur de localisation du centre de 0,82 mm. Cette précision, en particulier en ce qui concerne la localisation, permet de déterminer un plan de coupe à travers l'anévrisme adapté à une analyse clinique immédiate. De même, le cas d'un anévrisme plus grand, de forme sphérique, est illustré dans la Figure 6.11b. Notre méthode a estimé la pose avec une erreur d'orientation de 10,62° et une erreur de localisation du centre de 0,72 mm. Des erreurs d'orientation plus importantes ont été observées dans de rares cas, comme l'anévrisme de la Figure 6.11c. Nous avons relié ces erreurs (ici 41,54°) à la forme complexe de l'anévrisme, qui implique une incertitude dans l'annotation de l'orientation de l'axe. Plus généralement, les rares erreurs d'orientation étaient principalement dues à des anévrismes de petite taille et parfois de formes complexes, ce qui a conduit à des annotations de vérité terrain incertaines.



(a) Anévrisme de 1,97 mm
Erreur : 0,82mm/8,2°

(b) Anévrisme de 7,69 mm
Erreur : 0,72mm/10,62°

(c) Anévrisme de 3,52 mm
Erreur : 0,70mm/41,54°

FIGURE 6.11 – Résultats qualitatifs sur l'ensemble de données CHUV de Lausanne : les points prédites (bleu) et les vérités terrain associées (vert). Chaque plan de coupe reformaté a été déterminé en tournant autour de l'axe de l'anévrisme passant par les directions prédites.

Conclusion

Dans ce chapitre, nous avons exposé une méthode novatrice pour l'estimation de la pose des anévrismes à partir d'images TOF-MRA. Cette méthode se distingue par son approche basée sur des points de repère, simplifiant ainsi l'annotation des données et évitant la nécessité de paramétrer la rotation. Elle repose sur un réseau CNN simple, utilisant la similarité cosinus comme fonction de coût pour optimiser l'orientation des anévrismes. Notre méthode représente une avancée significative dans la détection automatique des anévrismes et leur analyse clinique, offrant plusieurs avantages par rapport aux méthodes existantes. Elle permet la localisation, l'estimation de la taille et de l'orientation des anévrismes en une seule étape. Nous avons également proposé une approche simple, accompagnée d'un outil adapté à l'analyse clinique de chaque anévrisme détecté.

Des analyses qualitatives préliminaires révèlent un fort potentiel en termes d'applications cliniques. Malgré les variations dues à l'incertitude de l'annotation, notre méthode a présenté de faibles erreurs, avec une médiane de 0,45 mm pour la localisation des anévrismes et une erreur modérée de 11,36 degrés

pour l'estimation de l'orientation. Pour réduire davantage ces erreurs, une définition clinique plus précise de l'axe de l'anévrisme serait bénéfique, permettant ainsi à notre réseau CNN d'apprendre des caractéristiques plus discriminantes. De plus, notre méthode offre la possibilité d'effectuer une évaluation précise de la pose des anévrismes, du collet au dôme. Pour des analyses plus approfondies, notre méthode peut également être adaptée pour estimer d'autres mesures morphologiques essentielles en pratique clinique, telles que la taille du collet des anévrismes, leur rapport d'aspect et d'autres paramètres significatifs.

Chapitre 7

Conclusion et perspectives

7.1 Résumé des contributions

Cette thèse s'est concentrée sur l'application des méthodes d'apprentissage profond à la détection des anévrismes intracrâniens non rompus, en utilisant des images obtenues par l'angiographie par résonance magnétique en temps de vol (TOF-MRA). Son objectif est de répondre à la demande croissante en matière d'outils automatisés pour le diagnostic de ces pathologies. En effet, l'analyse et l'interprétation manuelles de ces données, en raison de leur complexité, représentent un défi significatif pour les experts cliniciens. De plus, ce besoin est amplifié par l'augmentation continue du volume des données d'imagerie générées dans les établissements de santé. Deux objectifs particuliers ont été posés d'emblée à cette thèse. Tout d'abord améliorer la détection des anévrismes de petite taille (< 5mm). Ceux-ci sont particulièrement difficiles à identifier par les cliniciens, et pourtant essentiels à un dépistage précoce permettant la mise en place d'un suivi médical et une réduction des risques associés à un AVC hémorragique. En second lieu, les méthodes de l'état de l'art en début de thèse souffraient toutes d'un taux de FP/cas élevé, frein important à l'adoption de logiciels de détection automatique par les cliniciens. Nous avions donc en vue de faire baisser ce taux de FP/cas, sans impact négatif sur la sensibilité, qui doit rester de niveau au moins égal à celle des experts. Nous avons pour ce faire apporté des contributions à différents points de la chaîne des traitements mis en œuvre lors de l'exploitation de l'apprentissage profond.

Stratégie de gestion de données efficace

Notre première contribution dans le domaine de la détection des anévrismes porte sur une stratégie de gestion des données spécifiquement élaborée pour traiter les caractéristiques uniques associées à cette pathologie. Ces particularités englobent la disponibilité limitée des données, la rareté des cas d'anévrismes, le déséquilibre des classes, ainsi que la grande diversité des échelles et des formes des anévrismes. Notre stratégie se décompose en trois composantes fondamentales. La première implique l'adoption d'une approche d'annotation approximative mais efficace, ce qui a permis la construction rapide et l'annotation de la base de données du CHRU de Nancy, laquelle regroupe 132 patients. La seconde s'appuie sur l'utilisation de patchs de petite taille, une méthode bien établie qui favorise l'acquisition d'un plus grand nombre d'échantillons, notamment dans des contextes où les données sont limitées, et qui se révèle également pertinente dans des environnements caractérisés par des ressources de calcul restreintes. Enfin, la troisième composante propose des techniques d'échantillonnage et de génération de données spécifiquement adaptées pour compenser le déséquilibre et la rareté des signaux d'anévrismes dans les images médicales. Cela englobe une sélection guidée des échantillons sans anévrismes ainsi que la synthèse d'une variété de formes d'anévrismes.

Afin d'évaluer l'efficacité de notre stratégie de manière indépendante, nous l'avons associée à un réseau de segmentation classique, à travers une étude par ablation et une validation croisée. Les résultats obtenus ont été comparés à ceux de réseaux de segmentation plus sophistiqués, voire spécifiquement dédié à cette tâche. Une problématique d'incertitude dans l'annotation a été identifiée au sein de la base de données du CHRU de Nancy, où l'expert annotateur a atteint une sensibilité de 89,47%, bien qu'il fasse partie

de la même équipe que les experts qui ont établi la vérité terrain (biais possible vers une sensibilité plus grande). Cette situation met en évidence la complexité inhérente à la détection des anévrismes, même pour des experts. Parallèlement, des investigations ont été menées afin d'améliorer la détection des petits anévrismes et la réduction des faux positifs, notamment par l'intégration de connaissances médicales préalables dans les réseaux de neurones. Cependant, les limitations liées à la taille des ensembles de données et les contraintes en termes de ressources de calcul ont entravé la possibilité de réaliser des évaluations plus significatives. Cette démarche a mis en lumière les défis pratiques inhérents à l'application des algorithmes d'apprentissage dans ce domaine spécifique.

Approche de détection des anévrismes

Dans le chapitre 4, nous avons démontré que l'utilisation de réseaux de segmentation pour la détection d'objets nécessite des ajustements spécifiques. Ces réseaux demandent l'application d'opérations de post-traitement (comme le seuillage, le filtrage par taille, volume, localisation, etc.) sur les images prédites en vue d'extraire les détections sous forme de composantes connexes. Ces procédures incluent également la sélection d'un score de confiance parmi plusieurs valeurs de voxels. Cependant, ces opérations peuvent influencer significativement le nombre et les dimensions des détections obtenues. Par exemple, lorsque l'on traite des objets de petite taille, tels que les anévrismes, de légères variations peuvent avoir un impact significatif sur les métriques d'évaluation. Le choix de métriques appropriées est donc essentiel pour une évaluation précise. Dans notre travail, nous avons privilégié des critères basés sur l'intersection sur l'union (IoU). De plus, nous avons également utilisé la métrique de précision moyenne (AP) pour synthétiser les performances globales à différents seuils de classification, en accord avec les travaux actuels sur la détection d'objets en vision classique. Ces aspects, souvent négligés dans les travaux existants, sont importants pour une analyse complète et précise.

Nous avons adopté une approche différente de la plupart des travaux existants pour la détection des anévrismes. Alors que ces derniers s'appuient principalement sur des réseaux de segmentation, nous avons présenté une méthode basée sur un réseau de détection d'objets en une seule étape. Notre méthode consiste à localiser les anévrismes en utilisant des représentations sphériques. Cette méthode, adaptée à leur forme particulière, permet non seulement de localiser les anévrismes, mais aussi d'estimer leur taille. Un aspect clé de notre méthode est l'utilisation d'une approche sans ancres, simplifiant ainsi le processus de détection. Pour évaluer l'efficacité de notre approche, nous avons réalisé une analyse exhaustive en utilisant deux ensembles de données importants : les données privées du CHRU de Nancy et les données publiques du CHUV de Lausanne, publiées en fin 2022 [80]. Un effort a été fait pour annoter cette dernière en suivant notre approche d'annotation.

Notre évaluation comprenait plusieurs étapes pour assurer son efficacité et sa fiabilité. Nous avons commencé par une étude par ablation et une validation croisée à 5 plis. Ces étapes ont permis d'analyser l'influence de chaque composant de notre approche et de réaliser une évaluation approfondie de sa performance. Par la suite, une validation a été menée par deux experts cliniques en utilisant les données du CHUV de Lausanne. Au cours de cette évaluation, des défis liés à l'incertitude des annotations ont été confirmés, l'annotateur, cette fois-ci extérieur, atteignant une sensibilité de 82,7%. Ce taux est comparable, et même légèrement inférieur à celui obtenu par les méthodes de détection automatisée. Notre méthode, par exemple, a montré une sensibilité de 86,78% avec un taux de 0,53 FP/cas. Ces méthodes ont démontré leur efficacité à identifier les caractéristiques discriminantes des anévrismes, et ce, même en présence de bruit dans les annotations des données d'entraînement. Il a été démontré que les méthodes de détection automatique ont des performances compatibles avec un usage clinique, bien que la gestion du taux de faux positifs demeure un défi à relever.

Estimation de la pose des anévrismes

En plus de la détection des anévrismes, nous avons introduit dans le chapitre 6 une approche novatrice pour estimer la pose des anévrismes à partir d'images TOF-MRA. Cette méthode ne se limite pas à la localisation des anévrismes, mais englobe également l'estimation de leur orientation. Cette double fonctionnalité offre aux radiologues un outil pour la visualisation et l'analyse des anévrismes dans un plan de coupe reformaté immédiatement disponible et adapté à chaque angioarchitecture alors qu'aujourd'hui un tel plan n'est pas utilisé car sa détermination est trop chronophage. Des plans sub-optimaux, et parfois trompeurs, sont donc employés. Notre méthode d'annotation est également adaptée à cet objectif,

permettant une annotation rapide et simple de la position et de l'orientation de chaque anévrisme. Fondée sur un réseau CNN en une seule étape, notre modèle est optimisé par une fonction de coût impliquant la similarité cosinus. Elle a été adaptée et évaluée pour les tâches de détection et d'estimation de la position, démontrant des performances prometteuses dans les deux domaines, sur les deux ensembles de données. Les erreurs constatées sont faibles, avec une erreur médiane de 0,45 mm pour la localisation du centre des anévrismes, proche de la résolution des images IRM, ou modérées avec une erreur de 11,36° pour l'orientation. Les résultats qualitatifs sont encourageants et suggèrent une utilité clinique de notre méthode.

En parallèle à notre recherche sur les réseaux de neurones, cette thèse a conduit au développement de plusieurs plugins du logiciel Slicer 3D [92]. Ces outils comprennent un dispositif pour l'annotation des anévrismes, utilisant soit des sphères, soit des boîtes englobantes, un outil permettant aux experts d'évaluer les détections, ainsi qu'un outil de visualisation et de manipulation de plans de coupes reformatés, adaptés à la pose des anévrismes. Dans un esprit de partage scientifique, le code source de ces outils est rendu public et disponible à l'adresse suivante : `https://gitlab.inria.fr/yassis`. Cette initiative vise à surmonter la limitation des ressources rencontrée au début de la thèse et à faciliter l'accès aux outils développés. Elle encourage leur utilisation et leur amélioration par la communauté scientifique, contribuant ainsi au progrès dans les domaines de la détection et de l'analyse des anévrismes.

Ce site met également à disposition les modèles CNN développés dans le cadre de cette thèse, ainsi que la vérité terrain établie par nos deux experts pour la base publique du CHUV de Lausanne.

7.2 Perspectives futures

L'avancement des méthodes automatisées pour la détection des anévrismes présente un potentiel considérable en vue d'améliorer les procédures de diagnostic actuelles. Ces techniques ont le pouvoir d'accroître l'efficacité des processus médicaux tout en réduisant les biais associés à l'observation humaine. L'intérêt pour de telles approches est renforcé par l'anticipation d'une augmentation du dépistage préventif des anévrismes. Toutefois, pour que ces méthodes puissent être intégrées dans la pratique clinique, elles doivent surpasser la sensibilité diagnostique des professionnels de la santé, tout en maintenant un taux acceptable de faux positifs. Nos travaux de recherche ont démontré que les méthodes récentes atteignent des taux de sensibilité comparables à ceux des experts humains. Toutefois, les principaux défis identifiés dans nos recherches sont liés à la limitation des ensembles de données et à la fiabilité des données de référence, notamment en ce qui concerne le problème de l'incertitude des annotations. Ces défis restreignent à la fois la capacité d'entraînement des réseaux de neurones et la validité des évaluations. Même si la même tendance s'observe de manière consistante sur les valeurs moyennes des métriques (AP, sensibilité, taux de FP/cas, ...) les écarts-types estimés par validation croisée sont trop grands pour prétendre en général à une quelconque significativité de la comparaison des méthodes explorées.

Nos diverses investigations à ce sujet nous font d'ailleurs remettre sérieusement en doute la validité scientifique des résultats du challenge ADAM, qui permet selon nous au mieux de classer les méthodes en 3 grandes catégories, mais n'autorise par la comparaison fine qui a donné lieu au tableau de classement final. Ceci remet en question la fiabilité de l'évaluation des méthodes existantes. Ces considérations ouvrent donc des perspectives selon trois axes : la fiabilité des évaluations, à travers l'amélioration et l'extension des jeux de données ; l'amélioration des performances des réseaux de détection, en s'attaquant plus spécifiquement au problème de rareté des anévrismes, en particulier par des techniques de synthèse de données plus avancées ; et l'extension de nos travaux sur la détection à des analyses plus sophistiquées, afin de mesurer l'anévrisme et prédire son risque de rupture, améliorant ainsi sa prise en charge, y compris thérapeutique.

Amélioration et extension des ensembles de données

L'amélioration et l'élargissement des ensembles de données jouent un rôle crucial dans la fiabilité des méthodes de détection des anévrismes. La plupart des méthodes existantes sont restreintes à quelques centaines de patients et proviennent souvent de bases de données privées contenant quelques dizaines d'anévrismes. L'accroissement de la taille des ensembles de données offre la possibilité d'effectuer des évaluations plus solides et significatives. Renforcer la collaboration avec des professionnels de la santé

est essentielle. Il est ainsi impératif d'intensifier les efforts de collecte et d'annotation des données. Des initiatives en ce sens sont en cours, à l'instar du projet U-CAN [4] au CHU de Nantes, qui se focalise sur la prédiction de l'évolution des anévrismes intracrâniens. Encore plus récemment, le registre FRAN, soutenu par la Fondation des Spécialités Médicales [5] vient d'être initié afin de recueillir les informations médicales concernant tous les anévrismes, rompus ou non, traités ou non, dans tous les CHU de France. Un registre de pratique médicale en première instance, mais d'ores et déjà prévu pour alimenter la recherche dans le domaine.

Dans ce contexte, notre méthode d'annotation, caractérisée par sa rapidité et sa simplicité de mise en œuvre, peut être employée pour augmenter la quantité de données disponibles. Afin d'améliorer la fiabilité des annotations et d'augmenter le nombre de données d'apprentissage, il est envisageable de faire intervenir plusieurs experts pour l'annotation de chaque anévrisme, afin de mieux modéliser les variabilités intra- et inter-opérateurs. La première conséquence serait de pouvoir associer chaque mesure de performance à une incertitude, et ainsi prendre une décision statistiquement fondée quant à la supériorité d'une méthode, et/ou l'apport réel de tel ou tel composant d'un modèle. Les données annotées par plusieurs experts peuvent ensuite être utilisées pour former des réseaux de neurones, afin d'apprendre à modéliser la variabilité des annotations entre différents annotateurs [316]. Un travail de thèse a d'ailleurs été initié dans le sillon de notre travail, par un praticien hospitalier de Nancy, avec deux objectifs principaux. En premier lieu, mieux caractériser un anévrisme intracrânien, mieux cerner et quantifier les performances effectives des experts dans leur détection, préciser certains aspects morphologiques, tel que ce que nous avons appelé leur axe principal dans ce travail. En second lieu, mieux comprendre ce qui fait la variabilité d'annotation, en incluant la notion de choix médical, différente de celle de bruit d'annotation, et donc à prendre différemment en compte dans l'apprentissage d'un réseau de neurones.

Ces démarches soulignent l'importance d'une approche collaborative dans la collecte et l'annotation des données, contribuant à améliorer la précision des annotations tout en enrichissant la quantité et la diversité des ensembles de données disponibles.

Rareté des anévrismes et synthèse de données

La rareté des anévrismes dans les images médicales, associée à leur petite taille et à leur faible signal, pose un défi considérable pour l'entraînement efficace des modèles d'apprentissage profond. Pour surmonter cet obstacle, nous avons utilisé une stratégie de synthèse d'images, qui consiste à créer des formes variées d'anévrismes à partir de cas existants. Cette stratégie permet d'augmenter artificiellement la diversité et le volume des données d'entraînement sans compromettre leur pertinence clinique. Néanmoins, il est apparu qu'une synthèse excessive d'anévrismes basée sur des cas existants conduit à une stagnation dans l'amélioration des performances. Comme perspective, une autre approche prometteuse implique la synthèse de formes d'anévrismes à partir de vaisseaux sains, notamment au niveau des sites fréquents pour le développement d'anévrismes. Cette méthode nécessite la modélisation de structures vasculaires synthétiques basées sur des données réelles. Les atlas statistiques de vaisseaux jouent un rôle essentiel dans cette démarche [207]. Ces atlas aident à identifier les régions vasculaires les plus propices au développement de ces anomalies, comme le polygone de Willis [208]. Leur utilisation permet de cibler spécifiquement ces zones à risque pour une création plus pertinente de données vasculaires synthétiques.

Une alternative, non exclusive, consisterait à augmenter l'image d'un patient sain ou sans pathologie sur un site vasculaire donné, par l'image de l'anévrisme d'un autre patient. Des algorithmes de recalage déformable, exploitant éventuellement un atlas tel que mentionné ci-avant pour identifier les site d'implantation, permettraient cette augmentation et ainsi l'implantation d'un anévrisme donné dans des paysages angio-architecturaux variés, avec pour conséquence une meilleure discrimination entre les anévrismes et les vaisseaux sanguins. Néanmoins, la validation de la fiabilité de ces données synthétiques est d'une grande importance. Des études sont nécessaires pour développer des méthodes d'évaluation rigoureuses, assurant que les données synthétiques sont réalistes et utiles pour l'entraînement des modèles d'apprentissage profond. En outre, l'acquisition de données à grande échelle permet de garantir la représentativité et la fiabilité des ensembles de données.

4. `https://ichgcp.net/fr/clinical-trials-registry/NCT04578808`
5. voir `https://specialitesmedicales.org/registres-heberges/les-registres-heberges-sur-la-plateforme/`

Analyses morphologiques des anévrismes

L'analyse morphologique des anévrismes est une composante essentielle de leur évaluation qualitative. Cette analyse englobe plusieurs paramètres, notamment la taille de l'anévrisme, le rapport de taille par rapport au vaisseau parent, le rapport d'aspect, ainsi que le volume de l'anévrisme [34]. Notre méthode d'estimation de la pose permet non seulement de déterminer l'axe principal de l'anévrisme, mais aussi d'estimer sa taille. Elle peut être étendue pour mesurer d'autres aspects morphologiques des anévrismes. Toutefois, il convient de souligner que la fiabilité de ces mesures dépend largement de la modalité d'imagerie utilisée. Aujourd'hui, l'IRM ne permet d'effectuer au mieux que des mesures grossières, aidant dans la décision de poursuivre la prise en charge d'un anévrisme par un traitement endovasculaire ou un simple suivi. Lors du traitement, l'angiographie rotationnelle tridimensionnelle (3DRA) est considérée comme une modalité de référence en raison de sa précision et de sa haute résolution [9]. Les mesures de la forme de l'anévrisme, décrivant la taille du matériel à utiliser (par exemple diamètre d'enroulement et longueur des coils) sont donc effectuées sur cette imagerie. La segmentation de l'anévrisme ne fait selon sens d'un point de vue clinique que dans un volume 3DRA. Des techniques avancées de segmentation, telles que le Segment Anything Model (SAM) intégrées dans des interfaces interactives destinées aux radiologues pourrait optimiser le processus d'analyse des anévrismes [129, 189].

Mais des modèles de régression des mesures codifiées seraient aussi sans doute d'un grand intérêt à développer, dans la suite de que nous avons fait pour la pose, et donc l'axe principal de l'anévrisme. En perspective plus lointaine, se trouve l'estimation du risque de rupture d'un anévrisme découvert de manière fortuite. Si les aspects quantitatifs de la morphologie du sac anévrismal sont essentiels dans cette estimation, ils ne suffisent cependant pas. D'autres paramètres, biologiques, cliniques, historiques, génétiques, non exprimés dans l'image diagnostique, sont nécessaires à agréger pour cette prise de décision. Des modèles d'analyse de données hétérogènes, intégrant, au-delà de l'analyse radiomique, directement l'image ou sa représentation latente générée par réseau de neurones, devront être développés en vue de cet objectif.

Annexe A

Outil de vérification des détections d'anévrismes

Vérification des détections d'anévrismes par experts cliniques

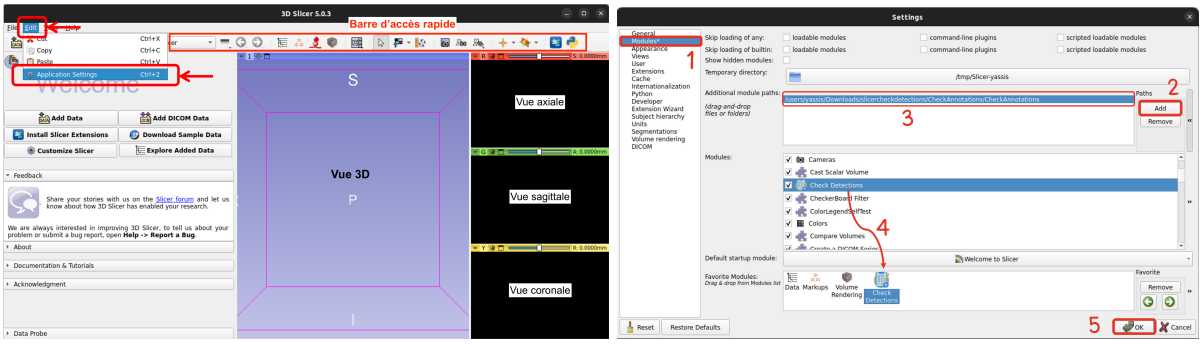
Cette annexe présente un guide d'utilisation pour le plugin du 3D Slicer spécialement conçu pour la validation des détections d'anévrismes réalisées par des algorithmes d'apprentissage et évaluées par des experts. Chaque détection est représentée sous la forme d'une sphère, caractérisée par ses paramètres : son centre, son diamètre, et un score de confiance indiquant la probabilité qu'il s'agisse d'un véritable anévrisme. L'objectif principal de cette démarche est d'évaluer chaque détection, de déterminer sa localisation au sein de l'arbre vasculaire (Acom, PCom, A1, M1, PCA, etc.) et de caractériser sa morphologie (artefact, bruit, post-coiling, dysplasie, etc.) dans le cas des détections incorrectes.

Le présent document est structuré en deux parties principales : la première partie décrit les procédures d'installation et de configuration de l'outil dans le logiciel 3D Slicer, tandis que la seconde partie est dédiée à l'exploration des fonctionnalités spécifiques de l'outil. Elle fournit des instructions sur l'utilisation de ses différents composants pour l'analyse des détections d'anévrismes.

Configuration et intégration de l'outil dans le logiciel 3D Slicer

La procédure initiale consiste à installer le logiciel 3D Slicer sur votre ordinateur. Après l'installation, notre outil d'outil peut être télecharger depuis l'adresse `https://gitlab.inria.fr/yassis/slicercheckdetections`. Une fois téléchargé, décompressez ce fichier et placez le dossier extrait dans un répertoire de votre choix, comme votre dossier `Documents`.

Afin d'accéder à l'outil à partir de l'interface graphique du logiciel 3D Slicer, il est impératif de configurer préalablement le chemin d'accès au répertoire de l'outil dans les paramètres du logiciel. Les étapes requises pour exécuter cette configuration sont exposées dans la Figure A.1. Après avoir cliqué sur le bouton `Add`, il faut naviguer successivement dans les répertoires "slicercheckdetections-v1.0.0", "CheckAnnotations", puis à nouveau "CheckAnnotations". Ensuite, il convient de choisir l'option `Choose`. Il est important de noter que le chemin spécifique indiqué à l'étape 3 dépendra de l'endroit où le fichier zip a été décompressé. Pour un accès rapide à l'outil via la barre d'accès rapide de l'application (tel que illustré dans la Figure A.1a), une alternative consiste simplement à faire glisser l'outil dans la liste des outils favoris, conformément à l'étape 4 de la Figure A.1b. Dans certains cas, il peut être nécessaire de redémarrer l'application si l'icône de l'outil n'apparaît pas immédiatement dans la barre d'accès rapide.



(a) Accès aux paramètres du logiciel. (b) Ajout du chemin de l'outil dans les paramètres.

FIGURE A.1 – Configuration de l'outil dans le logiciel 3D Slicer.

Une fois que la configuration est terminée, l'utilisateur a deux options pour accéder à l'interface principale de l'outil : soit en utilisant la barre d'accès rapide, identifiée par la flèche verte dans la Figure A.2, soit en effectuant une recherche de l'outil dans la liste des outils disponibles, en suivant les étapes 1 et 2 telles que indiquées dans la Figure A.2. Pour cela, les utilisateurs doivent d'abord cliquer sur le bouton "Bienvenue dans Slicer," puis sélectionner l'option "Check Detections".

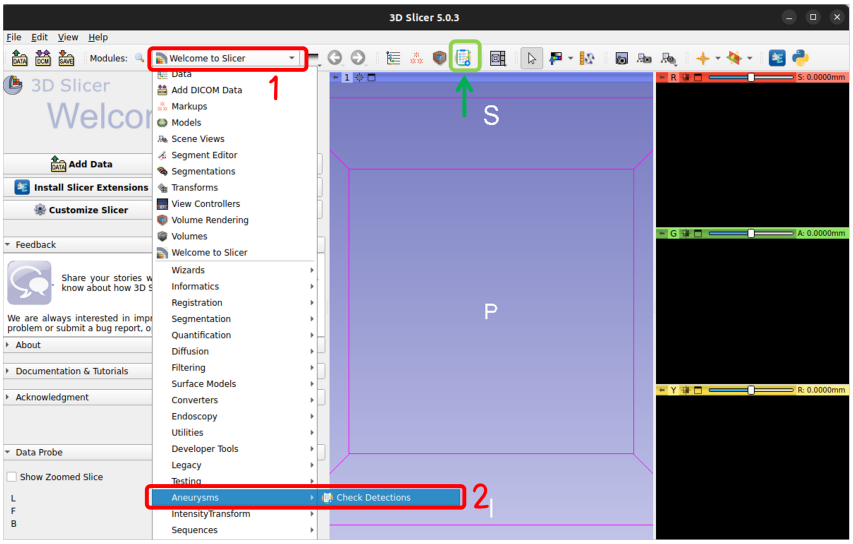


FIGURE A.2 – Deux méthodes pour accéder à l'outil depuis l'interface principale du logiciel 3D Slicer.

Composantes et fonctionnalités de l'outil

L'objectif de cet outil est de faciliter la vérification des détections d'anévrismes pour un ensemble de patients. À cette fin, les données sont organisées par patient, chacun correspondant à un répertoire avec deux types de fichiers distincts : un fichier de volume 3D au format (*.nii.gz*) et un fichier au format (*.json*) contenant les détections générées par le modèle. Cette structure permet une accessibilité aisée aux données de chaque patient dans le cadre du processus de vérification (voir la figure A.3). Il est à noter que dès qu'un patient est soumis à une vérification par l'utilisateur, il est automatiquement déplacé vers un nouveau répertoire nommé "Done", situé au sein du même répertoire principal. Cette démarche vise à maintenir l'ordre des données et à signaler quels patients ont déjà été soumis à une vérification par l'utilisateur. De plus, il est également possible de charger les données des patients précédemment vérifiés : les décisions prises par l'utilisateur seront automatiquement chargées et pourront être modifiées.

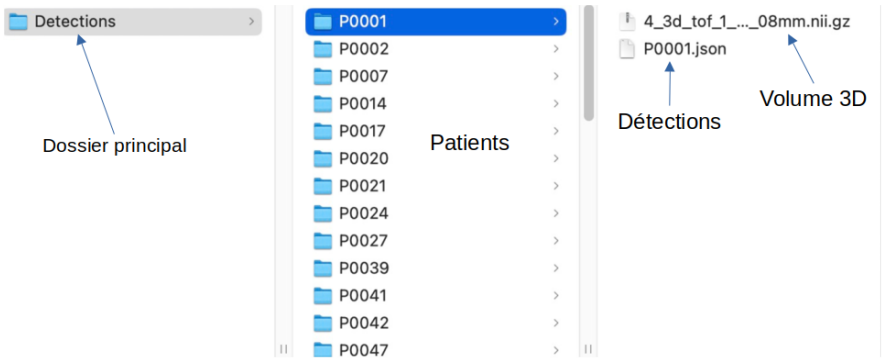


FIGURE A.3 – Structure du dossier des détections à vérifier.

La Figure A.4 présente l'interface principale de l'outil, qui se divise en deux sections distinctes. La première section est réservée au chargement et à la manipulation des données d'entrée, notamment le volume 3D et les détections. En contraste, la seconde section est principalement consacrée à la visualisation et à la vérification des détections de manière individuelle. Chaque composant de l'interface est numéroté dans la Figure, et son fonctionnement est expliqué en détail ci-après.

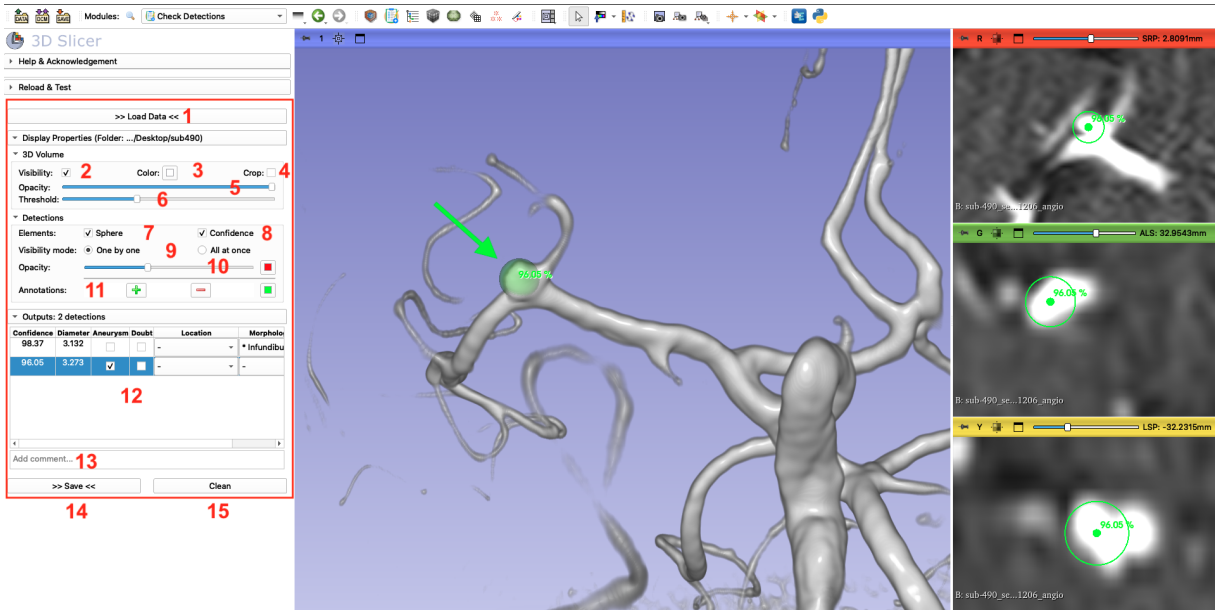


FIGURE A.4 – Composants et fonctionnalités de l'outil de vérification des détections.

1. Le bouton `Load Data` consiste à importer les données d'un patient en sélectionnant le répertoire correspondant (par exemple, P0001, comme illustré dans la figure A.3). Une fois cette opération effectuée, le module procédera automatiquement au chargement des détections (si disponibles) et affichera le volume 3D du patient dans les vues 3D, axiale, sagittale et coronale.
2. La case à cocher "Visibility" permet de réguler la visibilité du volume dans la vue 3D. En activant ou désactivant cette case, il est possible d'afficher ou de masquer le volume dans la vue 3D, une

opération qui peut se révéler bénéfique pour une visualisation optimale des détections.

3. Modifie la couleur du volume dans la vue 3D.
4. Limite l'affichage du volume 3D au voisinage de la détection choisie, au lieu de visualiser l'ensemble de l'arbre vasculaire.
5. Régule l'opacité du volume dans la vue 3D.
6. Le curseur "Threshold" permet d'appliquer un seuillage à la vue 3D, pour mieux visualiser les vaisseaux sanguins.
7. Active ou désactive la visibilité des détections. Par défaut, aucune détection n'est affichée, ce qui permet dans un premier temps de diagnostiquer des anévrismes sans préjuger des détections. Cette approche facilite également la localisation des anévrismes non détectés, le cas échéant.
8. Affiche ou masque les scores de confiance des détections.
9. Gère l'affichage des détections : individuellement ou en affichant la totalité des détections.
10. Régule l'opacité et la couleur des détections dans la vue 3D. Par défaut, toutes les détections sont en rouge, mais en cas de confirmation d'une détection ou d'annotation d'un nouvel anévrisme, sa couleur par défaut devient verte, ce qui peut être contrôlé dans l'option 11.
11. En cas d'absence d'une proposition d'anévrisme dans la table, le symbole "+" permet d'ajouter des points pour l'annotation (chaque anévrisme est annoté par deux points). Une fois l'annotation réalisée, la détection sera intégrée automatiquement dans la table, ce qui permet de visualiser son diamètre et de sélectionner sa morphologie.
12. La table affiche les informations relatives à chaque détection. Pour examiner individuellement chaque détection, l'utilisateur doit sélectionner la ligne correspondant à la détection en question. Cela entraîne automatiquement la suppression des autres détections de la vue 3D, tandis que les trois autres vues (axiale, sagittale, coronale) se concentrent sur la détection concernée. Après avoir vérifié cette détection, l'utilisateur doit déterminer si elle constitue effectivement un anévrisme réel ou non. Dans le cas d'un anévrisme réel, il sera automatiquement coloré en vert, tandis qu'un faux positif sera affiché en rouge. Cette classification s'effectue en cochant la case correspondante dans la colonne "Anévrisme". En cas d'incertitude concernant la détection proposée, l'utilisateur peut cocher la case dans la colonne "Doute". En outre, la colonne "Localisation" permet de spécifier la position de la détection au sein de l'arbre vasculaire, tandis que la colonne "Morphologie" est utilisée pour les détections erronées.
13. Un champ de texte libre permet d'ajouter des notes concernant les données du patient en question, par exemple la qualité de l'images. Ces notes seront enregistrées dans un fichier intitulé *description.txt* pour être consultées et analysées ultérieurement.
14. Une fois que l'utilisateur a examiné toutes les détections proposées, il est impératif de sauvegarder ses décisions afin de les préserver. Un nouveau fichier nommé *_output.json* sera généré dans le répertoire du patient. Il est essentiel de noter que l'utilisateur peut à tout moment recharger les données d'un patient qu'il avait préalablement vérifiées. Ses décisions seront automatiquement restaurées et affichées dans la table, ce qui simplifie la tâche de vérification. Cette fonctionnalité se révèle utile lorsque l'utilisateur souhaite réexaminer ses choix antérieurs ou les mettre à jour si nécessaire.
15. Le bouton "Clean" permet de réinitialiser l'outil et supprimer le volume 3D et les détections.

Protocole d'évaluation

Pour chaque patient :

1. Chargez les données du patient en cliquant sur le bouton `Load`.
2. Ajustez la visibilité de l'arbre vasculaire sur la vue 3D en modifiant le seuil et l'opacité.
3. Effectuez une première analyse pour repérer les emplacements potentiels des anévrismes.

4. Activez la visibilité des détections en cochant la ligne de la détection dans la table. **Il se peut qu'aucune détection ne soit présente**. Ce peut être le cas si la détection a échoué ou simplement parce que la base contient des patients sains.
5. Examinez chaque détection dans la table, en notant s'il s'agit d'un anévrisme ou non, ou si vous avez des doutes.
6. Ajoutez une annotation en cas de non-détection d'anévrisme.
7. Enregistrez les données.
8. Passez au patient suivant et répétez le processus à partir de l'étape 1.

Bibliographie

[1] BRATS Challenge 2023 - Cluster of Challenges. `https://www.med.upenn.edu/cbica/brats/`. Accessed : November 27, 2023.

[2] CHUV Lausanne. `https://www.chuv.ch/fr/rad/rad-home/patients-et-familles/nos-interventions`. Service de radiodiagnostic et radiologie interventionnelle.

[3] Front-matter. In Mark S. Nixon and Alberto S. Aguado, editors, *Feature Extraction & Image Processing for Computer Vision (Third Edition)*. Academic Press, Oxford, third edition edition, 2012.

[4] Byung Bok Ahn. The compact 3D convolutional neural network for medical images. *Standford University*, 2017.

[5] Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L Rubin, and Bradley J Erickson. Deep learning for brain MRI segmentation : state of the art and future directions. *Journal of digital imaging*, 30 :449–459, 2017.

[6] Rustam Al-Shahi and Charles Warlow. A systematic review of the frequency and prognosis of arteriovenous malformations of the brain in adults. *Brain*, 124(10) :1900–1926, 2001.

[7] Klaus Kaae Andersen, Tom Skyhøj Olsen, Christian Dehlendorff, and Lars Peter Kammersgaard. Hemorrhagic and ischemic strokes compared : stroke severity, mortality, and risk factors. *Stroke*, 40(6) :2068–2072, 2009.

[8] Michele Andreucci, Richard Solomon, Adis Tasanarong, et al. Side effects of radiographic contrast media : pathogenesis, risk factors, and prevention. *BioMed research international*, 2014, 2014.

[9] René Anxionnat, Serge Bracard, Xavier Ducrocq, Yves Trousset, Laurent Launay, Erwan Kerrien, Marc Braun, Regis Vaillant, Francesco Scomazzoni, Ariel Lebedinsky, et al. Intracranial aneurysms : clinical value of 3D digital subtraction angiography in the therapeutic decision and endovascular treatment. *Radiology*, 218(3) :799–808, 2001.

[10] Hidetaka Arimura, Qiang Li, Yukunori Korogi, Toshinori Hirai, Hiroyuki Abe, Yasuyuki Yamashita, Shigehiko Katsuragawa, Ryuji Ikeda, and Kunio Doi. Automated computerized scheme for detection of unruptured intracranial aneurysms in three-dimensional magnetic resonance angiography1. *Academic radiology*, 11(10) :1093–1104, 2004.

[11] Hidetaka Arimura, Qiang Li, Yukunori Korogi, Toshinori Hirai, Shigehiko Katsuragawa, Yasuyuki Yamashita, Kazuhiro Tsuchiya, and Kunio Doi. Computerized detection of intracranial aneurysms for three-dimensional MR angiography : Feature extraction of small protrusions based on a shape-based difference image technique. *Medical physics*, 33(2) :394–401, 2006.

[12] Sylvain Arlot and Alain Celisse. A survey of cross-validation procedures for model selection. 2010.

[13] Adrián Arnaiz Rodríguez. Deep convolutional autoencoders for reconstructing magnetic resonance images of the healthy brain. 2021.

[14] Adam S Arthur, Andy Molyneux, Alexander L Coon, Isil Saatci, Istvan Szikora, Feyyaz Baltacioglu, Ali Sultan, Daniel Hoit, Josser E Delgado Almandoz, Lucas Elijovich, et al. The safety and effectiveness of the Woven EndoBridge (WEB) system for the treatment of wide-necked bifurcation aneurysms : final 12-month results of the pivotal web intrasaccular therapy (WEB-IT) study. *Journal of neurointerventional surgery*, 11(9) :924–930, 2019.

[15] Youssef Assis, Liang Liao, Fabien Pierre, René Anxionnat, and Erwan Kerrien. An efficient data strategy for the detection of brain aneurysms from MRA with deep learning. In *Deep Generative Models, and Data Augmentation, Labelling, and Imperfections*, pages 226–234. Springer, 2021.

[16] Youssef Assis, Liang Liao, Fabien Pierre, René Anxionnat, and Erwan Kerrien. Aneurysm Pose Estimation with Deep Learning. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2023 : 26th International Conference, Vancouver, Canada, October 8–October 12, 2023, Proceedings*. Springer, 2023.

[17] Alexi Assmus. Early history of X rays. *Beam Line*, 25(2) :10–24, 1995.

[18] Mathieu Aubry, Daniel Maturana, Alexei A Efros, Bryan C Russell, and Josef Sivic. Seeing 3D chairs : exemplar part-based 2D-3D alignment using a large dataset of cad models. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3762–3769, 2014.

[19] Jongduk Baek, Angel R Pineda, and Norbert J Pelc. To bin or not to bin ? the effect of CT system limiting resolution on noise and detectability. *Physics in Medicine & Biology*, 58(5) :1433, 2013.

[20] Merih I Baharoglu, Clemens M Schirmer, Daniel A Hoit, Bu-Lang Gao, and Adel M Malek. Aneurysm inflow-angle as a discriminant for rupture in sidewall cerebral aneurysms : morphometric and computational fluid dynamic analysis. *Stroke*, 41(7) :1423–1430, 2010.

[21] Daniel Balouek, Alexandra Carpen Amarie, Ghislain Charrier, et al. Adding virtualization capabilities to the Grid'5000 testbed. In Ivan I. Ivanov, Marten van Sinderen, Frank Leymann, and Tony Shan, editors, *Cloud Computing and Services Science*, volume 367 of *Communications in Computer and Information Science*, pages 3–20. Springer International Publishing, 2013.

[22] Andriy I Bandos, Howard E Rockette, Tao Song, and David Gur. Area under the free-response ROC curve (FROC) and a related summary index. *Biometrics*, 65(1) :247–256, 2009.

[23] Isaac Bankman. *Handbook of medical image processing and analysis*. Elsevier, 2008.

[24] Ana Barragán-Montero, Umair Javaid, Gilmer Valdés, Dan Nguyen, Paul Desbordes, Benoit Macq, Siri Willems, Liesbeth Vandewinckele, Mats Holmström, Fredrik Löfman, et al. Artificial intelligence and machine learning for medical imaging : A technology review. *Physica Medica*, 83 :242–256, 2021.

[25] Christian F Baumgartner, Lisa M Koch, Marc Pollefeys, and Ender Konukoglu. An exploration of 2D and 3D deep learning techniques for cardiac MR image segmentation. In *Statistical Atlases and Computational Models of the Heart. ACDC and MMWHS Challenges : 8th International Workshop, STACOM 2017, Held in Conjunction with MICCAI 2017, Quebec City, Canada, September 10-14, 2017, Revised Selected Papers 8*, pages 111–119. Springer, 2018.

[26] M. Baumgartner, P.F. Jaeger, F. Isensee, et al. Retina U-Net for Aneurysm Detection in MR Images. In *Automatic Detection And SegMentation Challenge (ADAM)*, 2020.

[27] Michael Baumgartner, Paul F Jäger, Fabian Isensee, and Klaus H Maier-Hein. nnDetection : A self-configuring method for medical object detection. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 530–539. Springer, 2021.

[28] Thomas Benner, Jonathan J Wisco, André JW van der Kouwe, Bruce Fischl, Mark G Vangel, Fred H Hochberg, and A Gregory Sorensen. Comparison of manual and automatic section positioning of brain MR images. *Radiology*, 239(1) :246–254, 2006.

[29] Jorge Bernal, Nima Tajkbaksh, Francisco Javier Sanchez, Bogdan J Matuszewski, Hao Chen, Lequan Yu, Quentin Angermann, Olivier Romain, Bjørn Rustad, Ilangko Balasingham, et al. Comparative validation of polyp detection methods in video colonoscopy : results from the MICCAI 2015 endoscopic vision challenge. *IEEE transactions on medical imaging*, 36(6) :1231–1249, 2017.

[30] Leonard Berrada, Andrew Zisserman, and M Pawan Kumar. Smooth loss functions for deep top-k classification. *arXiv preprint arXiv :1802.07595*, 2018.

[31] Jeroen Bertels, Tom Eelbode, Maxim Berman, Dirk Vandermeulen, Frederik Maes, Raf Bisschops, and Matthew B Blaschko. Optimizing the dice score and jaccard index for medical image segmentation : Theory and practice. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019 : 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part II 22*, pages 92–100. Springer, 2019.

[32] Javier Oliván Bescós, Marjan J Slob, Cornelis H Slump, Menno Sluzewski, and Willem Jan van Rooij. Volume measurement of intracranial aneurysms from 3D rotational angiography : improvement of accuracy by gradient edge detection. *American journal of neuroradiology*, 26(10) :2569–2572, 2005.

[33] Roongroj Bhidayasiri, Michael FX Waters, and Christopher Giza. *Neurological differential diagnosis : a prioritized approach*. John Wiley & Sons, 2005.

[34] Philippe Bijlenga, Renato Gondar, Sabine Schilling, Sandrine Morel, Sven Hirsch, Johanna Cuony, Marco-Vincenzo Corniola, Fabienne Perren, Daniel Rüfenacht, and Karl Schaller. PHASES score for the management of intracranial aneurysm : a cross-sectional population-based retrospective study. *Stroke*, 48(8) :2105–2112, 2017.

[35] Ziga Bizjak, June Ho Choi, Wonhyoung Park, Franjo Pernuš, and Žiga Špiclin. Deep geometric learning for intracranial aneurysm detection : towards expert rater performance. *Journal of NeuroInterventional Surgery*, 2023.

[36] William C Black and H Gilbert Welch. Advances in diagnostic imaging and overestimations of disease prevalence and the benefits of therapy. *New England Journal of Medicine*, 328(17) :1237–1243, 1993.

[37] Alexey Bochkovskiy, Chien-Yao Wang, and Hong-Yuan Mark Liao. Yolov4 : Optimal speed and accuracy of object detection. *arXiv preprint arXiv :2004.10934*, 2020.

[38] Amelia K Boehme, Charles Esenwa, and Mitchell SV Elkind. Stroke risk factors, genetics, and prevention. *Circulation research*, 120(3) :472–495, 2017.

[39] Alexei Botchkarev. Performance Metrics (Error Measures) in Machine Learning Regression, Forecasting and Prognostics : Properties and Typology. *ArXiv*, abs/1809.03006, 2018.

[40] Andrew P Bradley. The use of the area under the ROC curve in the evaluation of machine learning algorithms. *Pattern recognition*, 30(7) :1145–1159, 1997.

[41] Adrian P Brady. Error and discrepancy in radiology : inevitable or avoidable ? *Insights into imaging*, 8 :171–182, 2017.

[42] M. Brett, A.P. Leff, C. Rorden, and J. Ashburner. Spatial normalization of brain images with focal lesions using cost function masking. *Neuroimage*, 14(2) :486–500, 2001.

[43] Jonathan L Brisman, Joon K Song, and David W Newell. Cerebral aneurysms. *New England journal of medicine*, 355(9) :928–939, 2006.

[44] Robert D Brown, John Huston, Richard Hornung, DR PH, Tatiana Foroud, David F Kallmes, Dawn Kleindorfer, Irene Meissner, Daniel Woo, Laura Sauerbeck, et al. Screening for brain aneurysm in the Familial Intracranial Aneurysm study : frequency and predictors of lesion detection. *Journal of neurosurgery*, 108(6) :1132–1138, 2008.

[45] Mateusz Buda, Atsuto Maki, and Maciej A Mazurowski. A systematic study of the class imbalance problem in convolutional neural networks. *Neural networks*, 106 :249–259, 2018.

[46] Daniel Bystrov, Vladimir Pekar, Stewart Young, Sebastian PM Dries, Harald S Heese, and Arianne M van Muiswinkel. Automated planning of MRI scans of knee joints. In *Medical Imaging 2007 : Visualization and Image-Guided Procedures*, volume 6509, pages 1023–1031. SPIE, 2007.

[47] Hu Cao, Yueyue Wang, Joy Chen, Dongsheng Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-Unet : Unet-like pure transformer for medical image segmentation. In *European conference on computer vision*, pages 205–218. Springer, 2022.

[48] M Jorge Cardoso, Wenqi Li, Richard Brown, Nic Ma, Eric Kerfoot, Yiheng Wang, Benjamin Murrey, Andriy Myronenko, Can Zhao, Dong Yang, et al. Monai : An open-source framework for deep learning in healthcare. *arXiv preprint arXiv :2211.02701*, 2022.

[49] Heang-Ping Chan, Lubomir M Hadjiiski, and Ravi K Samala. Computer-aided diagnosis in the era of deep learning. *Medical physics*, 47(5) :e218–e227, 2020.

[50] Siang-Hua Victor Chan, Kai-Sing Alain Wong, Yat-Ming Peter Woo, Kwong-Yau Chan, and Kar-Ming Leung. Volume measurement of the intracranial aneurysm : a discussion and comparison of the

alternatives to manual segmentation. *Journal of Cerebrovascular and Endovascular Neurosurgery*, 16(4) :358–363, 2014.

[51] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. SMOTE : synthetic minority over-sampling technique. *Journal of artificial intelligence research*, 16 :321–357, 2002.

[52] Geng Chen, Xia Wei, Huang Lei, Yang Liqin, Li Yuxin, Dai Yakang, and Geng Daoying. Automated computer-assisted detection system for cerebral aneurysms in time-of-flight magnetic resonance angiography using fully convolutional network. *BioMedical Engineering OnLine*, 19(1) :1–10, 2020.

[53] Jiancong Chen, Yingying Zhang, Jingyi Wang, Xiaoxue Zhou, Yihua He, and Tong Zhang. Ellipsenet : Anchor-free ellipse detection for automatic cardiac biometrics in fetal echocardiography. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 218–227. Springer, 2021.

[54] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yan Wang, Le Lu, Alan L Yuille, and Yuyin Zhou. Transunet : Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv :2102.04306*, 2021.

[55] Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Kevin Murphy, and Alan L Yuille. Deeplab : Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE transactions on pattern analysis and machine intelligence*, 40(4) :834–848, 2017.

[56] Long Chen, Hanwang Zhang, Jun Xiao, Liqiang Nie, Jian Shao, Wei Liu, and Tat-Seng Chua. SCA-CNN : Spatial and channel-wise attention in convolutional networks for image captioning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5659–5667, 2017.

[57] Ting Chen, Yiqiang Zhan, Shaoting Zhang, and Maneesh Dewan. Automatic alignment of brain MR scout scans using data-adaptive multi-structural model. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2011 : 14th International Conference, Toronto, Canada, September 18-22, 2011, Proceedings, Part II 14*, pages 574–581. Springer, 2011.

[58] Xiaodan Chen, Yun Liu, Huazhang Tong, Yonghai Dong, Dongyang Ma, Lei Xu, and Cheng Yang. Meta-analysis of computed tomography angiography versus magnetic resonance angiography for intracranial aneurysm. *Medicine*, 97(20), 2018.

[59] Yicheng Chen, Shengxiang Ji, Yuelin Xin, Kun Han, and Xiaohui Xie. On-the-Fly guidance Training for Medical Image Registration. 2023.

[60] Phillip M Cheng, Emmanuel Montagnon, Rikiya Yamashita, Ian Pan, Alexandre Cadrin-Chênevert, Francisco Perdigón Romero, Gabriel Chartrand, Samuel Kadoury, and An Tang. Deep learning : an update for radiologists. *Radiographics*, 41(5) :1427–1445, 2021.

[61] Veronika Cheplygina, Marleen de Bruijne, and Josien PW Pluim. Not-so-supervised : a survey of semi-supervised, multi-instance, and transfer learning in medical image analysis. *Medical image analysis*, 54 :280–296, 2019.

[62] Phillip Chlap, Hang Min, Nym Vandenberg, Jason Dowling, Lois Holloway, and Annette Haworth. A review of medical image data augmentation techniques for deep learning applications. *Journal of Medical Imaging and Radiation Oncology*, 65(5) :545–563, 2021.

[63] Dami Choi, Christopher J Shallue, Zachary Nado, Jaehoon Lee, Chris J Maddison, and George E Dahl. On empirical comparisons of optimizers for deep learning. *arXiv preprint arXiv :1910.05446*, 2019.

[64] François Chollet et al. Keras. `https://keras.io`, 2015.

[65] Ö. Çiçek, A. Abdulkadir, S. Lienkamp, et al. 3D U-Net : learning dense volumetric segmentation from sparse annotation. In *International conference on medical image computing and computer-assisted intervention (MICCAI)*, pages 424–432, 2016.

[66] Dan Ciregan, Ueli Meier, and Jürgen Schmidhuber. Multi-column deep neural networks for image classification. In *2012 IEEE conference on computer vision and pattern recognition*, pages 3642–3649. IEEE, 2012.

[67] J. Cohen. A coefficient of agreement for nominal scales. *Educational and psychological measurement*, 20(1) :37–46, 1960.

[68] Imperial College London. Information extraction from images (IXI). `https://brain-development.org/ixi-dataset/`, 2015. Accessed : November 27, 2023.

[69] Alvaro Collet, Manuel Martinez, and Siddhartha S Srinivasa. The moped framework : Object recognition and pose estimation for manipulation. *The international journal of robotics research*, 30(10) :1284–1306, 2011.

[70] Danny Coomans and Désiré Luc Massart. Alternative k-nearest neighbour rules in supervised pattern recognition : Part 1. k-nearest neighbour classification by using alternative voting rules. *Analytica Chimica Acta*, 136 :15–27, 1982.

[71] Amanda J Craig, Johann Von Felden, Teresa Garcia-Lezana, Samantha Sarcognato, and Augusto Villanueva. Tumour evolution in hepatocellular carcinoma. *Nature reviews Gastroenterology & hepatology*, 17(3) :139–152, 2020.

[72] Jifeng Dai, Haozhi Qi, Yuwen Xiong, Yi Li, Guodong Zhang, Han Hu, and Yichen Wei. Deformable convolutional networks. In *Proceedings of the IEEE international conference on computer vision*, pages 764–773, 2017.

[73] Yuchao Dai, Hongdong Li, and Laurent Kneip. Rolling shutter camera relative pose : Generalized epipolar geometry. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4132–4140, 2016.

[74] Onat Dalmaz, Mahmut Yurt, and Tolga Çukur. Resvit : Residual vision transformers for multimodal medical image synthesis. *IEEE Transactions on Medical Imaging*, 41(10) :2598–2614, 2022.

[75] Erik B Dam, Martin Koch, and Martin Lillholm. *Quaternions, interpolation and animation*, volume 2. Datalogisk Institut, Københavns Universitet Copenhagen, Denmark, 1998.

[76] R. De Feo, J. Kaiponen, and J. Tohka. Multi-loss CNN ensembles for aneurysm segmentation. In *Automatic Detection And SegMentation Challenge (ADAM)*, 2020.

[77] S Deepak and PM Ameer. Brain tumor classification using deep CNN features via transfer learning. *Computers in biology and medicine*, 111 :103345, 2019.

[78] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet : A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[79] Li Deng. The mnist database of handwritten digit images for machine learning research [best of the web]. *IEEE signal processing magazine*, 29(6) :141–142, 2012.

[80] Tommaso Di Noto, Guillaume Marie, Sebastien Tourbier, Yasser Alemán-Gómez, Oscar Esteban, Guillaume Saliou, Meritxell Bach Cuadra, Patric Hagmann, and Jonas Richiardi. Towards automated brain aneurysm detection in TOF-MRA : open data, weak labels, and anatomical knowledge. *Neuroinformatics*, pages 1–14, 2022.

[81] Munaib Din, Siddharth Agarwal, Mariusz Grzeda, David A Wood, Marc Modat, and Thomas C Booth. Detection of cerebral aneurysms using artificial intelligence : a systematic review and meta-analysis. *Journal of NeuroInterventional Surgery*, 15(3) :262–271, 2023.

[82] Jia Ding, Aoxue Li, Zhiqiang Hu, and Liwei Wang. Accurate pulmonary nodule detection in computed tomography images using deep convolutional neural networks. In *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017 : 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*, pages 559–567. Springer, 2017.

[83] Kai Dinghofer and Frank Hartung. Analysis of criteria for the selection of machine learning frameworks. In *2020 International Conference on Computing, Networking and Communications (ICNC)*, pages 373–377. IEEE, 2020.

[84] Alexey Dosovitskiy, Lucas Beyer, Alexander Kolesnikov, Dirk Weissenborn, Xiaohua Zhai, Thomas Unterthiner, Mostafa Dehghani, Matthias Minderer, Georg Heigold, Sylvain Gelly, et al. An image is worth 16x16 words : Transformers for image recognition at scale. *arXiv preprint arXiv :2010.11929*, 2020.

[85] Kaiwen Duan, Song Bai, Lingxi Xie, Honggang Qi, Qingming Huang, and Qi Tian. Centernet : Keypoint triplets for object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 6569–6578, 2019.

[86] Audrey Duran, Gaspard Dussert, Olivier Rouvière, Tristan Jaouen, Pierre-Marc Jodoin, and Carole Lartizien. ProstAttention-Net : A deep attention model for prostate cancer segmentation by aggressiveness in MRI scans. *Medical Image Analysis*, 77 :102347, 2022.

[87] Andre Esteva, Katherine Chou, Serena Yeung, Nikhil Naik, Ali Madani, Ali Mottaghi, Yun Liu, Eric Topol, Jeff Dean, and Richard Socher. Deep learning-enabled medical computer vision. *NPJ digital medicine*, 4(1) :5, 2021.

[88] Nima Etminan, Robert D Brown, Kerim Beseoglu, Seppo Juvela, Jean Raymond, Akio Morita, James C Torner, Colin P Derdeyn, Andreas Raabe, J Mocco, et al. The unruptured intracranial aneurysm treatment score : a multidisciplinary consensus. *Neurology*, 85(10) :881–889, 2015.

[89] Nima Etminan, Han-Sol Chang, Katharina Hackenberg, Nicolien K De Rooij, Mervyn DI Vergouwen, Gabriel JE Rinkel, and Ale Algra. Worldwide incidence of aneurysmal subarachnoid hemorrhage according to region, time period, blood pressure, and smoking prevalence in the population : a systematic review and meta-analysis. *JAMA neurology*, 76(5) :588–597, 2019.

[90] Mohammed Faluk and Orlando De Jesus. Saccular aneurysm. 2020.

[91] A. Faron, T. Sichtermann, N. Teichert, et al. Performance of a deep-learning neural network to detect intracranial aneurysms from 3D TOF-MRA compared to human readers. *Clinical neuroradiology*, 30(3) :591–598, 2020.

[92] A. Fedorov, R. Beichel, J. Kalpathy-Cramer, et al. 3D Slicer as an image computing platform for the Quantitative Imaging Network. *Magnetic resonance imaging*, 30(9) :1323–1341, 2012. PMID : 22770690.

[93] Enzo Ferrante and Nikos Paragios. Slice-to-volume medical image registration : A survey. *Medical image analysis*, 39 :101–123, 2017.

[94] Vittorio Ferrari, Tinne Tuytelaars, and Luc Van Gool. Simultaneous object recognition and segmentation from single or multiple model views. *International journal of computer vision*, 67(2) :159–188, 2006.

[95] Juhana Frösen, Riikka Tulamo, Anders Paetau, Elisa Laaksamo, Miikka Korja, Aki Laakso, Mika Niemelä, and Juha Hernesniemi. Saccular intracranial aneurysm : pathology and mechanisms. *Acta neuropathologica*, 123 :773–786, 2012.

[96] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, and Hanqing Lu. Dual attention network for scene segmentation. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 3146–3154, 2019.

[97] Kunihiko Fukushima. Neocognitron : A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position. *Biological cybernetics*, 36(4) :193–202, 1980.

[98] F Brock Fuller. The writhing number of a space curve. *Proceedings of the National Academy of Sciences*, 68(4) :815–819, 1971.

[99] David M Gauntt and Gary T Barnes. An automatic and accurate x-ray tube focal spot/grid alignment system for mobile radiography : System description and alignment accuracy a. *Medical Physics*, 37(12) :6402–6410, 2010.

[100] Mariana-Iuliana Georgescu, Radu Tudor Ionescu, Andreea-Iuliana Miron, Olivian Savencu, Nicolae-Cătălin Ristea, Nicolae Verga, and Fahad Shahbaz Khan. Multimodal multi-head convolutional attention with various kernel sizes for medical image super-resolution. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pages 2195–2205, 2023.

[101] Ross Girshick. Fast r-cnn. In *Proceedings of the IEEE international conference on computer vision*, pages 1440–1448, 2015.

[102] Ross Girshick, Jeff Donahue, Trevor Darrell, and Jitendra Malik. Rich feature hierarchies for accurate object detection and semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 580–587, 2014.

[103] Xavier Glorot and Yoshua Bengio. Understanding the difficulty of training deep feedforward neural networks. In *Proceedings of the thirteenth international conference on artificial intelligence and statistics*, pages 249–256. JMLR Workshop and Conference Proceedings, 2010.

[104] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings, 2011.

[105] Ron Goldman. Understanding quaternions. *Graphical models*, 73(2) :21–49, 2011.

[106] Ian Goodfellow, Yoshua Bengio, and Aaron Courville. *Deep learning*. MIT press, 2016.

[107] Philip B Gorelick, Ka Sing Wong, Hee-Joon Bae, and Dilip K Pandey. Large artery intracranial occlusive disease : a large worldwide burden but a relatively neglected frontier. *Stroke*, 39(8) :2396–2399, 2008.

[108] Akhilesh Gotmare, Nitish Shirish Keskar, Caiming Xiong, and Richard Socher. A closer look at deep learning heuristics : Learning rate restarts, warmup and distillation. *arXiv preprint arXiv :1810.13243*, 2018.

[109] Priya Goyal, Piotr Dollár, Ross Girshick, Pieter Noordhuis, Lukasz Wesolowski, Aapo Kyrola, Andrew Tulloch, Yangqing Jia, and Kaiming He. Accurate, large minibatch sgd : Training imagenet in 1 hour. *arXiv preprint arXiv :1706.02677*, 2017.

[110] Jiuxiang Gu, Zhenhua Wang, Jason Kuen, Lianyang Ma, Amir Shahroudy, Bing Shuai, Ting Liu, Xingxing Wang, Gang Wang, Jianfei Cai, et al. Recent advances in convolutional neural networks. *Pattern recognition*, 77 :354–377, 2018.

[111] Hao Guan and Mingxia Liu. Domain adaptation for medical image analysis : a survey. *IEEE Transactions on Biomedical Engineering*, 69(3) :1173–1185, 2021.

[112] Odd Erik Gundersen and Sigbjørn Kjensmo. State of the art : Reproducibility in artificial intelligence. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.

[113] Meng-Hao Guo, Tian-Xing Xu, Jiang-Jiang Liu, Zheng-Ning Liu, Peng-Tao Jiang, Tai-Jiang Mu, Song-Hai Zhang, Ralph R Martin, Ming-Ming Cheng, and Shi-Min Hu. Attention mechanisms in computer vision : A survey. *Computational visual media*, 8(3) :331–368, 2022.

[114] Michael P Hartung, Thomas M Grist, and Christopher J François. Magnetic resonance angiography : current status and future directions. *Journal of Cardiovascular Magnetic Resonance*, 13(1) :1–11, 2011.

[115] Md Kamrul Hasan, Lilian Calvet, Navid Rabbani, and Adrien Bartoli. Detection, segmentation, and 3D pose estimation of surgical tools using convolutional neural networks and algebraic geometry. *Medical Image Analysis*, 70 :101994, 2021.

[116] Haibo He, Yang Bai, Edwardo A Garcia, and Shutao Li. ADASYN : Adaptive synthetic sampling approach for imbalanced learning. In *2008 IEEE international joint conference on neural networks (IEEE world congress on computational intelligence)*, pages 1322–1328. Ieee, 2008.

[117] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Delving deep into rectifiers : Surpassing human-level performance on ImageNet classification. In *Proceedings of the IEEE international conference on computer vision*, pages 1026–1034, 2015.

[118] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[119] Clemens M Hentschke, Oliver Beuing, Rosa Nickl, and Klaus D Tönnies. Detection of cerebral aneurysms in MRA, CTA and 3D-RA data sets. In *Medical Imaging 2012 : Computer-Aided Diagnosis*, volume 8315, pages 447–453. SPIE, 2012.

[120] Clemens M Hentschke, Oliver Beuing, Harald Paukisch, Cordula Scherlach, Martin Skalej, and Klaus D Tönnies. A system to detect cerebral aneurysms in multimodality angiographic data sets. *Medical physics*, 41(9) :091904, 2014.

[121] Suzana Herculano-Houzel. The human brain in numbers : a linearly scaled-up primate brain. *Frontiers in human neuroscience*, page 31, 2009.

[122] Steven A Hicks, Inga Strümke, Vajira Thambawita, Malek Hammou, Michael A Riegler, Pål Halvorsen, and Sravanthi Parasa. On evaluation metrics for medical applications of artificial intelligence. *Scientific reports*, 12(1) :5979, 2022.

[123] Lars B Hindenes, Asta K Håberg, Liv Hege Johnsen, Ellisiv B Mathiesen, David Robben, and Torgil R Vangberg. Variations in the Circle of Willis in a large population sample using 3D TOF angiography : The Tromsø Study. *PLoS One*, 15(11) :e0241373, 2020.

[124] Sven Hirsch, Johannes Reichold, Matthias Schneider, Gábor Székely, and Bruno Weber. Topology and hemodynamics of the cortical cerebrovascular system. *Journal of Cerebral Blood Flow & Metabolism*, 32(6) :952–967, 2012.

[125] Wenguang Hou, Shaojie Mei, Qiuling Gui, Yingcheng Zou, Yifan Wang, Xianbo Deng, and Qimin Cheng. 1d CNN-based intracranial aneurysms detection in 3D TOF-MRA. *Complexity*, 2020 :1–13, 2020.

[126] Jie Hu, Li Shen, Samuel Albanie, Gang Sun, and Andrea Vedaldi. Gather-excite : Exploiting feature context in convolutional neural networks. *Advances in neural information processing systems*, 31, 2018.

[127] Jie Hu, Li Shen, and Gang Sun. Squeeze-and-excitation networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7132–7141, 2018.

[128] Xiaoming Hu, T Michael De Silva, Jun Chen, and Frank M Faraci. Cerebral vascular disease and neurovascular injury in ischemic stroke. *Circulation research*, 120(3) :449–471, 2017.

[129] Yuhao Huang, Xin Yang, Lian Liu, Han Zhou, Ao Chang, Xinrui Zhou, Rusi Chen, Junxuan Yu, Jiongquan Chen, Chaoyu Chen, et al. Segment anything model for medical images ? *Medical Image Analysis*, 92 :103061, 2024.

[130] Ziyan Huang, Zehua Wang, Zhikai Yang, and Lixu Gu. AdwU-Net : adaptive depth and width U-Net for medical image segmentation by differentiable neural architecture search. In *International Conference on Medical Imaging with Deep Learning*, pages 576–589. PMLR, 2022.

[131] Haruki Imai, Samuel Matzek, Tung D Le, Yasushi Negishi, and Kiyokuni Kawachiya. Fast and accurate 3D medical image segmentation with data-swapping method. *arXiv preprint arXiv :1812.07816*, 2018.

[132] Takashi Inoue, Hiroaki Shimizu, Miki Fujimura, Atsushi Saito, and Teiji Tominaga. Annual rupture risk of growing unruptured cerebral aneurysms detected by magnetic resonance angiography. *Journal of neurosurgery*, 117(1) :20–25, 2012.

[133] Sergey Ioffe and Christian Szegedy. Batch normalization : Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. pmlr, 2015.

[134] Jeremy Irvin, Pranav Rajpurkar, Michael Ko, Yifan Yu, Silviana Ciurea-Ilcus, Chris Chute, Henrik Marklund, Behzad Haghgoo, Robyn Ball, Katie Shpanskaya, et al. Chexpert : A large chest radiograph dataset with uncertainty labels and expert comparison. In *Proceedings of the AAAI conference on artificial intelligence*, volume 33, pages 590–597, 2019.

[135] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnU-Net : a self-configuring method for deep learning-based biomedical image segmentation. *Nature methods*, 18(2) :203–211, 2021.

[136] Ali Iskurt, Yasar Becerikli, and Kamran Mahmutyazicioglu. Automatic identification of landmarks for standard slice positioning in brain MRI. *Journal of Magnetic Resonance Imaging*, 34(3) :499–510, 2011.

[137] Matthias Ivantsits, Leonid Goubergrits, Jan-Martin Kuhnigk, Markus Huellebrand, Jan Bruening, Tabea Kossen, Boris Pfahringer, Jens Schaller, Andreas Spuler, Titus Kuehne, et al. Detection and analysis of cerebral aneurysms based on X-ray rotational angiography-the CADA 2020 challenge. *Medical image analysis*, 77 :102333, 2022.

[138] Matthias Ivantsits, Jan-Martin Kuhnigk, Markus Huellebrand, Titus Kuehne, and Anja Hennemuth. Deep learning-based 3D U-Net Cerebral Aneurysm Detection. In *Cerebral Aneurysm Detection and Analysis : First Challenge, CADA 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 1*, pages 31–38. Springer, 2021.

[139] Paul F Jaeger, Simon AA Kohl, Sebastian Bickelhaupt, Fabian Isensee, Tristan Anselm Kuder, Heinz-Peter Schlemmer, and Klaus H Maier-Hein. Retina U-Net : Embarrassingly simple exploitation of segmentation supervision for medical object detection. In *Machine Learning for Health Workshop*, pages 171–183. PMLR, 2020.

[140] M. Jang, J.H. Kim, J.W. Park, et al. Features of "false positive" unruptured intracranial aneurysms on screening magnetic resonance angiography. *PloS one*, 15(9) :e0238597, 2020.

[141] Tim Jerman, Aichi Chien, Franjo Pernuš, Boštjan Likar, and Žiga Špiclin. Automated cutting plane positioning for intracranial aneurysm quantification. *IEEE Transactions on Biomedical Engineering*, 67(2) :577–587, 2019.

[142] Tim Jerman, Franjo Pernuš, Boštjan Likar, and Žiga Špiclin. Computer-aided detection and quantification of intracranial aneurysms. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015 : 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part II 18*, pages 3–10. Springer, 2015.

[143] Andrew M Jersey and David M Foster. Cerebral aneurysm. 2018.

[144] Ge-Peng Ji, Guobao Xiao, Yu-Cheng Chou, Deng-Ping Fan, Kai Zhao, Geng Chen, and Luc Van Gool. Video polyp segmentation : A deep learning perspective. *Machine Intelligence Research*, 19(6) :531–549, 2022.

[145] Yizhuan Jia, Weibin Liao, Yi Lv, Ziyu Su, Jiaqi Dou, Zhongwei Sun, and Xuesong Li. Detect and Identify Aneurysms Based on Adjusted 3D Attention UNet. In *Cerebral Aneurysm Detection and Analysis : First Challenge, CADA 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 1*, pages 39–48. Springer, 2021.

[146] Cheng Jin, Zhengrui Guo, Yi Lin, Luyang Luo, and Hao Chen. Label-efficient deep learning in medical image analysis : Challenges and future directions. *arXiv preprint arXiv :2303.12484*, 2023.

[147] Ze Jin, Hidetaka Arimura, Shingo Kakeda, Fumio Yamashita, Makoto Sasaki, and Yukunori Korogi. An ellipsoid convex enhancement filter for detection of asymptomatic intracranial aneurysm candidates in CAD frameworks. *Medical physics*, 43(2) :951–960, 2016.

[148] J.M. Johnson and T.M. Khoshgoftaar. Survey on deep learning with class imbalance. *Journal of Big Data*, 6(1) :1–54, 2019.

[149] Joshua D Jones, Pedro Castanho, Peter Bazira, and Katherine Sanders. Anatomical variations of the circle of Willis and their prevalence, with a focus on the posterior communicating artery : A literature review and meta-analysis. *Clinical Anatomy*, 34(7) :978–990, 2021.

[150] Bio Joo, Sung Soo Ahn, Pyeong Ho Yoon, Sohi Bae, Beomseok Sohn, Yong Eun Lee, Jun Ho Bae, Moo Sung Park, Hyun Seok Choi, and Seung-Koo Lee. A deep learning algorithm may automate intracranial aneurysm detection on MR angiography with high diagnostic performance. *European Radiology*, 30(11) :5785–5793, 2020.

[151] Wiard Jorritsma, Fokie Cnossen, and Peter MA van Ooijen. Improving the radiologist–CAD interaction : designing for appropriate trust. *Clinical radiology*, 70(2) :115–122, 2015.

[152] Willi A Kalender. X-ray computed tomography. *Physics in Medicine & Biology*, 51(13) :R29, 2006.

[153] Konstantinos Kamnitsas, Christian Ledig, Virginia FJ Newcombe, Joanna P Simpson, Andrew D Kane, David K Menon, Daniel Rueckert, and Ben Glocker. Efficient multi-scale 3D CNN with fully connected CRF for accurate brain lesion segmentation. *Medical image analysis*, 36 :61–78, 2017.

[154] Michael Kampffmeyer, Arnt-Borre Salberg, and Robert Jenssen. Semantic segmentation of small objects and modeling of uncertainty in urban remote sensing images using deep convolutional neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 1–9, 2016.

[155] Davood Karimi, Haoran Dou, Simon K Warfield, and Ali Gholipour. Deep learning with noisy labels : Exploring techniques and remedies in medical image analysis. *Medical image analysis*, 65 :101759, 2020.

[156] Lisa Kausch, Sarina Thomas, Holger Kunze, Jan Siad El Barbari, and Klaus H Maier-Hein. Shape-based pose estimation for automatic standard views of the knee. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 476–486. Springer, 2023.

[157] Keisuke Kawauchi, Sho Furuya, Kenji Hirata, Chietsugu Katoh, Osamu Manabe, Kentaro Kobayashi, Shiro Watanabe, and Tohru Shiga. A convolutional neural network-based system to classify patients using FDG PET/CT examinations. *BMC cancer*, 20 :1–10, 2020.

[158] Alexander Keedy. An overview of intracranial aneurysms. *McGill Journal of Medicine : MJM*, 9(2) :141, 2006.

[159] Hoel Kervadec, Jihene Bouchtiba, Christian Desrosiers, Eric Granger, Jose Dolz, and Ismail Ben Ayed. Boundary loss for highly unbalanced segmentation. In *International conference on medical imaging with deep learning*, pages 285–296. PMLR, 2019.

[160] Diederik P Kingma and Jimmy Ba. Adam : A method for stochastic optimization. *arXiv preprint arXiv :1412.6980*, 2014.

[161] Syoji Kobashi, Katsuya Kondo, and Yutaka Hata. Computer-aided diagnosis of intracranial aneurysms in MRA images with case-based reasoning. *IEICE TRANSACTIONS on Information and Systems*, 89(1) :340–350, 2006.

[162] Mona Köhler, Markus Eisenbach, and Horst-Michael Gross. Few-shot object detection : a comprehensive survey. *IEEE Transactions on Neural Networks and Learning Systems*, 2023.

[163] Alex Krizhevsky, Geoffrey Hinton, et al. Learning multiple layers of features from tiny images. 2009.

[164] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. ImageNet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[165] Anderson H Kuo, Prashant Nagpal, Brian B Ghoshhajra, and Sandeep S Hedgire. Vascular magnetic resonance angiography techniques. *Cardiovascular Diagnosis and Therapy*, 9(Suppl 1) :S28, 2019.

[166] Rohan R Lall, Christopher S Eddleman, Bernard R Bendok, and H Hunt Batjer. Unruptured intracranial aneurysms and the assessment of rupture risk based on anatomical and morphological factors : sifting through the sands of data. *Neurosurgical focus*, 26(5) :E2, 2009.

[167] Alexandra Lauric, Eric Miller, Sarah Frisken, and Adel M Malek. Automated detection of intracranial aneurysms based on parent vessel 3D analysis. *Medical image analysis*, 14(2) :149–159, 2010.

[168] Hei Law and Jia Deng. Cornernet : Detecting objects as paired keypoints. In *Proceedings of the European conference on computer vision (ECCV)*, pages 734–750, 2018.

[169] Frederic E Lecouvet, Jérôme Claus, Patrick Schmitz, Vincent Denolin, Clemens Bos, and Bruno C Vande Berg. Clinical evaluation of automated scan prescription of knee MR images. *Journal of Magnetic Resonance Imaging : An Official Journal of the International Society for Magnetic Resonance in Medicine*, 29(1) :141–145, 2009.

[170] Yann LeCun, Léon Bottou, Yoshua Bengio, and Patrick Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11) :2278–2324, 1998.

[171] Dawa Chyophel Lepcha, Bhawna Goyal, Ayush Dogra, Kanta Prasad Sharma, and Deena Nath Gupta. A deep journey into image enhancement : A survey of current and emerging trends. *Information Fusion*, 93 :36–76, 2023.

[172] Vincent Lepetit, Pascal Fua, et al. Monocular model-based 3d tracking of rigid objects : A survey. *Foundations and Trends® in Computer Graphics and Vision*, 1(1) :1–89, 2005.

[173] Xuecan Liang, Pengyu Hu, Liguo Zhang, Jianguo Sun, and Guisheng Yin. MCFNet : Multi-layer concatenation fusion network for medical images fusion. *IEEE Sensors Journal*, 19(16) :7107–7119, 2019.

[174] Fangzhou Liao, Ming Liang, Zhe Li, Xiaolin Hu, and Sen Song. Evaluate the malignancy of pulmonary nodules using the 3-d deep leaky noisy-or network. *IEEE transactions on neural networks and learning systems*, 30(11) :3484–3495, 2019.

[175] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, and Serge Belongie. Feature pyramid networks for object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2117–2125, 2017.

[176] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, and Piotr Dollár. Focal loss for dense object detection. In *Proceedings of the IEEE international conference on computer vision*, pages 2980–2988, 2017.

[177] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. Microsoft coco : Common objects in context. In *Computer Vision–ECCV 2014 : 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part V 13*, pages 740–755. Springer, 2014.

[178] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen Awm Van Der Laak, Bram Van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42 :60–88, 2017.

[179] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. Ssd : Single shot multibox detector. In *Computer Vision–ECCV 2016 : 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part I 14*, pages 21–37. Springer, 2016.

[180] X Liu, H Haraldsson, Y Wang, E Kao, M Ballweber, AJ Martin, CE McCulloch, F Faraji, D Saloner, UCSF Intracranial Aneurysm Monitoring Group, et al. A volumetric metric for monitoring intracranial aneurysms : repeatability and growth criteria in a longitudinal MR imaging study. *American Journal of Neuroradiology*, 42(9) :1591–1597, 2021.

[181] Yang Liu, Peng Sun, Nickolas Wergeles, and Yi Shang. A survey and performance evaluation of deep learning methods for small object detection. *Expert Systems with Applications*, 172 :114602, 2021.

[182] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer : Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022, 2021.

[183] Xavier Lladó, Onur Ganiler, Arnau Oliver, Robert Martí, Jordi Freixenet, Laia Valls, Joan C Vilanova, Lluís Ramió-Torrentà, and Àlex Rovira. Automated detection of multiple sclerosis lesions in serial brain MRI. *Neuroradiology*, 54 :787–807, 2012.

[184] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015.

[185] Xiangde Luo, Tao Song, Guotai Wang, Jieneng Chen, Yinan Chen, Kang Li, Dimitris N Metaxas, and Shaoting Zhang. SCPM-net : An anchor-free 3d lung nodule detection network using sphere representation and center points matching. *Medical Image Analysis*, 75 :102287, 2022.

[186] J. Ma and X. An. Loss Ensembles for Intracranial Aneurysm Segmentation : An Embarrassingly Simple Method. In *Automatic Detection And SegMentation Challenge (ADAM)*, 2020.

[187] Jun Ma. Cutting-edge 3D medical image segmentation methods in 2020 : Are happy families all alike ? *arXiv preprint arXiv :2101.00232*, 2021.

[188] Jun Ma, Jianan Chen, Matthew Ng, Rui Huang, Yu Li, Chen Li, Xiaoping Yang, and Anne L. Martel. Loss Odyssey in Medical Image Segmentation. *Medical Image Analysis*, 71 :102035, 2021.

[189] Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1) :654, 2024.

[190] James MacQueen et al. Some methods for classification and analysis of multivariate observations. In *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, volume 1, pages 281–297. Oakland, CA, USA, 1967.

[191] Batta Mahesh. Machine learning algorithms-a review. *International Journal of Science and Research (IJSR).[Internet]*, 9(1) :381–386, 2020.

[192] Lena Maier-Hein, Bjoern Menze, et al. Metrics reloaded : Pitfalls and recommendations for image analysis validation. *arXiv. org*, (2206.01653), 2022.

[193] Inderjeet Mani and I Zhang. kNN approach to unbalanced data distributions : a case study involving information extraction. In *Proceedings of workshop on learning from imbalanced datasets*, volume 126, pages 1–7. ICML, 2003.

[194] Eric Marchand, Hideaki Uchiyama, and Fabien Spindler. Pose estimation for augmented reality : a hands-on survey. *IEEE transactions on visualization and computer graphics*, 22(12) :2633–2651, 2015.

[195] Primoz Markelj, Dejan Tomaževič, Bostjan Likar, and Franjo Pernuš. A review of 3D/2D registration methods for image-guided interventions. *Medical image analysis*, 16(3) :642–661, 2012.

[196] Clémence Maupu, Héloïse Lebas, and Yacine Boulaftali. Imaging modalities for intracranial aneurysm : more than meets the eye. *Frontiers in Cardiovascular Medicine*, 9, 2022.

[197] O. Maysam and Z. Chengcheng. nnU-Net with Multiple Loss Ensembles for Aneurysm Segmentation. In *Automatic Detection And SegMentation Challenge (ADAM)*, 2020.

[198] Robert J McDonald, Kara M Schwartz, Laurence J Eckel, Felix E Diehn, Christopher H Hunt, Brian J Bartholmai, Bradley J Erickson, and David F Kallmes. The effects of changes in utilization and technological advancements of cross-sectional imaging on radiologist workload. *Academic radiology*, 22(9) :1191–1198, 2015.

[199] Mary L McHugh. Interrater reliability : the kappa statistic. *Biochemia medica*, 22(3) :276–282, 2012.

[200] Shun Miao, Joseph Lucas, and Rui Liao. Automatic pose initialization for accurate 2D/3D registration applied to abdominal aortic aneurysm endovascular repair. In *Medical Imaging 2012 : Image-Guided Procedures, Robotic Interventions, and Modeling*, volume 8316, pages 243–250. SPIE, 2012.

[201] Fausto Milletari, Nassir Navab, and Seyed-Ahmad Ahmadi. V-net : Fully convolutional neural networks for volumetric medical image segmentation. In *2016 fourth international conference on 3D vision (3DV)*, pages 565–571. Ieee, 2016.

[202] Shervin Minaee, Yuri Boykov, Fatih Porikli, Antonio Plaza, Nasser Kehtarnavaz, and Demetri Terzopoulos. Image segmentation using deep learning : A survey. *IEEE transactions on pattern analysis and machine intelligence*, 44(7) :3523–3542, 2021.

[203] Mitsue Miyazaki and Vivian S Lee. Nonenhanced MR angiography. *Radiology*, 248(1) :20–43, 2008.

[204] Andrew J Molyneux, Richard SC Kerr, Ly-Mee Yu, Mike Clarke, Mary Sneade, Julia A Yarnold, and Peter Sandercock. International subarachnoid aneurysm trial (ISAT) of neurosurgical clipping versus endovascular coiling in 2143 patients with ruptured intracranial aneurysms : a randomised comparison of effects on survival, dependency, seizures, rebleeding, subgroups, and aneurysm occlusion. *The Lancet*, 366(9488) :809–817, 2005.

[205] Patrice Monkam, Shouliang Qi, He Ma, Weiming Gao, Yudong Yao, and Wei Qian. Detection and classification of pulmonary nodules using convolutional neural networks : a survey. *Ieee Access*, 7 :78075–78091, 2019.

[206] Reza Moradi, Reza Berangi, and Behrouz Minaei. A survey of regularization strategies for deep models. *Artificial Intelligence Review*, 53 :3947–3986, 2020.

[207] Pauline Mouches and Nils D Forkert. A statistical atlas of cerebral arteries generated using multicenter MRA datasets from healthy subjects. *Scientific data*, 6(1) :29, 2019.

[208] Rafic Nader, Romain Bourcier, and Florent Autrusseau. Using deep learning for an automatic detection and classification of the vascular bifurcations along the Circle of Willis. *Medical Image Analysis*, page 102919, 2023.

[209] Mahesh B Nagarajan, Markus B Huber, Thomas Schlossbauer, Gerda Leinsinger, Andrzej Krol, and Axel Wismüller. Classification of small lesions in dynamic breast MRI : eliminating the need for precise lesion segmentation through spatio-temporal analysis of contrast enhancement. *Machine vision and applications*, 24(7) :1371–1381, 2013.

[210] Toshio Nakagawa and Kazuo Hashi. The incidence and treatment of asymptomatic, unruptured cerebral aneurysms. *Journal of neurosurgery*, 80(2) :217–223, 1994.

[211] T. Nakao, S. Hanaoka, Y. Nomura, et al. Deep neural network-based computer-assisted detection of cerebral aneurysms in MR angiography. *Journal of Magnetic Resonance Imaging*, 47(4) :948–953, 2018.

[212] Mathias Neugebauer, Volker Diehl, Martin Skalej, and Bernhard Preim. Geometric Reconstruction of the Ostium of Cerebral Aneurysms. In *VMV*, pages 307–314, 2010.

[213] Ethan H Nguyen, Haichun Yang, Ruining Deng, Yuzhe Lu, Zheyu Zhu, Joseph T Roland, Le Lu, Bennett A Landman, Agnes B Fogo, and Yuankai Huo. Circle Representation for Medical Object Detection. *IEEE transactions on medical imaging*, 41(3) :746–754, 2021.

[214] Ha Q Nguyen, Khanh Lam, Linh T Le, Hieu H Pham, Dat Q Tran, Dung B Nguyen, Dung D Le, Chi M Pham, Hang TT Tong, Diep H Dinh, et al. Vindr-CXR : An open dataset of chest x-rays with radiologist's annotations. *Scientific Data*, 9(1) :429, 2022.

[215] Uli Niemann, Philipp Berg, Annika Niemann, Oliver Beuing, Bernhard Preim, Myra Spiliopoulou, and Sylvia Saalfeld. Rupture status classification of intracranial aneurysms using morphological parameters. In *2018 IEEE 31st International Symposium on Computer-Based Medical Systems (CBMS)*, pages 48–53. IEEE, 2018.

[216] Hidehisa Nishi, Nicole M Cancelliere, Ariana Rustici, Guillaume Charbonnier, Vanessa Chan, Julian Spears, Thomas R Marotta, and Vitor Mendes Pereira. Deep learning-based cerebral aneurysm segmentation and morphological analysis with three-dimensional rotational angiography. *Journal of NeuroInterventional Surgery*, 2023.

[217] Mika Okahara, Hiro Kiyosue, Masanori Yamashita, Hirohumi Nagatomi, Hiroyuki Hata, Toshiyuki Saginoya, Yoshiko Sagara, and Hiromu Mori. Diagnostic accuracy of magnetic resonance angiography for cerebral aneurysms in correlation with 3D–digital subtraction angiographic images : a study of 133 aneurysms. *Stroke*, 33(7) :1803–1808, 2002.

[218] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, Matthew Lee, Mattias Heinrich, Kazunari Misawa, Kensaku Mori, Steven McDonagh, Nils Y Hammerla, Bernhard Kainz, et al. Attention u-net : Learning where to look for the pancreas. *arXiv preprint arXiv :1804.03999*, 1804.

[219] Nobuyuki Otsu. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics*, 9(1) :62–66, 1979.

[220] Niall O'Mahony, Sean Campbell, Anderson Carvalho, and others Harapanahalli. Deep learning vs. traditional computer vision. In *Advances in Computer Vision : Proceedings of the 2019 Computer Vision Conference (CVC), Volume 1 1*, pages 128–144. Springer, 2020.

[221] Rafael Padilla, Sergio L Netto, and Eduardo AB Da Silva. A survey on performance metrics for object-detection algorithms. In *2020 international conference on systems, signals and image processing (IWSSIP)*, pages 237–242. IEEE, 2020.

[222] A. Park, C. Chute, P. Rajpurkar, et al. Deep Learning–Assisted Diagnosis of Cerebral Aneurysms Using the HeadXNet Model. *JAMA network open*, 2(6) :e195600–e195600, 2019.

[223] Kiru Park, Timothy Patten, and Markus Vincze. Pix2pose : Pixel-wise coordinate regression of objects for 6D pose estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 7668–7677, 2019.

[224] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch : An imperative style, high-performance deep learning library. *Advances in neural information processing systems*, 32, 2019.

[225] Josh Patterson and Adam Gibson. *Deep learning : A practitioner's approach.* " O'Reilly Media, Inc.", 2017.

[226] Dabal Pedamonti. Comparison of non-linear activation functions for deep neural networks on MNIST classification task. *arXiv preprint arXiv :1804.02763*, 2018.

[227] David L Penn, Ricardo J Komotar, and E Sander Connolly. Hemodynamic mechanisms underlying cerebral aneurysm pathogenesis. *Journal of Clinical Neuroscience*, 18(11) :1435–1438, 2011.

[228] Fernando Pérez-García, Rachel Sparks, and Sébastien Ourselin. TorchIO : a python library for efficient loading, preprocessing, augmentation and patch-based sampling of medical images in deep learning. *Computer Methods and Programs in Biomedicine*, 208 :106236, 2021.

[229] Olivier Petit, Nicolas Thome, Clement Rambour, Loic Themyr, Toby Collins, and Luc Soler. U-net transformer : Self and cross attention for medical image segmentation. In *Machine Learning in Medical Imaging : 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12*, pages 267–276. Springer, 2021.

[230] Alexander R Podgorsak, Ryan A Rava, Mohammad Mahdi Shiraz Bhurwani, Anusha R Chandra, Jason M Davies, Adnan H Siddiqui, and Ciprian N Ionita. Automatic radiomic feature extraction using deep learning for angiographic parametric imaging of intracranial aneurysms. *Journal of neurointerventional surgery*, 12(4) :417–421, 2020.

[231] Michael B Pritz. Cerebral aneurysm classification based on angioarchitecture. *Journal of Stroke and Cerebrovascular Diseases*, 20(2) :162–167, 2011.

[232] Ines Rahmany, Mohamed El Arbi Nemmala, Nawres Khlifa, and Houda Megdiche. Automatic detection of intracranial aneurysm using LBP and fourier descriptor in angiographic images. *International journal of computer assisted radiology and surgery*, 14 :1353–1364, 2019.

[233] Kristóf Ralovich, Matthias John, Estelle Camus, Nassir Navab, and Tobias Heimann. 6Dof catheter detection, application to intracardiac echocardiography. In *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2014 : 17th International Conference, Boston, MA, USA, September 14-18, 2014, Proceedings, Part II 17*, pages 635–642. Springer, 2014.

[234] Ramin Ranjbarzadeh, Abbas Bagherian Kasgari, Saeid Jafarzadeh Ghoushchi, Shokofeh Anari, Maryam Naseri, and Malika Bendechache. Brain tumor segmentation based on deep learning and an attention mechanism using MRI multi-modalities brain images. *Scientific Reports*, 11(1) :10930, 2021.

[235] Sebastian Raschka. Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv :1811.12808*, 2018.

[236] Muhammad Imran Razzak, Saeeda Naz, and Ahmad Zaib. Deep learning for medical image processing : Overview, challenges and the future. *Classification in BioApps : Automation of Decision Making*, pages 323–350, 2018.

[237] Joseph Redmon, Santosh Divvala, Ross Girshick, and Ali Farhadi. You only look once : Unified, real-time object detection. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 779–788, 2016.

[238] Joseph Redmon and Ali Farhadi. Yolov3 : An incremental improvement. *arXiv preprint arXiv :1804.02767*, 2018.

[239] Shaoqing Ren, Kaiming He, Ross Girshick, and Jian Sun. Faster r-cnn : Towards real-time object detection with region proposal networks. *Advances in neural information processing systems*, 28, 2015.

[240] Yun Ren, Changren Zhu, and Shunping Xiao. Small object detection in optical remote sensing images via modified faster R-CNN. *Applied Sciences*, 8(5) :813, 2018.

[241] Félix Renard, Soulaimane Guedria, Noel De Palma, and Nicolas Vuillerme. Variability and reproducibility in deep learning for medical image segmentation. *Scientific Reports*, 10(1) :13724, 2020.

[242] Hamid Rezatofighi, Nathan Tsoi, JunYoung Gwak, Amir Sadeghian, Ian Reid, and Silvio Savarese. Generalized intersection over union : A metric and a loss for bounding box regression. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 658–666, 2019.

[243] Zakaria Rguibi, Abdelmajid Hajami, Dya Zitouni, Amine Elqaraoui, and Anas Bedraoui. Cxai : Explaining convolutional neural networks for medical imaging diagnostic. *Electronics*, 11(11) :1775, 2022.

[244] Leslie Rice, Eric Wong, and Zico Kolter. Overfitting in adversarially robust deep learning. In *International Conference on Machine Learning*, pages 8093–8104. PMLR, 2020.

[245] O. Ronneberger, P. Fischer, and T. Brox. U-net : Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, 2015.

[246] Saikat Roy, Gregor Koehler, Constantin Ulrich, Michael Baumgartner, Jens Petersen, Fabian Isensee, Paul F Jaeger, and Klaus H Maier-Hein. Mednext : transformer-driven scaling of convnets for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 405–415. Springer, 2023.

[247] Raphael H Sacho, Guillaume Saliou, Alex Kostynskyy, Ravi Menezes, Michael Tymianski, Timo Krings, Ivan Radovanovic, Karel Terbrugge, Gabriel JE Rinkel, and Robert Willinsky. Natural history and outcome after treatment of unruptured intradural fusiform aneurysms. *Stroke*, 45(11) :3251–3256, 2014.

[248] Chander Sadasivan, David J Fiorella, Henry H Woo, and Baruch B Lieber. Physical factors effecting cerebral aneurysm pathophysiology. *Annals of biomedical engineering*, 41 :1347–1365, 2013.

[249] Omer Sagi and Lior Rokach. Ensemble learning : A survey. *Wiley Interdisciplinary Reviews : Data Mining and Knowledge Discovery*, 8(4) :e1249, 2018.

[250] Rjiba Saifeddine, Urruty Thierry, Bourdon Pascal, Maloigne-Fernandez Christine, Delepaul Roger, and Remy Guillevin. Team Xlim. In *ADAM*, 2020.

[251] Seyed Sadegh Mohseni Salehi, Deniz Erdogmus, and Ali Gholipour. Tversky loss function for image segmentation using 3D fully convolutional deep networks. In *International workshop on machine learning in medical imaging*, pages 379–387. Springer, 2017.

[252] Sebastian Sanchez, Meredith Hickerson, Rishi R Patel, Dana Ghazaleh, Ranita Tarchand, Geeta S Paranjape, Holly Pope, Santiago Ortega-Gutierrez, John M Pederson, David Hasan, et al. Morphological Characteristics of Ruptured Brain Aneurysms : A Systematic Literature Review and Meta-Analysis. *Stroke : Vascular and Interventional Neurology*, 3(2) :e000707, 2023.

[253] IH Sarker. Machine learning : algorithms, real-world applications and research directions. SN computer science 2 : 160, 2021.

[254] Marcel P Schilling, Tim Scherr, Friedrich R Münke, Oliver Neumann, Mark Schutera, Ralf Mikut, and Markus Reischl. Automated annotator variability inspection for biomedical image segmentation. *IEEE access*, 10 :2753–2765, 2022.

[255] P Thomas Schoenemann. Evolution of the size and functional areas of the human brain. *Annu. Rev. Anthropol.*, 35 :379–406, 2006.

[256] Bernhard Schölkopf and Alexander J Smola. *Learning with kernels : support vector machines, regularization, optimization, and beyond*. MIT press, 2002.

[257] Laligam N Sekhar and Roberto C Heros. Origin, growth, and rupture of saccular aneurysms : a review. *Neurosurgery*, 8(2) :248–260, 1981.

[258] Arnaud Arindra Adiyoso Setio, Alberto Traverso, Thomas De Bel, Moira SN Berens, Cas Van Den Bogaard, Piergiorgio Cerello, Hao Chen, Qi Dou, Maria Evelina Fantacci, Bram Geurts, et al. Validation, comparison, and combination of algorithms for automatic detection of pulmonary nodules in computed tomography images : the LUNA16 challenge. *Medical image analysis*, 42 :1–13, 2017.

[259] Daniel M Sforza, Christopher M Putman, and Juan Raul Cebral. Hemodynamics of cerebral aneurysms. *Annual review of fluid mechanics*, 41 :91–107, 2009.

[260] Fahad Shamshad, Salman Khan, Syed Waqas Zamir, Muhammad Haris Khan, Munawar Hayat, Fahad Shahbaz Khan, and Huazhu Fu. Transformers in medical imaging : A survey. *Medical Image Analysis*, page 102802, 2023.

[261] Yoli Shavit and Ron Ferens. Introduction to camera pose estimation with deep learning. *arXiv preprint arXiv :1907.05272*, 2019.

[262] Micah J Sheller, Brandon Edwards, G Anthony Reina, Jason Martin, Sarthak Pati, Aikaterini Kotrotsou, Mikhail Milchenko, Weilin Xu, Daniel Marcus, Rivka R Colen, et al. Federated learning in medicine : facilitating multi-institutional collaborations without sharing patient data. *Scientific reports*, 10(1) :12598, 2020.

[263] Zhao Shi, Chongchang Miao, U Joseph Schoepf, Rock H Savage, Danielle M Dargis, Chengwei Pan, Xue Chai, Xiu Li Li, Shuang Xia, Xin Zhang, et al. A clinically applicable deep-learning model for detecting intracranial aneurysm in computed tomography angiography images. *Nature communications*, 11(1) :6090, 2020.

[264] Takashi Shirasaka, Akio Hiwatashi, Koji Yamashita, Masatoshi Kondo, Hiroshi Hamasaki, Yamato Shimomiya, Yasuhiko Nakamura, Yoshinori Funama, and Hiroshi Honda. Optimal scan timing for artery–vein separation at whole-brain CT angiography using a 320-row MDCT volume scanner. *The British journal of radiology*, 90(1070) :20160634, 2017.

[265] Suprosanna Shit, Ivan Ezhov, Johannes C Paetzold, and Bjoern Menze. A ν-net : Automatic detection and segmentation of aneurysm. In *International workshop on Cerebral Aneurysm Detection*, pages 51–57. Springer, 2020.

[266] Ajay Shrestha and Ausif Mahmood. Review of deep learning algorithms and architectures. *IEEE access*, 7 :53040–53065, 2019.

[267] Abhinav Shrivastava, Abhinav Gupta, and Ross Girshick. Training region-based object detectors with online hard example mining. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 761–769, 2016.

[268] T Sichtermann, A Faron, R Sijben, et al. Deep learning–Based Detection of Intracranial Aneurysms in 3D TOF-MRA. *American Journal of Neuroradiology*, 40(1) :25–32, 2019.

[269] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv :1409.1556*, 2014.

[270] Satya P Singh, Lipo Wang, Sukrit Gupta, Haveesh Goli, Parasuraman Padmanabhan, and Balázs Gulyás. 3D deep learning on medical images : a review. *Sensors*, 20(18) :5097, 2020.

[271] Tao Song, Jieneng Chen, Xiangde Luo, Yechong Huang, Xinglong Liu, Ning Huang, Yinan Chen, Zhaoxiang Ye, Huaqiang Sheng, Shaoting Zhang, et al. CPM-net : A 3D center-points matching network for pulmonary nodule detection in CT scans. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 550–559. Springer, 2020.

[272] Jennifer Soun, Lu-Aung Yosuke Masudathaya, Arabdha Biswas, and Daniel S Chow. The Role of Artificial Intelligence in Neuro-oncology Imaging. *Machine Learning for Brain Disorders*, pages 963–976, 2023.

[273] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout : a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1) :1929–1958, 2014.

[274] J.N. Stember, P. Chang, D.M. Stember, et al. Convolutional neural networks for the detection and measurement of cerebral aneurysms on magnetic resonance angiography. *Journal of digital imaging*, 32(5) :808–815, 2019.

[275] Carole H Sudre, Wenqi Li, Tom Vercauteren, Sebastien Ourselin, and M Jorge Cardoso. Generalised dice overlap as a deep learning loss function for highly unbalanced segmentations. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : Third International Workshop, DLMIA 2017, and 7th International Workshop, ML-CDS 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, Proceedings 3*, pages 240–248. Springer, 2017.

[276] Cecilia Summers and Michael J Dinneen. Nondeterminism and instability in neural network optimization. In *International Conference on Machine Learning*, pages 9913–9922. PMLR, 2021.

[277] Shiliang Sun, Zehui Cao, Han Zhu, and Jing Zhao. A survey of optimization methods from a machine learning perspective. *IEEE transactions on cybernetics*, 50(8) :3668–3681, 2019.

[278] Christian Szegedy, Wei Liu, Yangqing Jia, Pierre Sermanet, Scott Reed, Dragomir Anguelov, Dumitru Erhan, Vincent Vanhoucke, and Andrew Rabinovich. Going deeper with convolutions. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1–9, 2015.

[279] S. A. Taghanaki, K. Abhishek, J.P. Cohen, et al. Deep semantic segmentation of natural and medical images : a review. *Artificial Intelligence Review*, pages 1–42, 2020.

[280] A.A. Taha and A. Hanbury. Metrics for evaluating 3D medical image segmentation : analysis, selection, and tool. *BMC medical imaging*, 15(1) :1–28, 2015.

[281] Nima Tajbakhsh, Jae Y Shin, Suryakanth R Gurudu, R Todd Hurst, Christopher B Kendall, Michael B Gotway, and Jianming Liang. Convolutional neural networks for medical image analysis : Full training or fine tuning ? *IEEE transactions on medical imaging*, 35(5) :1299–1312, 2016.

[282] Yee Liang Thian, Yiting Li, Pooja Jagmohan, David Sia, Vincent Ern Yao Chan, and Robby T Tan. Convolutional neural networks for automated fracture detection and localization on wrist radiographs. *Radiology : Artificial Intelligence*, 1(1) :e180001, 2019.

[283] James H Thrall, Xiang Li, Quanzheng Li, Cinthia Cruz, Synho Do, Keith Dreyer, and James Brink. Artificial intelligence and machine learning in radiology : opportunities, challenges, pitfalls, and criteria for success. *Journal of the American College of Radiology*, 15(3) :504–508, 2018.

[284] Zhi Tian, Chunhua Shen, Hao Chen, and Tong He. Fcos : Fully convolutional one-stage object detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 9627–9636, 2019.

[285] Kimberley M Timmins, Irene C van der Schaaf, Edwin Bennink, Ynte M Ruigrok, Xingle An, Michael Baumgartner, Pascal Bourdon, Riccardo De Feo, Tommaso Di Noto, Florian Dubost, et al. Comparing methods of detecting and segmenting unruptured intracranial aneurysms on TOF-MRAs : The ADAM challenge. *Neuroimage*, 238 :118216, 2021.

[286] R Tongdee, VR Narra, EP Oliveira, W Chapman, KM Elsayes, and JJ Brown. Utility of 3D magnetic resonance imaging in preoperative evaluation of hepatobiliary diseases. *HPB*, 8(4) :311–317, 2006.

[287] Hugo Touvron, Matthieu Cord, Matthijs Douze, Francisco Massa, Alexandre Sablayrolles, and Hervé Jégou. Training data-efficient image transformers & distillation through attention. In *International conference on machine learning*, pages 10347–10357. PMLR, 2021.

[288] Jonathan Tremblay, Thang To, Balakumar Sundaralingam, Yu Xiang, Dieter Fox, and Stan Birchfield. Deep object pose estimation for semantic robotic grasping of household objects. *arXiv preprint arXiv :1809.10790*, 2018.

[289] Nicholas J Tustison, Brian B Avants, Philip A Cook, Yuanjie Zheng, Alexander Egan, Paul A Yushkevich, and James C Gee. N4ITK : improved N3 bias correction. *IEEE transactions on medical imaging*, 29(6) :1310–1320, 2010.

[290] D. Ueda, A. Yamamoto, M. Nishimori, et al. Deep learning for MR angiography : automated detection of cerebral aneurysms. *Radiology*, 290(1) :187–194, 2018.

[291] Jasper RR Uijlings, Koen EA Van De Sande, Theo Gevers, and Arnold WM Smeulders. Selective search for object recognition. *International journal of computer vision*, 104 :154–171, 2013.

[292] Maria Vakalopoulou, Stergios Christodoulidis, Ninon Burgos, Olivier Colliot, and Vincent Lepetit. *Deep Learning : Basics and Convolutional Neural Networks (CNNs)*, pages 77–115. Springer US, New York, NY, 2023.

[293] Jeya Maria Jose Valanarasu, Poojan Oza, Ilker Hacihaliloglu, and Vishal M Patel. Medical transformer : Gated axial-attention for medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021 : 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*, pages 36–46. Springer, 2021.

[294] André JW van der Kouwe, Thomas Benner, Bruce Fischl, Franz Schmitt, David H Salat, Martin Harder, A Gregory Sorensen, and Anders M Dale. On-line automatic slice positioning for brain MR imaging. *Neuroimage*, 27(1) :222–230, 2005.

[295] Willem Jan van Rooij, ME Sprengers, Anjob N de Gast, JPP Peluso, and Menno Sluzewski. 3D rotational angiography : the new gold standard in the detection of additional intracranial aneurysms. *American Journal of Neuroradiology*, 29(5) :976–979, 2008.

[296] Gaël Varoquaux and Veronika Cheplygina. Machine learning for medical imaging : methodological failures and recommendations for the future. *NPJ digital medicine*, 5(1) :48, 2022.

[297] Charles Vega, Jeremiah V Kwoon, and Sean D Lavine. Intracranial aneurysms : current evidence and clinical practice. *American family physician*, 66(4) :601–609, 2002.

[298] Andrey Verendeev and Chet C Sherwood. Human brain evolution. *Current opinion in behavioral sciences*, 16 :41–45, 2017.

[299] Murali Vijayan and P Hemachandra Reddy. Stroke, vascular dementia, and Alzheimer's disease : molecular links. *Journal of Alzheimer's disease*, 54(2) :427–443, 2016.

[300] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. YOLOv7 : Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 7464–7475, 2023.

[301] Qilong Wang, Banggu Wu, Pengfei Zhu, Peihua Li, Wangmeng Zuo, and Qinghua Hu. ECA-net : Efficient channel attention for deep convolutional neural networks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11534–11542, 2020.

[302] Wenhai Wang, Enze Xie, Xiang Li, Deng-Ping Fan, Kaitao Song, Ding Liang, Tong Lu, Ping Luo, and Ling Shao. Pyramid vision transformer : A versatile backbone for dense prediction without convolutions. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 568–578, 2021.

[303] Xiaolong Wang, Ross Girshick, Abhinav Gupta, and Kaiming He. Non-local neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 7794–7803, 2018.

[304] Yaqing Wang, Quanming Yao, James T Kwok, and Lionel M Ni. Generalizing from a few examples : A survey on few-shot learning. *ACM computing surveys (csur)*, 53(3) :1–34, 2020.

[305] Jürgen Weese and Cristian Lorenz. Four challenges in medical image analysis from an industrial perspective, 2016.

[306] Dominik Weishaupt, Victor D Köchli, Borut Marincek, Johannes M Froehlich, Daniel Nanz, and Klaas P Pruessmann. *How does MRI work ? : an introduction to the physics and function of magnetic resonance imaging*, volume 2. Springer, 2006.

[307] Daniel Wesierski and Anna Jezierska. Instrument detection and pose estimation with rigid part mixtures model in video-assisted surgeries. *Medical image analysis*, 46 :244–265, 2018.

[308] Philip M White, Evelyn M Teasdale, Joanna M Wardlaw, and Valerie Easton. Intracranial aneurysms : CT angiography and MR angiography for detection—prospective blinded comparison in a large patient cohort. *Radiology*, 219(3) :739–749, 2001.

[309] Ken CL Wong, Mehdi Moradi, Hui Tang, and Tanveer Syeda-Mahmood. 3d segmentation with exponential logarithmic loss for highly unbalanced object sizes. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018 : 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part III 11*, pages 612–619. Springer, 2018.

[310] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. Cbam : Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*, pages 3–19, 2018.

[311] David Wooden, Matthew Malchano, Kevin Blankespoor, Andrew Howardy, Alfred A Rizzi, and Marc Raibert. Autonomous navigation for bigdog. In *2010 IEEE international conference on robotics and automation*, pages 4736–4741. Ieee, 2010.

[312] Xindong Wu, Vipin Kumar, J Ross Quinlan, Joydeep Ghosh, Qiang Yang, Hiroshi Motoda, Geoffrey J McLachlan, Angus Ng, Bing Liu, Philip S Yu, et al. Top 10 algorithms in data mining. *Knowledge and information systems*, 14 :1–37, 2008.

[313] Yu Xiang, Tanner Schmidt, Venkatraman Narayanan, and Dieter Fox. Posecnn : A convolutional neural network for 6D object pose estimation in cluttered scenes. *arXiv preprint arXiv :1711.00199*, 2017.

[314] Han Xiao, Kashif Rasul, and Roland Vollgraf. Fashion-mnist : a novel image dataset for benchmarking machine learning algorithms. *arXiv preprint arXiv :1708.07747*, 2017.

[315] Chang-Yang Xing, Takashi Tarumi, Jie Liu, Yinan Zhang, Marcel Turner, Jonathan Riley, Cynthia Duron Tinajero, Li-Jun Yuan, and Rong Zhang. Distribution of cardiac output to the brain across the adult lifespan. *Journal of Cerebral Blood Flow & Metabolism*, 37(8) :2848–2856, 2017.

[316] Yan Yan, Rómer Rosales, Glenn Fung, Ramanathan Subramanian, and Jennifer Dy. Learning from multiple annotators with varying expertise. *Machine learning*, 95 :291–327, 2014.

[317] Haichun Yang, Ruining Deng, Yuzhe Lu, Zheyu Zhu, Ye Chen, Joseph T Roland, Le Lu, Bennett A Landman, Agnes B Fogo, and Yuankai Huo. Circlenet : Anchor-free glomerulus detection with circle representation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 35–44. Springer, 2020.

[318] Jiehua Yang, Mingfei Xie, Canpei Hu, Osamah Alwalid, Yongchao Xu, Jia Liu, Teng Jin, Changde Li, Dandan Tu, Xiaowu Liu, et al. Deep learning for detecting cerebral aneurysms with CT angiography. *Radiology*, 298(1) :155–163, 2021.

[319] Xi Yang, Ding Xia, Taichi Kin, and Takeo Igarashi. Intra : 3D intracranial aneurysm dataset for deep learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2656–2666, 2020.

[320] Xiaojiang Yang, Daniel J Blezek, Lionel TE Cheng, William J Ryan, David F Kallmes, and Bradley J Erickson. Computer-aided detection of intracranial aneurysms in MR angiography. *Journal of digital imaging*, 24(1) :86–95, 2011.

[321] Y. Yang, Y. Lin, Y. Li, et al. Automatic Aneurysm Segmentation via 3D U-Net Ensemble. In *Automatic Detection And SegMentation Challenge (ADAM)*, 2020.

[322] Yuan Yao, Lorenzo Rosasco, and Andrea Caponnetto. On early stopping in gradient descent learning. *Constructive Approximation*, 26 :289–315, 2007.

[323] Jingru Yi, Pengxiang Wu, Bo Liu, Qiaoying Huang, Hui Qu, and Dimitris Metaxas. Oriented object detection in aerial images with box boundary-aware vectors. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 2150–2159, 2021.

[324] Tianwei Yin, Xingyi Zhou, and Philipp Krahenbuhl. Center-based 3d object detection and tracking. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 11784–11793, 2021.

[325] Stewart Young, Daniel Bystrov, Thomas Netsch, Rene Bergmans, Arianne van Muiswinkel, Fredy Visser, Rudolf Sprigorum, and Jürgen Gieseke. Automated planning of MRI neuro scans. In *Medical Imaging 2006 : Image Processing*, volume 6144, pages 551–558. SPIE, 2006.

[326] H. Yu, Y. Fan, and H. Shi. Team abc. In *ADAM*. ADAM Challenge, MICCAI, 2020.

[327] Jianhui Yu, Chaoyi Zhang, Heng Wang, Dingxin Zhang, Yang Song, Tiange Xiang, Dongnan Liu, and Weidong Cai. 3d medical point transformer : Introducing convolution to attention networks for medical point cloud analysis. *arXiv preprint arXiv :2112.04863*, 2021.

[328] Yiqiang Zhan, Maneesh Dewan, Martin Harder, Arun Krishnan, and Xiang Sean Zhou. Robust automatic knee MR slice positioning through redundant and hierarchical anatomy detection. *IEEE transactions on medical imaging*, 30(12) :2087–2100, 2011.

[329] Shifeng Zhang, Cheng Chi, Yongqiang Yao, Zhen Lei, and Stan Z Li. Bridging the gap between anchor-based and anchor-free detection via adaptive training sample selection. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 9759–9768, 2020.

[330] Yue-Hui Zhang, Chang-Qing Zhao, Lei-Sheng Jiang, Xiao-Dong Chen, and Li-Yang Dai. Modic changes : a systematic review of the literature. *European Spine Journal*, 17(10) :1289–1299, 2008.

[331] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018.

[332] Hengshuang Zhao, Yi Zhang, Shu Liu, Jianping Shi, Chen Change Loy, Dahua Lin, and Jiaya Jia. Psanet : Point-wise spatial attention network for scene parsing. In *Proceedings of the European conference on computer vision (ECCV)*, pages 267–283, 2018.

[333] Yu Zhao, Ke Zeng, Yiyuan Zhao, Parmeet Bhatia, Mahesh Ranganath, Muhammed Labeeb Kozhikkavil, Chen Li, and Gerardo Hermosillo. Deep learning solution for medical image localization and orientation detection. *Medical Image Analysis*, 81 :102529, 2022.

[334] Li Zhou and Margarita Sordo. Chapter 5 - Expert systems in medicine. In Lei Xing, Maryellen L. Giger, and James K. Min, editors, *Artificial Intelligence in Medicine*, pages 75–100. Academic Press, 2021.

[335] Xingyi Zhou, Dequan Wang, and Philipp Krähenbühl. Objects as points. *arXiv preprint arXiv :1904.07850*, 2019.

[336] Yi Zhou, Connelly Barnes, Jingwan Lu, Jimei Yang, and Hao Li. On the continuity of rotation representations in neural networks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5745–5753, 2019.

[337] Zongwei Zhou, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, and Jianming Liang. Unet++ : A nested u-net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support : 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, pages 3–11. Springer, 2018.

[338] Guangyu Zhu, Qi Yuan, Jian Yang, and Joon Hock Yeo. Experimental study of hemodynamics in the circle of Willis. *Biomedical engineering online*, 14 :1–15, 2015.

[339] Jiuwen Zhu, Yuexiang Li, Yifan Hu, Kai Ma, S Kevin Zhou, and Yefeng Zheng. Rubik's cube+ : A self-supervised feature learning framework for 3D medical image analysis. *Medical image analysis*, 64 :101746, 2020.

[340] Xiangxin Zhu, Carl Vondrick, Charless C Fowlkes, and Deva Ramanan. Do we need more training data ? *International Journal of Computer Vision*, 119(1) :76–92, 2016.

[341] Martin Zlocha, Qi Dou, and Ben Glocker. Improving RetinaNet for CT lesion detection with dense masks from weak RECIST labels. In *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019 : 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22*, pages 402–410. Springer, 2019.

[342] Zhengxia Zou, Keyan Chen, Zhenwei Shi, Yuhong Guo, and Jieping Ye. Object detection in 20 years : A survey. *Proceedings of the IEEE*, 2023.

Résumé

Les anévrismes intracrâniens sont des dilatations locales des vaisseaux sanguins cérébraux, présentant un risque significatif de rupture, susceptible d'entraîner des conséquences graves. La détection précoce des anévrismes non rompus revêt donc une importance cruciale pour prévenir des complications potentiellement mortelles. Cependant, l'analyse des images médicales pour localiser ces anévrismes est une tâche complexe et chronophage, demandant du temps et d'expertise, et malgré cela reste susceptible d'erreurs d'interprétation. Face à ces défis, cette thèse explore des méthodes automatisées pour la détection des anévrismes, visant à faciliter le travail des radiologues et à améliorer l'efficacité du diagnostic. Notre approche se concentre sur l'utilisation de techniques d'intelligence artificielle, en particulier les réseaux de neurones profonds, pour la détection des anévrismes à partir d'images d'IRM angiographique par temps de vol (TOF-MRA). Nos travaux de recherche se sont articulés autour de plusieurs axes principaux. Tout d'abord, en raison de la rareté des données d'entraînement dans le domaine médical, nous adoptons une méthode d'annotation rapide, bien qu'approximative, afin de faciliter la collection de données. Par ailleurs, nous proposons une stratégie basée sur des petits patchs. En association avec de la synthèse de données, les échantillons sont démultipliés dans la base d'apprentissage. En sélectionnant les échantillons, leur distribution est ajustée pour faciliter l'optimisation. Ensuite, pour la détection automatisée des anévrismes, nous étudions diverses architectures de réseaux de neurones. Une première approche explore les réseaux de segmentation d'images. Nous proposons dans un second temps une architecture innovante inspirée des travaux en détection d'objets. Ces architectures, notamment la dernière, obtiennent des résultats compétitifs, en particulier en termes de sensibilité par rapport aux experts. Enfin, au-delà de la détection des anévrismes, nous étendons notre modèle à l'estimation de la pose des anévrismes dans les images 3D. Ceci peut grandement faciliter leur analyse et leur interprétation dans des plans de coupes reformatés. Une évaluation approfondie des modèles proposés est systématiquement réalisée, incluant des études par ablation, l'exploitation de métriques adaptées au problème de la détection et des évaluations menées par des experts cliniciens, permettant d'apprécier leur efficacité potentielle pour une utilisation clinique. Nous mettons en particulier en évidence les problèmes liés à l'incertitude dans l'annotation des bases de données existantes.

Mots-clés: Anévrismes intracrâniens, Angiographie par résonance magnétique, Détection d'objets, Estimation de pose, Apprentissage profond.

Abstract

Intracranial aneurysms are local dilatations of cerebral blood vessels, presenting a significant risk of rupture, which can lead to serious consequences. Early detection of unruptured aneurysms is therefore crucial to prevent potentially fatal complications. However, analyzing medical images to locate these aneurysms is a complex and time-consuming task, requiring time and expertise, and yet remains prone to errors in interpretation. Faced with these challenges, this thesis explores automated methods for the detection of aneurysms, aiming to facilitate the work of radiologists and improve diagnostic efficiency. Our approach focuses on the use of artificial intelligence techniques, particularly deep neural networks, for the detection of aneurysms from time-of-flight magnetic resonance angiography (TOF-MRA) images. Our research work is centered around several main axes. Firstly, due to the scarcity of training data in the medical field, we adopt a rapid, although approximate, annotation method to facilitate data collection. Furthermore, we propose a strategy based on small patches. In association with data synthesis, the samples are multiplied in the training database. By selecting the samples, their distribution is adjusted to facilitate optimization. Secondly, for the automated detection of aneurysms, we investigate various neural network architectures. An initial approach explores image segmentation networks. Then, we propose an innovative architecture inspired by object detection methods. These architectures, especially the latter, lead to competitive results, particularly in terms of sensitivity compared to experts. Thirdly, beyond the detection of aneurysms, we extend our model to estimate the pose of aneurysms in 3D images. This can greatly facilitate their analysis and interpretation in reformatted cross-sectional plans. A thorough evaluation of the proposed models is systematically carried out, including ablation studies, the use of metrics adapted to the problem of detection, and evaluations conducted by clinical experts, allowing us to assess their potential effectiveness for clinical use. In particular, we highlight the issues related to uncertainty in the annotation of existing databases.

Keywords: Intracranial aneurysms, Magnetic resonance angiography, Detection, Pose estimation, Neural networks, Deep learning.