THÈSE POUR OBTENIR LE GRADE DE DOCTEUR DE L'UNIVERSITÉ DE MONTPELLIER

En Informatique

École doctorale : Information, Structures, Systèmes

Unité de recherche : DOI Ifremer, IROKO Inria

Multi-scale mapping of changes in tropical reefs

Présentée par Matteo CONTINI
Le 21 Novembre 2025

Sous la direction d'Alexis JOLY
et Sylvain BONHOMMEAU

Devant le jury composé de

Alexandra BRANZAN ALBU, Professeur, University of Victoria	Rapportrice
Luigi CECCARONI, Directeur de recherche, Earthwatch	Rapporteur
Laura MANNOCCI, Chargée de recherche, IRD	Examinatrice
Dino IENCO, Directeur de recherche, INRAE	Président
Alexis JOLY, Directeur de recherche, INRIA	Directeur de thèse
Sylvain BONHOMMEAU, Cadre scientifique des EPIC, IFREMER	Co-directeur
Julien BARDE, Ingénieur de recherche, IRD	Encadrant (invité)
Serge BERNARD, Directeur de recherche, CNRS	Encadrant (invité)

UNIVERSITÉ DE MONTPELLIER

Abstract

The monitoring of coral reef's health and their biodiversity in shallow waters over large spatial extents represents a major challenge in marine biology. Advances in marine technologies and citizen science initiatives have led to a growing availability of georeferenced image data, creating low-cost opportunities for data collection. However, extracting and mapping ecological information from this large and heterogeneous data stream requires automated and reproducible data processing frameworks.

This thesis contributes to this field through the development of a series of deep learning models aimed at supporting coral reef monitoring from image-level classification to semantic segmentation tasks. Central to this work is the creation of Seatizen Atlas: an open-access dataset of georeferenced underwater and aerial marine imagery curated according to Findable Accessible Interoperable Reusable (FAIR) principles and built to facilitate interdisciplinary research. Using this dataset, we first present a vision transformer trained to classify multiple benthic categories from underwater images, including coral morphotypes, seagrass and associated fauna. We then explore two complementary strategies to extend fine-scale underwater classifications to broader spatial scales using aerial drone imagery. The first leverages knowledge distillation techniques to train an aerial multilabel classification model using predictions from the underwater classifier, enabling broad-area classification without requiring manual aerial annotations. The second approach uses dense georeferenced underwater presence probabilities as weak annotations to train a segmentation model on aerial imagery. Finally, we explore how segmentation models trained on low-altitude drone imagery can be transferred to aerial images captured from higher altitudes.

This thesis demonstrates the potential of integrating diverse data sources with precise georeferencing techniques and innovative Artificial Intelligence (AI) methodologies to provide a scalable framework for the monitoring of coral reefs. It provides a foundation for future developments in coral reef mapping and highlights the importance of open data, interdisciplinary collaboration and methodological transparency in building reproducible ecological observation systems.

Sommaire

Le suivi de l'état de santé des récifs coralliens et de leur biodiversité sur de grandes étendues spatiales représente un défi majeur en biologie marine. Les avancées en matière de technologies marines et les initiatives de sciences participatives ont conduit à une disponibilité croissante d'images, offrant des opportunités de collecte de données à faible coût. Cependant, extraire des informations écologiques de ce flux de données volumineux et hétérogène nécessite des cadres de traitement automatisés et reproductibles.

Cette thèse contribue à ce domaine par le développement d'une série de modèles d'apprentissage profond conçus pour permettre le suivi des récifs coralliens, depuis la classification d'images jusqu'à la segmentation sémantique. Un élément central de ce travail est la création de Seatizen Atlas : un jeu de données en libre accès composé d'images sous-marines et aériennes, structuré selon les principes Findable Accessible Interoperable Reusable (FAIR) et conçu pour faciliter la recherche interdisciplinaire. À partir du jeu de données d'images sous-marines à haute résolution, nous présentons d'abord un réseau de neurones entraîné à classer plusieurs catégories benthiques sur des images sous-marines, incluant des morphotypes coralliens, des herbiers marins et la faune associée. Nous explorons ensuite deux stratégies complémentaires pour étendre les prédictions à fine échelle des images sous-marines à des échelles spatiales plus larges à l'aide d'images aériennes de drones. La première repose sur des techniques de distillation des connaissances pour entraîner un modèle de classification multilabel sur des images de drone aérien à faible altitude, en utilisant les prédictions du modèle sous-marin, sans avoir besoin d'annotation manuelle sur les images aériennes. La seconde approche utilise des prédictions de présence d'espèces géoréférencées issues des images sous-marines comme annotations faibles pour entraîner un modèle de segmentation sur ces images aériennes. Enfin, nous montrons comment le modèle de segmentation entraîné sur des images de drones à basse altitude peut être adapté à des images aériennes acquises à plus haute altitude.

Cette thèse montre le potentiel de l'intégration de sources de données diversifiées, associée à des techniques de géoréférencement et à des approches d'Artificial Intelligence (AI), pour développer un cadre évolutif de suivi des récifs coralliens. Elle constitue une base pour de futurs développements en suivi écologique et souligne l'importance des données ouvertes, de la collaboration interdisciplinaire et de la transparence méthodologique dans la construction de systèmes d'observation reproductibles.

Acknowledgements

Merci à Sylvain et à Alexis. Travailler avec vous pendant ces cinq dernières fantastiques années a été une super aventure et j'en garde énormément. J'admire votre façon d' observer le monde et de chercher par pur plaisir de la découverte.
Merci aussi à Julien: si je suis devenu accro au doi, aux principes fair, aux données interopérables et à l'open science c'est grâce à toi. Ce sont des valeurs que je vais adopter et que je garderai précieusement. Merci à Serge car sans toi je n'aurais pas eu l' occasion de faire l'upscaling et toute la deuxième partie de la thèse n'aurait pas eu lieu.

Mi romersi' Victor aka Grosderg car t'as été le meilleur compagnon de route de cette thèse. A partir du moment où je t'ai vu à l'entretien sur ta grosse chaise de geek et ta présentation en LaTeX, j'ai su que t'étais LE Crack. J'espère que tu pourras trouver de quoi satisfaire la curiosité pour tout et n'importe quoi que tu portes dans l'adn. Merci à Sunnith pour avoir transformé Plancha d'un énorme paddleboard sans prénom, en une gracieuse demoiselle maniable nommée Paola. Merci à Magali pour avoir géré toute la partie communication autour de la thèse. Merci à La Poule pour tes réponses aussi opportunes que cryptiques pendant toutes ces années et pour ta gentillesse. Merci à Crics pour ta joie de vivre et ta spontanéité et à Denis, per essere il mio Grande Gigante Gentile. Un grand merci à Ewoinj et Vevette car même si nous ne nous étions pas vus depuis des années, notre relation n'a pas changé du tout. Merci à Jerem pour être vrai, à Thomas pour son altruisme et à Laura pour savoir écouter. Merci Gaétan pour être toujours (ou presque) partant pour une bière et pour tes valeurs. Merci Mohan pour tes conseils. Merci à tous les amis de l'Ifremer Fonnker, de l'Ird et de l'Inria qui m'ont aidé au cours de ces dernières années, en ordre aléatoire merci Hugues, Blandine, Tristan, Nina, Betty, Julie2, Eric, Julie et Max (et Bernie), Cam Ly, Léanne, Pierre, Alois, Victor2, Arthur, Alexandre, Justine, Hervé, Cathy, Jules, IlyassKhoya, Lukas, Amoros et Aina. Merci à tous ceux qui ont partagé un café avec moi. Merci enfin au roi des codardi Cesarone, pour m'avoir donné la possibilité de trouver un ami au lirmm. Merci LIVE, Yummy et Le Camion pour avoir nourri mes journées.

En ouvrant le chapitre des dalons, la liste est infinie et le risque de s'éterniser est grand. Trugarez dit Abelou aka Habiboumiaimeaou, qui m'a appris ce que signifie être breton et avec lequel j'ai partagé tout depuis qu'on s'est connu. Mi romersi' Sunnith (encore) qui m'a introduit à la religion du matossage (et plein d'autres défauts que je garde précieusement). Comme pour Abel, notre amitié a été un coup de foudre. Merci à Marichou qui m'a fait redécouvrir la mer et m'a introduit aux secrets de l'apnée tout

en écoutant du Bob. A Coco pour m'avoir montré qu'il existe des princesses belles à l'intérieur et pour prendre la vie avec une saine légèreté. Merci à Lolo de m'avoir fait découvrir le canyoning et pour être un ami loyal et généreux comme j'en ai rarement rencontré. Merci à Phoenix, Isis, Nenena, Giorlay, Simon et JeanneLaMachine pour avoir partagé un bout de chemin sous le même toit. Merci Marla de Laura pour tous les carrys feu de bois et pour nous avoir accompagnés (poussés?) dans nos délires. Merci Pulk RIP. Merci aux meilleurs voisins que j'ai jamais eu, Antoine, Marion, Lola et Martin. Merci à Arnaud pour m'avoir fait tomber amoureux de l'apnée. Merci à Pierrick pour avoir été là le jour de la compétition. Un merci aussi au frère de Polo pour avoir appris *Mort aux vaches, version corse* à tout le monde. Merci à Laura Barquette pour ton courage et pour rester toujours à côté de moi. Merci à Jean René pour les poules. Une douce pensée à tous les fanfarons, ceux avec qui j'ai gardé contact (Kkb, Brz, Skw, Radio) et ceux avec qui je l'ai perdu. Si je suis encore en France c'est aussi parce que le premier impact que j'ai eu avec ce pays a été magique grâce à vous.

Un grand merci aussi à toutes les colocs montpelliéraines qui m'ont accueillies. Aux Courlitos, car la caz avec jardin et piscine ne me rendait pas nostalgique de la Run (surtout que le rhum arrangé et la zik ne manquaient pas). Merci Bruno, Emma, Max et Mathilde. Aux Miguelitos da Michele pour le ptit coin de paix. Shukran Mabrouk et Max. Pour finir merci à la coloc des Sottoni pour m'avoir permis de créer des lacs sur votre canap.

Grazie alla famiglia che ha sempre accettato di buon grado il fatto che abitassi dall'altra parte del mondo senza farmelo mai pesare. A Katan che voleva che mi fermassi con gli studi alla scuola dell'obbligo e andassi a vendere panini per strada, cosa che mi ha spinto a fare l'esatto contrario e continuare gli studi fino al dottorato. Forza cricio! A Gaspi che è la colonna portante della baracca e dei burattini (marionette?). A Brusio che ha avuto il coraggio di cambiare vita e che viene a trovarmi ovunque vada e Nicolisio che alla fin fine ci ha bagnato il naso a tutti. Un Seheee speciale a Lisaia aka Pauline per aver fatto in modo che Kramek non arrivasse sempre ultimo ai tornei. Grazie Jonas aka Jhonny aka Jennifer per continuare a starmi affianco e esser sempre disponibile nel momento del bisogno. Grazie alla Nuna per l'incredibile forza che trasmette e per ricordarmi sempre quale parola non posso dire. Un grazie pure al Nonno da lassu' che credo sarebbe stato fiero di questo traguardo. Infine grazie Rospa per saltare al mio fianco da stagno in stagno, spero non smetteremo mai. Per non perder le buone abitudini ti invito a cercare anche stavolta tra le righe.

Passiamo ora ai ritals emigrati (e non). Grazie Giulione aka Sottone perché (mi ripetero' dopo tre anni ma non trovo altre parole) sei un vero amico, uno di quelli che si contano sulle dita d'una mano. Grazie Zozzi per esser diventata un'amica veramente beeene. Grazie al furgoncino per le scorribande. Un grazie agli amici di vecchia data: Zicci, Carlone, Lendry, Leddy, Piva, Errichinio, Pergolone. Un grazie a Manfra perché è la prova vivente che i milanesi possono strapparsi di dosso il cemento. Grazie Sofaia per rimanere la mia amicran e per le perle (e non) di saggezza. Grazie a Hnari aka Natiportz aka Illuminatiportz perché ci lega ben piu' di una cicatrice, fratello mio. Un grazie a Ruffo per gli abbracci e per non cambiare mai (e cambiare sempre allo stesso tempo).

A Francone aka 'Uozz perché ogni anno che passa ci scopriamo di piu' e le tue risate mi arrivano dritte al cuore. Grazie Giopi per i consigli e per aprirmi la vi(t)a en tête da quando ci conosciamo. Grazie Albertino per ricordarti di noi. Grazie alla Toscana (terra natale o forse no) e ad Ocha perché c'è. Un grazie a JD per farsi in quattro (in silenzio) per i butei. Grazie a Angioletto perché grazie a te ho almeno un buon motivo per aprire le macchine. Grazie a Frufri aka Lux per andare oltre le nostre differenze e apprezzarmi (spero) per quel che sono. Un grazie infine a tutti i GLS che ho perso un po' di vista negli ultimi anni ma con cui ho condiviso tanto in passato e spero di riprendere in futuro: Aio, Nene, Bardo, Ganja, Camriup, Nd, Bea, Farhatauei, Apoz, Bubi, Dipa, Fva, Eug, Abba, Tower.

"Asper comme dit granper, i sa dousman, i sa sirman, i sa lontan", Faux train, Jessica Persée

Contents

Contents **14**

1 Introduction **19**
1.1 Marine ecosystems: richness, risk and responsibility 20
1.1.1 The ocean as the cradle of life 20
1.1.2 Coral reefs: vital ecosystems under threat 21
1.2 Marine monitoring techniques 23
1.2.1 Traditional survey techniques 23
1.2.2 Novel survey techniques 26
1.2.3 Citizen science: a complementary approach 31
1.2.4 Photogrammetry from images collected by the different platforms 32
1.3 Artificial Intelligence: a new frontier in marine monitoring 34
1.3.1 A short history of artificial intelligence 34
1.3.2 Deep Learning architectures 36
1.3.3 Computer vision tasks 37
1.3.4 Trends in deep learning and computer vision 40
1.3.5 The cost of understanding: interpretability in high-performance models 42
1.4 AI in marine vision 43
1.5 Aims of the thesis and thesis structure 45

2 Seatizen atlas: a collaborative dataset of underwater and aerial marine imagery **63**
2.1 Introduction 65
2.2 Materials and Methods 69
2.2.1 Instrumentation 69
2.2.2 Data acquisition system 69
2.2.3 Video / camera system 70
2.2.4 Data synchronization 70
2.2.5 Image geolocation 71
2.2.6 Data annotation 72
2.3 Data Records 74
2.3.1 Session deposit 74

2.3.2 Global deposit . . . 78
2.3.3 Annotation deposit . . . 80
2.4 Technical Validation . . . 81
2.4.1 Annotation methodology . . . 81
2.4.2 Spatial data accuracy . . . 81

3 Underwater marine biodiversity mapping in remote islands 87
3.1 Introduction . . . 89
3.2 Materials and methods . . . 89
3.2.1 Study sites . . . 89
3.2.2 Underwater image acquisition . . . 93
3.2.3 Image classification . . . 96
3.2.4 Photogrammetry survey . . . 101
3.3 Results . . . 101
3.3.1 Image classification . . . 101
3.3.2 Prediction maps . . . 103
3.3.3 ASV bathymetry . . . 104
3.3.4 Photogrammetry survey . . . 104
3.4 Discussion . . . 105
3.4.1 Image classification . . . 105
3.4.2 Prediction maps . . . 108
3.4.3 ASV bathymetry . . . 109
3.4.4 Photogrammetry survey . . . 110
3.4.5 Potential improvements . . . 111

4 From underwater to drone: a novel multi-scale knowledge distillation approach for coral reef monitoring 117
4.1 Introduction . . . 119
4.2 Materials and methods . . . 122
4.2.1 Underwater image acquisition . . . 122
4.2.2 Aerial image acquisition . . . 122
4.2.3 Multi-scale positioning . . . 124
4.2.4 Underwater image classification . . . 125
4.2.5 Upscaling predictions . . . 125
4.2.6 Test zone and model evaluation . . . 133
4.3 Results . . . 134
4.3.1 Upscaling process evaluation . . . 134
4.3.2 Prediction maps . . . 135
4.4 Discussion . . . 135
4.4.1 Transferring information across scales: model independence and flexibility . . . 136
4.4.2 Georeferencing challenges in multi-scale monitoring . . . 137
4.4.3 Expanding spatial coverage and species identification . . . 138
4.4.4 AI and biodiversity monitoring . . . 139

5 The point is the mask: scaling coral reef segmentation with weak supervision 145
5.1 Introduction . . . 147
5.2 Materials and methods . . . 148
5.2.1 Underwater data acquisition and image processing . . . 149
5.2.2 Interpolation & rasterization . . . 151
5.2.3 Aerial data acquisition and image processing . . . 152
5.2.4 Aerial segmentation dataset . . . 152
5.2.5 Aerial deep learning model (student model) . . . 157
5.2.6 Mask refinement and model retraining . . . 158
5.2.7 Test zone and model evaluation . . . 158
5.3 Results . . . 158
5.3.1 Evaluation of model variants during pipeline development . . . 158
5.3.2 Model performance on annotated reef zones . . . 159
5.3.3 Predicted habitat composition across reef zones . . . 160
5.4 Discussion . . . 161
5.4.1 WSSS loss function . . . 162
5.4.2 Aerial model evaluation . . . 163
5.4.3 Expanding to new ecological classes . . . 164
5.4.4 Comparing multi-label classification and semantic segmentation approaches . . . 166

6 Assessing the potential of upscaling from low- to high-altitude aerial imagery 173
6.1 Introduction . . . 175
6.2 Materials and Methods . . . 177
6.2.1 Aerial imagery data sources . . . 177
6.2.2 Semantic segmentation . . . 177
6.3 Results . . . 179
6.3.1 High-altitude model evaluation . . . 179
6.3.2 Predicted habitat composition in the Hermitage / La Saline lagoon 180
6.4 Discussion . . . 181

7 Conclusion and perspectives 189
7.1 Synthesis of results and contributions . . . 190
7.2 Limits and future directions . . . 192
7.2.1 Data advancements . . . 193
7.2.2 Technique advancements . . . 194
7.2.3 Ecological advancements . . . 196

Appendices 205
1 Supplementary material Chapter 2: Seatizen atlas: a collaborative dataset of underwater and aerial marine imagery . . . 206
1.1 Annotated classes . . . 206

2 Supplementary material Chapter 3: Underwater marine biodiversity mapping in remote islands . . . 223
2.1 ASV specifications . . . 223
2.2 ASV camera settings . . . 223
2.3 ASV bathymetry . . . 224
2.4 Remote data georeferencing . . . 225
2.5 Per-class F1 scores . . . 226
3 Supplementary material Chapter 4: From underwater to drone: a novel multi-scale knowledge distillation approach for coral reef monitoring . . . 227
3.1 UAV mission planning . . . 227
3.2 UAV mission execution . . . 227

Acronyms

AI Artificial Intelligence. iii, v, 20, 31, 32, 37, 38, 43, 45, 46, 64, 65, 68, 73, 80, 83, 88, 96, 109, 112, 118, 135, 137, 139, 141, 164, 177, 184, 191, 193, 194, 196

ASV Autonomous Surface Vehicle. 27, 29–31, 33, 43–46, 63–67, 69, 71, 74, 75, 77, 83, 87, 89, 90, 92–94, 96, 101, 103–106, 110–112, 118, 119, 121, 122, 125, 126, 128, 137, 147–158, 160, 161, 164, 166–168, 174, 184, 191, 194, 204, 223–226

AUC Area Under the Curve. 134, 136

AUV Autonomous Underwater Vehicle. 26, 27, 29

BCE Binary Cross-Entropy. 100, 132, 155

CE Cross-Entropy. 130, 162, 163

CNN Convolutional Neural Network. 36, 37, 43, 98, 119, 157

COCO Common Objects in Context. 66, 69, 80

CV Computer Vision. 20, 36–38, 40, 43, 89, 98, 158, 190, 192, 194, 196

DL Deep Learning. 35–38, 42, 43, 46, 77, 83, 88, 89, 96, 112, 118, 120, 121, 133, 137, 149, 158, 168, 192

DOI Digital Object Identifier. 45, 64, 65, 74, 79, 80

eDNA Environmental DNA. 23, 193

EXIF Exchangeable Image File Format. 66, 69, 71, 76, 95, 139

FAIR Findable Accessible Interoperable Reusable. iii, v, 31, 32, 45, 46, 64, 65, 69, 83, 139, 191

FOV Field of View. 25, 26, 101, 126

GBIF Global Biodiversity Information Facility. 31, 66, 67, 73, 80, 81, 139

GCP Ground Control Point. 72, 83, 110, 111, 122–124, 137, 138, 152, 177, 204

GCRMN Global Coral Reef Monitoring Network. 65, 66, 69, 72, 81, 96, 97, 139

GIS Geographic Information System. 66, 78, 104, 119, 124, 125, 192, 225

GNSS Global Navigation Satellite System. 64, 79, 81, 82, 95, 110, 195, 225, 226

GPU Graphics Processing Unit. 36, 99

GSD Ground Sampling Distance. 29, 110, 122, 126, 152, 160, 177

IFRECOR French Initiative for Coral Reefs. 91

IGN French National Institute of Geographic and Forest Information. 28, 31, 47, 124, 125, 137, 174–183, 185

IMU Inertial Measurement Unit. 96, 224

IoU Intersection over Union. 159, 163, 165, 179, 180, 182

KL Kullback-Leibler. 132, 136

LiDAR Light Detection and Ranging. 27, 28

MAE Mean Absolute Error. 132, 136

MARCSI Marine Citizen Science Initiatives. 31

ML Machine Learning. 34–38, 42, 43, 119

MLCM Multi Label Confusion Matrix. 101, 102, 105, 107

NLP Natural Language Processing. 36, 37, 40, 89

OBIA Object-Based Image Analysis. 119, 120

OGC Open Geospatial Consortium. 66, 69

PPK Post-Processing Kinematic. 71, 81, 83, 94, 95, 110, 137, 138, 224, 226

PPP Precise Point Positioning. 83, 95, 110, 226

RMSE Root Mean Square Error. 123, 132, 136

RNN Recurrent Neural Network. 37

ROV Remotely Operated Vehicle. 26, 27, 29

RTK Real Time Kinematic. 76, 83, 94

SAM Segment Anything Model. 40, 41, 147, 158, 164, 194

SDM Species Distribution Model. 134

SfM Structure from Motion. 32–34, 43, 72, 77, 120, 122, 125, 126, 152, 177, 227

SOTA State Of The Art. 35, 89, 98, 137, 196

SVM Support Vector Machines. 119, 120

UAV Unmanned Aerial Vehicle. 27–30, 43–45, 47, 64–67, 69, 71, 72, 74, 75, 77, 83, 89, 92, 108, 111, 112, 118, 119, 121, 122, 138, 141, 146–152, 154, 157, 159–161, 163–168, 174–185, 191, 192, 194–197

USV Unmanned Surface Vehicle. 27, 119, 120

UVC Underwater Visual Census. 23–25, 33, 67, 74, 108, 109, 118, 119, 190

VHR Very High Resolution. 175, 176, 182–185

ViT Vision Transformer. 37, 89

WSSS Weakly Supervised Semantic Segmentation. 41, 44, 146–148, 157, 162, 163, 168, 174, 176, 185, 191, 194, 197

XAI Explainable Artificial Intelligence. 42, 43

XMP Extensible Metadata Platform. 66, 69, 139

Chapter 1

Introduction



Figure 1.1: *Fishing community in the southeastern coast of Madagascar (Manakara). Ocean is a vital source of food, income and cultural identity for coastal communities in tropical regions, ensuring their livelihoods and daily subsistence.*
Image credit: Gaétan Morand, November 2022.

BEFORE delving into the scientific background, it is useful to briefly recall the origins and objectives of this thesis. Between 2020 and 2022, I completed a two-year civic service at Ifremer (Réunion Island), working on a project that combined citizen science with Artificial Intelligence (AI), focusing on Computer Vision (CV) for image recognition. Building on this experience, the idea naturally emerged to design a multidisciplinary PhD project that would continue this line of research. The main objective of this thesis is to monitor coral reef health at large spatial scales, while pursuing the highest possible taxonomic resolution and reducing the costs of data collection and image annotation. This ambition led to the creation of the present project, which was officially launched in October 2022 after obtaining funding from the PPR Océan et climat program.

In the following chapter, we introduce the ecological and technological challenges of coral reef monitoring and review the state of the art in data collection and analysis methods. The chapter concludes with a detailed description of the aims of this PhD and the proposed methodological framework, followed by an outline of the thesis structure.

1.1 Marine ecosystems: richness, risk and responsibility

1.1.1 The ocean as the cradle of life

Oceans span 71% of Earth's surface and contain 97% of its water, making it the largest and most diverse habitat on the planet (Inaba and Hall-Spencer, 2020).

From phytoplankton and photosynthetic bacteria that form the base of the ocean food web to complex marine species such as fish and mammals that regulate ecosystem stability, marine life exhibits a remarkable capacity to adapt across an extensive range of ecological niches. These range from equatorial waters to polar regions, from coral reefs and kelp forests in shallow coastal waters to the cold, dark depths of the abyssal trenches (Snelgrove, 1999). Each species contributes to maintaining the delicate balance of life, emphasizing the critical role of marine ecosystems in supporting the evolution and resilience of Earth's biodiversity (Chapin et al., 1997).

These ecosystems also provide essential resources and services to humanity. Oceans have absorbed the equivalent of 39% of anthropogenic carbon dioxide emissions since the beginning of the industrial era (1750), helping to mitigate climate change (McKinley et al., 2020). In addition, oceans are a vital source of food, providing sustenance for billions of people worldwide and have the potential to meet the food demand of a projected population of 9.2 billion humans by 2050 (Duarte et al., 2009); see Figure 1.1. Finally, phytoplankton, the ecological group of microalgae adapted to live in apparent suspension in water masses, are responsible for about half (49%) of the global net primary production per year of atmospheric oxygen (Naselli-Flores and Padisák, 2023). In other words, one out of every two breaths we take during our lifetime is made possible by phytoplankton. However, marine ecosystems are under increasing pressure from human activities (Duarte et al., 2020); see Figure 1.2.

As reported by the FAO (Food and Agriculture Organization of the United Nations),

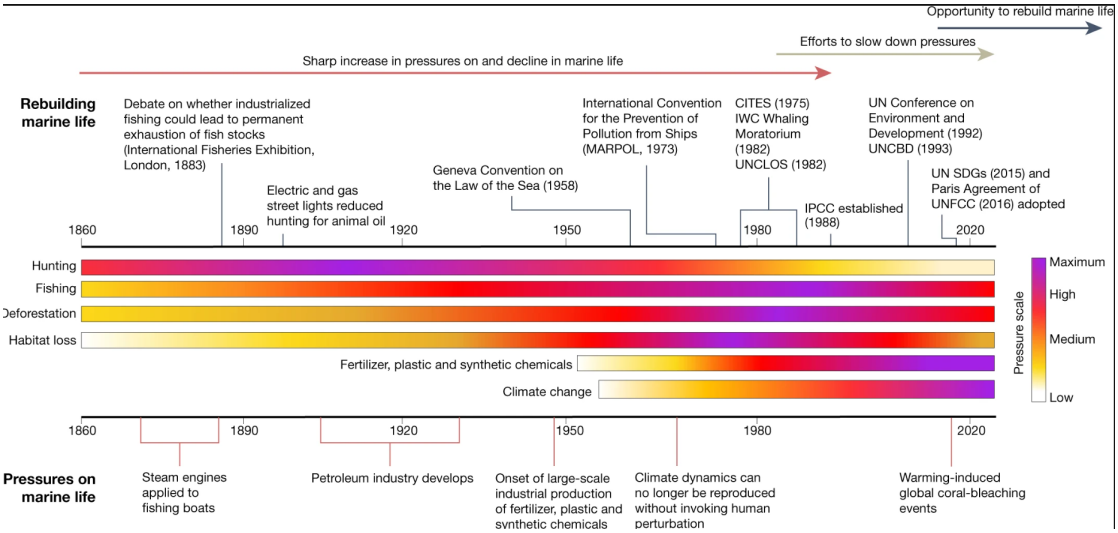


Figure 1.2: *Human pressures on marine ecosystems since 1860. A number of those pressures peaked in the 1980s and are slowing down at present (with great regional variation), with the notable exceptions of pollution and climate change. Figure extracted from Duarte et al. (2020).*

the combined effects of overfishing and habitat destruction through fishing techniques have resulted in the depletion of approximately one third of global fish stocks (Gréboval et al., 2002; Nieto et al., 2015). In the Mediterranean Sea, it is estimated that more than 90% of the stocks are overexploited (Colloca et al., 2017). Among the direct consequences of overfishing, are not only the truncation of the food web (and the corresponding loss of biodiversity as outlined by Pauly et al. (2005)) but also the risk of collapse of the fisheries sector, which would have serious economic repercussions on the livelihoods of millions of people who depend on fishing (Robinson et al., 2017).

In addition to overfishing, marine ecosystems are threatened by human pollution, including plastic waste, metals and chemical contaminants (Cinner et al., 2018; Graham et al., 2011). Microplastics, originating primarily from degraded larger plastics or industrial discharges, directly impact all trophic levels of the marine food web from phytoplankton (Sjollema et al., 2016) to apex predators (McHale and Sheehan, 2024). Finally, anthropogenic climate change is causing significant alterations in marine habitats (Hoegh-Guldberg and Bruno, 2010). Rising greenhouse gas concentrations are leading to ocean warming, acidification and deoxygenation, with cascading effects on marine biodiversity and ecosystem functioning (Figuerola et al., 2021; Lotze et al., 2019; Su et al., 2021).

1.1.2 Coral reefs: vital ecosystems under threat

Coral reefs are underwater ecosystems formed by colonies of tiny animals known as coral polyps. These polyps secrete calcium carbonate to form a hard exoskeleton, which builds

up to create the complex structures we recognize as coral reefs.

Coral reefs are among the most biologically diverse and valuable ecosystems on Earth (Roberts et al., 2002). Despite representing only 0.1% of the ocean's surface, they harbour over 25% of all marine species, making them an essential ecological hotspot (Pinsky et al., 2023).

These factors have led to the establishment of numerous ecological niches and marked geographical isolation, fostering an extraordinary variety of species across all taxa. Surveys estimate that coral reefs are home to more than 950,000 species, including 600 species of scleractinian (hard corals), around 2,000 species of fish and approximately 5,000 species of molluscs (Reaka-Kudla et al., 1997).

Coral reefs also support a wide array of complex food webs and diverse biotic interactions, such as symbiosis, commensalism and parasitism, further promoting biodiversity. These complex relationships between organisms, from tiny plankton to large predators, form the backbone of marine ecosystems (Figure 1.3[1]).

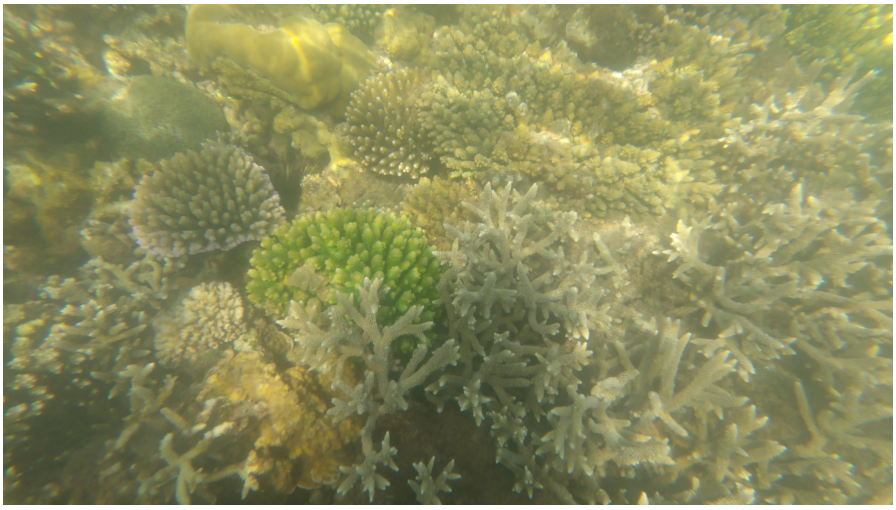


Figure 1.3: *Coral reef community in a tropical marine zone (Ifaty, Madagascar). Reef-building corals create complex habitats that host diverse marine species.*
Image credit: Matteo Contini, May 2023.

With their unique ecological and biological value, coral reefs are indispensable to global marine biodiversity, playing a pivotal role in sustaining life across the oceans. Beyond their ecological significance, coral reefs provide key services to over 500 million people, including food and income (Burke et al., 2011). Moreover, they act as natural barriers, protecting coastlines from erosion and storm surges (Hoegh-Guldberg et al., 2007; Costanza et al., 2014; Ferrario et al., 2014).

However, these ecosystems are under severe threat from climate change, ocean acidification, pollution and overexploitation (Mellin et al., 2022; Buddemeier, 2004). The

[1]The image can be found in Contini et al. (2025) under the relative path:
`20230507_MDG-IFATY_SCUBA-1_00/PROCESSED_DATA/FRAMES/20230507_MDG-IFATY_SCUBA-1_00_8_019.jpeg`

consensus among climate scientists is that, in the absence of prompt and decisive action to curtail the emission of greenhouse gases, the majority of ecosystem services provided by coral reefs will be irreversibly compromised by the end of the century (Rogers et al., 2015).

Protecting coral reefs is not just about preserving biodiversity but also about safeguarding the vital ecosystem services they provide to humanity.

1.2 Marine monitoring techniques

The rapid changes affecting coral reefs and other marine ecosystems in recent decades demand an unprecedented effort in monitoring and surveillance to inform conservation and restoration strategies.

Monitoring the composition of marine communities typically involves two distinct phases:

- Data acquisition: involves collecting information on species composition, abundance and distribution. This phase can be carried out using various methods, including visual surveys, remote sensing and genetic analysis.
- Data analysis: focuses on processing and interpreting the collected data to identify trends, patterns and potential threats to marine communities.

A wide range of methodologies exists for observing marine communities, each suited to different spatial scales, species or environments. Broadly, these methods can be classified into two major categories:

- Direct observation methods: These rely on direct observation of marine organisms and habitats, either by divers (Caldwell et al., 2016), fixed cameras (Pelletier et al., 2021), or autonomous vehicles (Gogendeau et al., 2025; Øie Kvile et al., 2024).
- Indirect detection methods: they use indirect detection of marine life or environmental parameters using advanced technologies, such as Environmental DNA (eDNA) sampling (Lamperti et al., 2023, 2024), acoustic monitoring (Elise et al., 2019a,b) and genetic sequencing (Chevrier et al., 2025).

In this work, we concentrate on **observation-based methods**. In the following sections, we will analyse the most common methods used for monitoring marine communities.

1.2.1 Traditional survey techniques

Underwater visual census

The Underwater Visual Census (UVC) is one of the most widely used methods for assessing communities in marine ecosystems (Stuart-Smith et al., 2017). This method

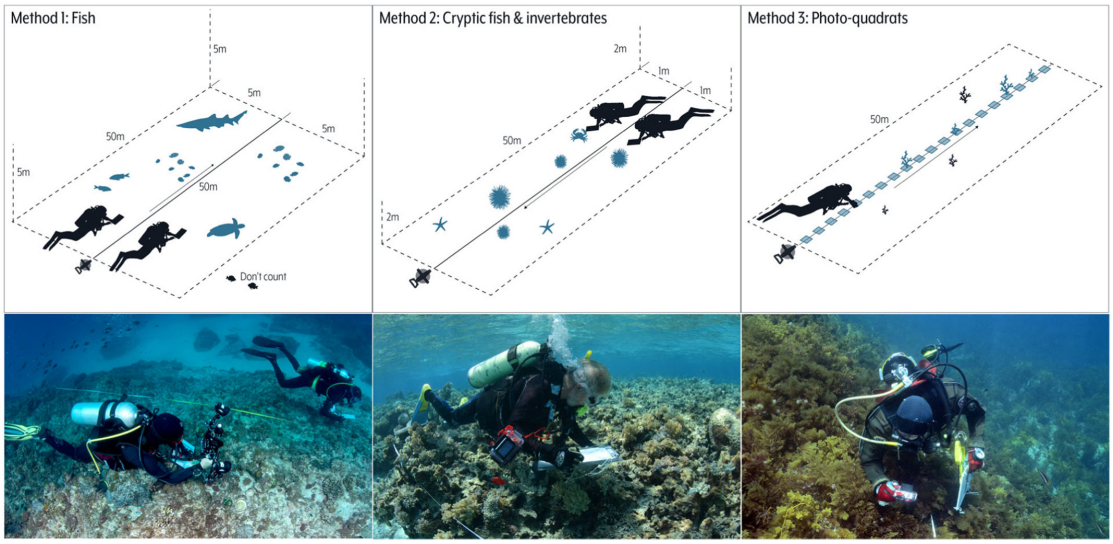


Figure 1.4: *Representation of an underwater visual census. Figure extracted from Edgar et al. (2020).*

involves direct observation by experienced divers who record data on species composition, abundance and size (Emslie et al., 2020).

At each site, a 200-meter transect line is laid out, divided into four 50-meter sections, placed along a specific depth contour (Edgar et al., 2020). Three different surveys are then conducted along this transect (Figure 1.4), each targeting distinct taxonomic groups:

- **Survey 1:** Focuses on fish species. Surveys are conducted within two bands measuring 5 m wide and 5 m high, running parallel to each 50 m section of the transect line.
- **Survey 2:** Targets macro-invertebrates and cryptic fish. Observations are made within a 1 m wide and 2 m high band located on one side of the transect line.
- **Survey 3:** Examines sessile invertebrates and macroalgae. Data are collected using 50 cm × 50 cm quadrats placed at 10 m intervals along the transect line, with 5 quadrats per 50 m segment.

UVC provides direct, real-time observation of marine organisms, enabling data acquisition and analysis in a single step. Its non-destructive nature makes it particularly suitable for fragile ecosystems, such as coral reefs. One of the method's key strengths is its high taxonomic resolution (Pelletier et al., 2011). Experienced divers can identify organisms up to the species level, enabling a precise assessment of biodiversity.

Nevertheless, UVC has several limitations (Cardenas et al., 2024). One major drawback is its dependence on the physiological limitations of divers. Surveys are restricted

by dive duration, depth and the number of dives possible per day, which limits the amount of data that can be collected. Moreover, observer bias is a significant issue. The accuracy of species identification, counts and size estimates can vary depending on the diver's expertise, physical condition, or environmental visibility, leading to inconsistencies and challenges in reproducibility. Additionally, as explained by Dickens et al. (2011), diver presence can change the behaviour of some species (typically fishes), causing individuals to flee or approach and introducing biases that are difficult to quantify. Finally, as noted by Broudic et al. (2025), traditional diver-based methods face inherent trade-offs between spatial coverage, sampling density and operational costs. In addition to technical and methodological limitations, UVC also involves high operational costs. Training and employing expert divers requires major investment. Costs include not only salaries and insurance but also travel, equipment maintenance, safety protocols and support logistics such as boats and crews. These financial demands can limit the frequency and spatial coverage of surveys, particularly in remote or resource-limited regions.

Fixed cameras

Fixed cameras are a widely used tool in marine ecosystem monitoring, particularly for observing fish communities and benthic organisms (Murphy and Jenkins, 2010). Cameras are deployed in a stationary position on the seafloor or attached to fixed structures, allowing researchers to capture video data over a defined area for extended periods. Camera systems may include additional features, such as:

- Baited Remote Underwater Video Station: Attract fish to the camera by deploying bait, providing a standardized method for assessing fish abundance and diversity (Cappo et al., 2006).
- Stereo-video systems: Capture 3D images of marine organisms, enabling accurate size measurements and species identification (Harvey et al., 2002).
- Rotational mechanisms: Allow the camera to rotate 360 degrees, expanding the Field of View (FOV) and capturing a more complete picture of the ecosystem (Pelletier et al., 2021).

Fixed cameras present several advantages for marine monitoring, making them a valuable tool in studying underwater ecosystems. One of the key benefits of fixed cameras is their non-intrusive nature. By being stationary, these cameras allow marine organisms to resume their natural behaviours without the disturbances typically caused by divers or mobile systems. This minimizes the observer effect and provides more accurate and reliable data on species interactions. Additionally, fixed cameras facilitate long-term data collection, as they can be deployed for extended periods allowing researchers to gather temporal data on species abundance, behaviour and environmental changes. Another significant advantage of fixed cameras lies in their capacity for data reproducibility. The video footage captured by these systems is archived, enabling multiple researchers to re-analyse the same dataset, ensuring consistency and robustness in the interpretation of

findings. Moreover, fixed cameras are cost-effective in the long term since once deployed, they require minimal human intervention.

Despite these benefits, fixed cameras also have limitations. One major drawback is their static FOV, which confines observations to a specific area (except for the Stavirolike rotational systems). This can result in the under-representation of highly mobile or widely dispersed species that move beyond the camera's range or in over-representation of species that frequent the monitored area. Furthermore, the significant volumes of video data generated by fixed cameras present challenges in data processing. Analysing these datasets can be time intensive and resource demanding, thus necessitating the use of automated tools for analysis.

1.2.2 Novel survey techniques

Recent advancements in robotics have introduced new tools for marine biodiversity monitoring that overcome many of the limitations associated with traditional methods (Liu et al., 2024). These modern survey techniques allow for expanded spatial coverage, improved temporal resolution and reduced dependence on human observers (Cardenas et al., 2024). In the following sections, we will explore four key categories of modern survey techniques: underwater, surface, low-altitude aerial and high-altitude aerial platforms and space platforms.

Underwater platforms

Mobile underwater platforms such as Remotely Operated Vehicles (ROVs) and Autonomous Underwater Vehicles (AUVs) have become essential tools in modern marine monitoring, enabling systematic surveys in deep (Brignone et al., 2023) and challenging environments (Peck et al., 2024). ROVs are tethered underwater robots controlled from a surface vessel. Typically equipped with cameras, lights and a range of sensors, they provide real-time, high-resolution visual and environmental data. Their manoeuvrability makes them ideal for surveys, close-up inspections of reef structures and operations in areas that are too deep or dangerous for divers. Moreover, when equipped with manipulator arms or specific tools, ROVs can perform tasks such as sample collection, bringing specimens to the surface for further analysis (Chaloux et al., 2021) or even conducting minor repairs on underwater infrastructure. Beyond their technical capabilities, industrial ROVs (originally developed for sectors like oil and gas) have demonstrated great potential for advancing ocean science by collecting vast amounts of ecological data during routine operations (Macreadie et al., 2018). The main drawback is that they require significant logistical infrastructure, including support vessels and trained operators. These requirements make ROV operations costly and limit their accessibility, particularly in remote or resource-limited regions

In contrast, AUVs operate independently of a power cable and are pre-programmed to follow specific survey paths. This autonomy allows them to cover greater distances and access confined or hard-to-reach environments with minimal disturbance. AUVs can carry similar sensor payloads to ROVs (including imaging systems, sonar and water-

quality sensors) but lack real-time feedback unless periodically recovered or paired with acoustic communication systems. Projects like ReefScan CoralAUV illustrate the use of advanced AUV platforms for detailed coral reef assessments (Dobson et al., 2024).

Surface platforms

Autonomous Surface Vehicles (ASVs) have emerged as valuable tools for ecological and environmental monitoring, particularly due to their versatility and affordability (Ackleson et al., 2017). ASVs operate on the water surface, capable of autonomously collecting high-resolution, geolocated data over extensive areas that are often inaccessible or dangerous for human divers. ASVs can integrate diverse sensors specific to various ecological monitoring tasks, including acoustic tracking, photogrammetric surveys and bathymetric mapping (Gogendeau et al., 2025). An important advantage of ASVs compared to Unmanned Surface Vehicle (USV) such as AUVs and ROVs is their ability to directly access GPS signals. Because ASVs operate on the water surface, their antennas remain above water, enabling precise geolocation and simplified navigation. This capability significantly improves spatial accuracy and reduces the complexity and costs associated with underwater positioning systems required by submerged vehicles.

Despite these advantages, ASVs have operational constraints, including sensitivity to wave conditions and potential navigation challenges in congested waterways. Moreover, in deep waters or areas of significant bathymetric variability, imagery collected by ASVs can be excessively bluish and lose detail, primarily due to increased depth and water turbidity. Therefore, while ASVs perform well in shallow-water environments, such as lagoons (e.g., Réunion Island), image acquisition at depths beyond 10–12 meters generally benefits from underwater platforms such as ROVs or AUVs, which can maintain consistent proximity to the seabed and provide higher-quality imagery under challenging conditions.

Low-altitude aerial platforms

Unmanned Aerial Vehicles (UAVs) have become a widely used tool for monitoring shallow-water habitats such as coral reefs (Valdazo et al., 2024), seagrass beds (Amone-Mabuto et al., 2024) and mangroves (Basyuni et al., 2025). UAVs can rapidly collect geo-referenced imagery across extensive and often inaccessible areas, offering a cost-effective and flexible alternative to traditional diver-based or boat-based surveys. Beyond standard rgb cameras, UAV platforms can be equipped with advanced sensors, including Light Detection and Ranging (LiDAR) for capturing detailed elevation data and multispectral or hyperspectral cameras for flood modelling, mapping marine macrophyte meadows or assessing benthic composition (Li et al., 2021; Román et al., 2021; Yang et al., 2022). When combined with machine learning algorithms, UAV imagery facilitates the automated classification of benthic environments, significantly accelerating data processing workflows (Doukari and Topouzelis, 2022). Broad habitat categories such as sand, seagrass and hard bottom can generally be identified with high reliability from aerial imagery (Qu et al., 2024; Øie Kvile et al., 2024).

However, UAV-based monitoring faces several constraints. Environmental conditions such as weather, water clarity and operational factors like battery life and regulatory restrictions can significantly influence data quality and coverage. Additionally, the limited spatial resolution of UAV imagery often makes it challenging to accurately identify finer-scale features, such as specific coral morphotypes or small benthic organisms, especially in complex or mixed assemblages. Therefore, careful ground-truth validation and integration with complementary sources (like underwater imagery or *in situ* human observations) are essential to ensure ecological accuracy.

High-altitude aerial and space platforms

Satellite remote sensing has been a key tool in marine and terrestrial habitat monitoring, providing continuous, large-scale data for monitoring the evolution of ecosystems (Nomenisoa et al., 2024). Optical satellite sensors such as Sentinel-2 deliver imagery suitable for producing continental-scale species distribution maps, biodiversity indicator maps and habitat maps (Leblanc et al., 2022, 2025). For example, Sentinel or Pleiades data have been used to measure the spatial footprint of cyclones (Mouquet et al., 2020), to map coral reefs (B. Lyons et al., 2020) or to monitor the coverage evolution of seagrass meadows (Hedley et al., 2016). While these sensors provide excellent coverage, their resolution may not be sufficient to detect fine-scale habitat structures (although this is gradually improving, with recent advances over the past decades taking us from Landsat's 30 m, to Sentinel-2's 10 m and down to Pléiades 0.5 m resolution).

Airborne remote sensing platforms, on the other hand, can offer a compromise between the high spatial resolution of UAVs and the large coverage of satellite imagery. Indeed, with a resolution that can reach down to a few tens of centimetres, airborne platforms can capture detailed habitat features while covering larger areas than UAVs. In previous works, airborne hyperspectral imagery has been combined with LiDAR data to quantify the geographic distribution of live corals up to 16 m water depth (Asner et al., 2020) or combined with satellite imagery to identify marine aquaculture cages (Quaade et al., 2024). In this last study, the authors used among other data sources, aerial images operated by the French National Institute of Geographic and Forest Information (IGN) from the *BD Ortho* series. These images are orthorectified aerial photographs captured for each french department with a cadence of 2 to 4 years and a resolution ranging from 50 cm px^{-1} for most departments before 2014 to 15 cm px^{-1} in later years.

However, satellite and airborne imagery also have inherent limitations. Atmospheric and sea surface conditions, including cloud cover and atmospheric scattering, can significantly affect image quality and reliability. In addition, resolution constraints directly influence the types of ecological features that can be detected. For instance, identifying small-scale benthic features such as *Acropora Digitate* or *Non-acropora Encrusting* corals, *Algae Turf* or *Sea urchin* typically requires sub-centimetre resolution, often only achievable with underwater or very-low-altitude UAV imagery. When the objective is to detect larger organisms such as *Sea cucumber*, *Non-acropora Massive*, or *Acropora Tabular*, a centimetre-level resolution is generally sufficient. Finally, for broader-scale classification tasks, such as mapping coral assemblages, sand patches or

seagrass meadows, metric resolution data is often adequate. These examples highlight the importance of selecting imagery according to the ecological targets and integrating multiple observation scales to ensure detailed habitat assessments.

Comparison of marine monitoring platforms

As we have seen, the selection of a platform for marine monitoring depends on several factors, including the ecological context, the spatial and temporal scale of the survey, the available budget and the type of data required. In the following, we will refer to data collected by underwater or surface platforms (e.g., divers, ROVs, ASVs) as fine-scale data, typically characterized by a Ground Sampling Distance (GSD) (the length on the seafloor corresponding to a single image pixel) on the order of millimetres. Medium-scale data correspond to imagery acquired by aerial platforms such as UAVs, with a GSD of a few centimetres. Finally, large-scale data are collected from high-altitude aerial or airborne platforms, with GSD typically in the range of tens of centimetres.

In Table 1.1 we summarize the main advantages and limitations of each platform type across classical ecological monitoring criteria.

Table 1.1: *Comparison of platform types across selected ecological monitoring criteria. Symbols: ++ (strong advantage), + (moderate), – (limitation), – – (strong limitation).*

Criteria	**ROVs/AUVs**	**ASVs**	**UAVs**	**Aircraft/ Satellite**
Image resolution	++	+	–	– –
Spatial coverage	– –	–	+	++
Deployment cost	– – / –	+	++	– –
Human deployment	– –	–	+	++
Weather sensitivity	++	+	–	– –
Operational depth	++	+	–	– –
Real-time data access	++ / – –	+	+	– –
Sensor payload flexibility	++	++	–	– –

For instance, satellite imagery may be best suited for monitoring large-scale changes in coastal ecosystems, while underwater platforms such as ROVs or AUVs are more appropriate for detailed studies of specific habitats or species, particularly in deep or hard-to-access environments. However, ecological questions are often complex and multifaceted, requiring a combination of data types and spatial scales to be fully understood. In such cases, no single platform can provide a complete picture. A combined approach integrating data from different scales is then required.

In the context of this PhD, the main objective was to monitor coral reef health at large spatial scales, while pursuing the highest possible taxonomic resolution and reducing the costs of both data collection and image annotation. Among the criteria presented in Table 1.1, the first three were particularly important: *image resolution*, which directly affects the precision of benthic class identification; *spatial coverage*, nec-

essary for extending monitoring to regional and national scales and *deployment cost*, which determines the feasibility of long-term monitoring in resource-limited contexts. The strategy developed in this work seeks to combine the advantages of ASVs for their fine-scale image resolution with those of high-altitude aircraft and satellite platforms for their cost-efficiency and broad coverage, using UAV imagery as a bridging scale between the two.

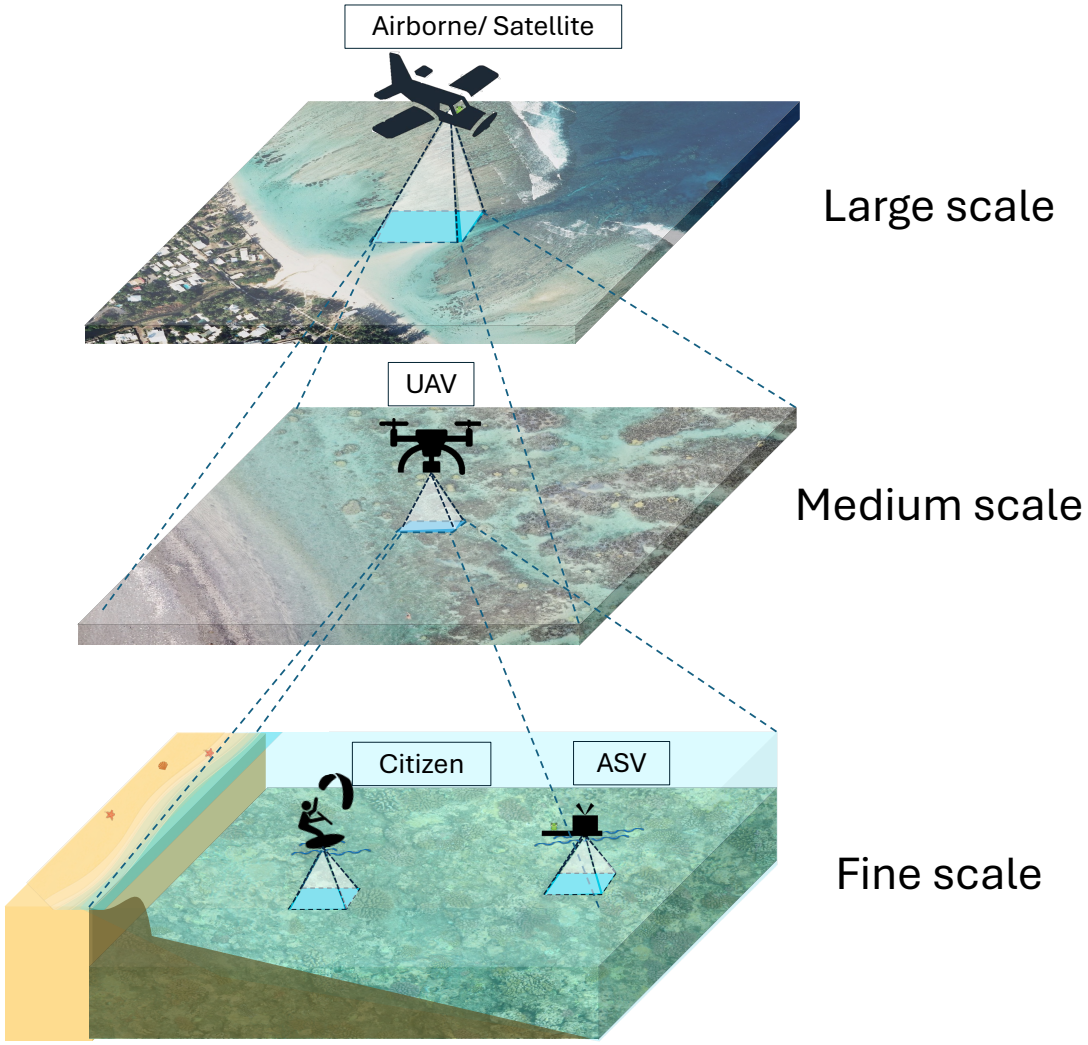


Figure 1.5: *Schema of the different platforms used in this thesis, showcasing the integration of multiple platforms to obtain complementary data. Citizen and ASV underwater images capture fine-scale information about benthic communities, which are passed to aerial drones and then to airborne images thanks to bathymetric and geolocation data. Image credit: Letizia Lamperti.*

In Figure 1.5 we present the different platforms used in this thesis, showcasing the integration of multiple platforms to obtain complementary data. Underwater images captured either through a participatory science project or ASVs provide fine-scale information about coral morphotypes, benthic communities and habitats, which is then passed to aerial drones thanks to bathymetric and geolocation data. Repeating this process with airborne images gathered by the IGN *BD Ortho* series allows for the monitoring of larger areas and the detection of changes over time.

1.2.3 Citizen science: a complementary approach

Citizen science has emerged as a valuable tool for expanding the reach and impact of ecological monitoring by enabling non-experts to contribute to scientific research (Fraisl et al., 2022). By leveraging the participation of the public, citizen science not only generates large volumes of observation data but also raises awareness, engagement and education within local communities (Kullenberg and Kasperowski, 2016; Vohland et al., 2021). In the context of conservation, such initiatives have been particularly useful, given that data collection is frequently constrained by limited resources and personnel and they have been applied in various ecological fields. For instance, land citizen scientists have contributed to the monitoring of protected areas (Mandeville et al., 2023; Bonnet et al., 2020), the identification of invasive species (Espitalier et al., 2025) and the evaluation of migration timing and habitat occupancy of migratory bird species (de Camargo Barbosa et al., 2021).

Despite the increasing interest in marine science, Marine Citizen Science Initiatives (MARCSI) are often viewed as trailing behind their terrestrial counterparts, particularly in biodiversity-related citizen science efforts (Wehn et al., 2025). However, the spread of identified MARCSI has been accelerating since the early 2000s, with 1,267 of them that can be ascertained (e.g., (Matabos et al., 2025; Anton et al., 2021)). Unfortunately, in terms of how open and Findable Accessible Interoperable Reusable (FAIR) these data are, there is still a long way to go. A recent study showed that only 20% of Marine Citizen Science Initiatives (MARCSI) projects make their data fully open and an additional 23% partially publish their datasets: significantly below the 57% openness rate observed in citizen science initiatives overall (Wehn et al., 2025; Davis et al., 2023). In some cases this is due to the lack of a permanent identifier for the data, which is essential for ensuring long-term accessibility and findability. But even when data are available, they often lack a semantic description of the observed metadata, with only 15.7% of the non-closed datasets providing expected metadata.

Nevertheless, some initiatives stand out for their ability to provide or support the production of FAIR-compliant data. One major example is iNaturalist, a widely used international platform that enables users to geolocate, document and identify species observations with the help of AI and expert community validation. Integrated into several MARCSI projects (Roberts et al., 2022; Rocha et al., 2024), this platform facilitates standardized observations, the assignment of persistent identifiers and interoperability with other scientific databases. It is worth noting that research-grade iNaturalist observations are regularly shared with the Global Biodiversity Information Facility (GBIF).

This ensures that citizen-contributed data are accessible and reusable via established biodiversity data infrastructure.

In this thesis, a large portion (97.5%) of the annotated underwater images in the Seatizen Atlas image dataset come from the citizen science project *Seatizen*, developed by Ifremer in 2020. This project aimed to establish a large-scale, continuous environmental monitoring program for marine ecosystems by involving recreational sea users (such as snorkellers, kitesurfers and paddleboarders) in the collection of geo-referenced imagery. These data contributed directly to the creation of the Seatizen Atlas (see Chapter 2), a large-scale, standardized dataset of underwater and aerial marine imagery designed to support ecological monitoring, geospatial analysis and AI-based image recognition. This integration demonstrates the potential of well-designed citizen science initiatives to generate data that is both compliant with the FAIR principles and suitable for training modern AI models for ecological monitoring.

1.2.4 Photogrammetry from images collected by the different platforms

In this thesis, images collected from the underwater and aerial platforms described above were also leveraged for 3D reconstruction tasks. By merging these multi-view images, we aimed to create coherent spatial representations of surveyed environments through photogrammetric techniques.

Photogrammetry is a technique that enables three-dimensional (3D) reconstruction and mapping of a scene with overlapping images from different viewpoints. For each image, points corresponding to the same physical features of the object being photographed are identified. By measuring the positions of corresponding points across multiple images, it is possible to accurately determine the spatial coordinates of those features in a three-dimensional space.

A widely used approach in modern photogrammetry is Structure from Motion (SfM), which automates this process by combining feature matching, camera pose estimation and dense reconstruction to generate 3D point clouds from a set of images. In SfM surveys, overlap and sidelap refer to how much each image overlaps with the preceding image and the images on adjacent flight lines. In order to obtain a good 3D model, it is important to have a high overlap and sidelap between images. At least 70% overlap and sidelap are recommended for a good 3D model (Westoby et al., 2012). The final orthophotos, derived from the computed 3D point clouds, are resampled photographs in which the position of a point in the image is directly related to the position of a detail on the ground.

Its use in marine environments dates back to 1968, initially applied by underwater archaeologists (Drap, 2012; Pollio, 2006). Traditionally, expert divers photograph the area of interest from different angles to ensure sufficient coverage for accurate 3D reconstruction. In recent years photogrammetry has gained popularity in coral reef studies, characterizing habitat complexity and monitoring fine-scale changes (Ferrari et al., 2017; Urbina-Barreto et al., 2021; Million et al., 2021). Wang et al. (2024) have also explored the integration of photogrammetry with traditional ecological survey techniques. They

used SfM photogrammetry to assess the consistency and reliability of UVC-based methods for coral cover estimation, highlighting how 3D reconstructions can validate and complement in-situ monitoring protocols. However, most of these studies have focused on surveys performed by divers. In the following sections, we explore how photogrammetry can be performed using autonomously collected data (both underwater and aerial) to reduce dependence on human divers, increase survey repeatability and expand spatial coverage.

Underwater autonomous photogrammetry

Recently, some studies have explored the use of autonomous vehicles for photogrammetry surveys (Johnson-Roberson et al., 2010; Lubczonek et al., 2021). A key challenge in large-area photogrammetry is maintaining positional accuracy. This can be addressed by coupling images with high-resolution GPS and orientation data, improving the georeferencing of the resulting 3D models (Gogendeau et al., 2025). Integrating these external data sources increases both the efficiency and accuracy of the photogrammetric reconstruction.

While ASVs offer a promising platform for conducting photogrammetric surveys, their use is constrained by several environmental factors (see Section 1.2.2). One major limitation is the restricted depth at which the seafloor can be effectively mapped, primarily due to the dependency on water clarity. This significantly affects both benthic visibility and reconstruction accuracy of the resulting reconstructions, limiting the applicability of ASV-based photogrammetry in turbid or low-light conditions.

Aerial photogrammetry

Imaging the seabed, even in tropical environments with very clear waters, is often complicated by natural factors such as wind, waves or reflections of sunlight on the water surface. Direct sunlight reflections can be extremely bright and cause overexposed areas in aerial images. These reflections overexpose the image, making it difficult to see the seafloor structure, as shown in Figure 1.6.

If the seafloor is not visible in the image, SfM algorithms cannot map these areas effectively. Therefore, it is necessary to avoid such reflections in the images. A solution is to take images when the sun is at a low angle, which reduces the reflection of sunlight on the water surface, as suggested by (Slocum et al., 2019). This can be achieved by surveying the desired area early in the morning or late in the afternoon, or on a cloudy day. In our case, images were taken between daybreak and sunrise, which is the best time to avoid reflections on the water surface on the west part of Réunion Island.

Regarding flight altitude, a lower one implies a higher resolution point cloud in the SfM process since more details of the seabed will be visible but, on the other hand, wave-induced refraction becomes more visible. On the contrary, setting the flight altitude too high will reduce the resolution of the point cloud, making it difficult to distinguish characteristics of study objects (e.g., coral colonies, habitats, etc.).

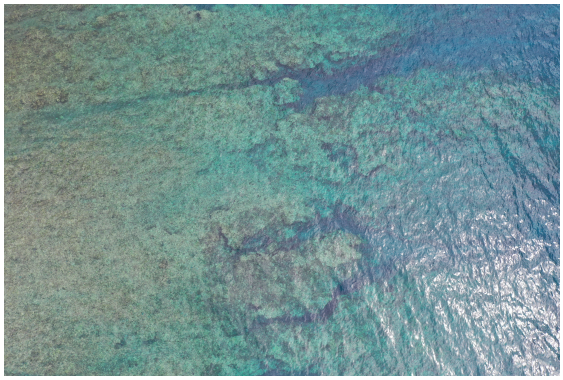


Figure 1.6: *Example of an aerial image with localized overexposure caused by sunlight reflections on the water surface, taken by a Mavic 2 Pro drone in the Saint-Leu lagoon in Réunion Island. The sun reflection on the water surface makes it impossible to see the seafloor structure.*

Photogrammetric processing

Many photogrammetry software are available to build a 3D model from images. Since our goal was to create an open-source pipeline, the software package `OpenDroneMap` was used. `OpenDroneMap` is a commercial-grade open-source software package for SfM photogrammetric processing (initially developed for aerial images) used to generate georeferenced orthophotos, point clouds, elevation models and textured 3D models from aerial images (OpenDroneMap Development Team, 2025). Moreover, its flexibility to run locally or on web servers makes it well-suited for integration into automated post-processing pipelines.

1.3 Artificial Intelligence: a new frontier in marine monitoring

Before explaining the principles behind deep learning's efficiency, we will first review the historical development of machine learning and deep learning. The section concludes with a discussion of the trade-off between model performance and interpretability.

1.3.1 A short history of artificial intelligence

Creating machines capable of autonomous reasoning has fascinated scientists for centuries (Russell and Norvig, 2016). But it was only in the 19th century, with the advent of the logical reasoning formalized by George Boole and the creation of the Turing machine by Alan Turing, that the first steps towards the concept of a modern computer were taken (Benko and Lányi, 2009). Machine Learning (ML) originated alongside the first computers in the 1950s and has evolved in tandem with advances in computational power, as represented in Figure 1.7.

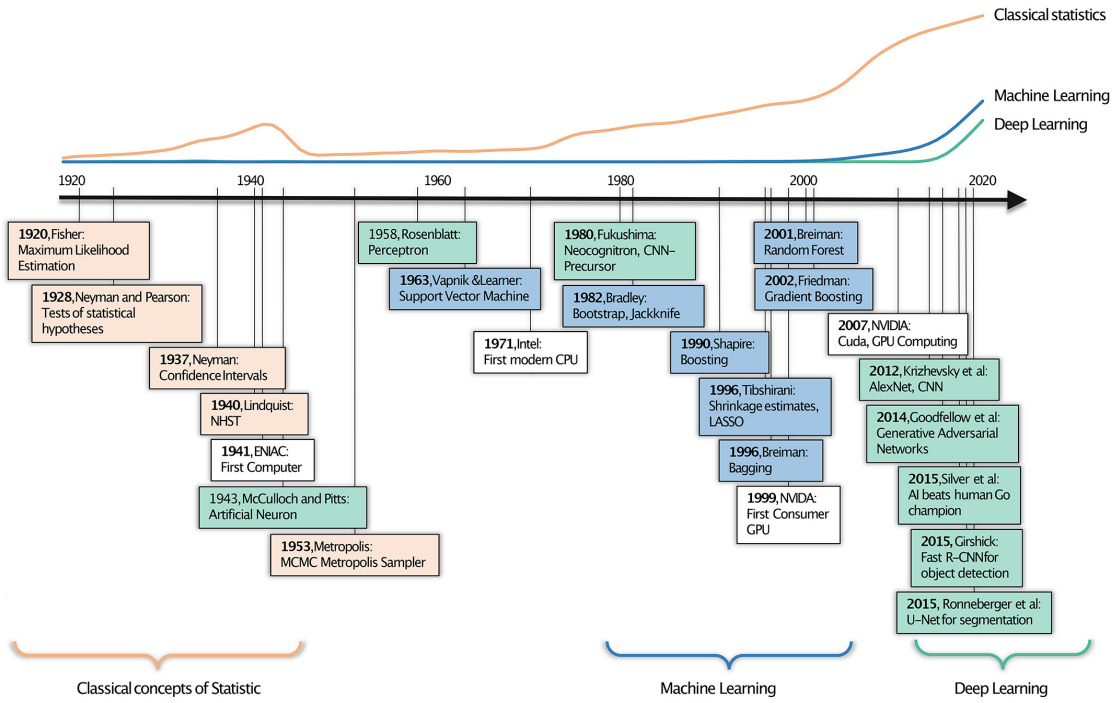


Figure 1.7: *The three eras of statistical learning reflect distinct phases in methodological development. Foundational statistical concepts emerged in 1920. ML saw a surge of innovation from 1980. While the theoretical foundation for Deep Learning (DL) was postulated in the '60s, it has only gained popularity in recent years. The trend lines above the timeline correspond to the frequency of the occurrence of classical statistics (orange), machine learning (blue) and deep learning (green) terms in the scientific literature. Since the publication of Pichler and Hartig (2023), the field has seen the rise of new paradigms such as transformer-based architectures and the proliferation of open platforms like Hugging Face, which have democratized access to State Of The Art (SOTA) models. Figure extracted from Pichler and Hartig (2023).*

The 1990s marked the field's first major breakthrough, characterized by the development of foundational algorithms such as boosting, bagging, shrinkage estimation and random forests, which began to challenge the dominance of traditional probability-based statistical models for data analysis and prediction (Pichler and Hartig, 2023). What these techniques have in common is that they abandon the classical idea of a probabilistic data-generating model in favour of generic algorithmic structures. The aim of these ML models is to minimise a loss function that is not necessarily linked to the probability of the observations (Shmueli, 2010). This can be formally expressed as the minimization of the expected value of a function $L(y_i, \hat{f}(x_i))$ over the data, where y_i is the true value and $\hat{f}(x_i)$ is the model's prediction.

The goal of ML is then to build a predictive model $f : \mathcal{X} \to \mathcal{Y}$, where $\mathcal{X}$ represents the input space and $\mathcal{Y}$ represents the output or target space, that performs well on unseen

data. If a ML model is trained with labelled data $\{(x_i, y_i)\}_{i=1}^{N}$ (a pair of input and output), the task is referred to as supervised learning, where the objective is to learn the relationship between inputs and their corresponding target output. In contrast, if a model is trained on unlabelled data, it falls under unsupervised learning, where the goal is to identify hidden patterns, structures, or groupings within the dataset. A key aspect of traditional ML is the need for feature engineering, meaning that humans must decide which input variables are most relevant and informative for the model to learn from (Olden et al., 2008) [2]. If the model adapts too closely to the training data, capturing noise or random fluctuations instead of general patterns, it will result in poor performance when applied to new, unseen data. This phenomenon is known as overfitting.

A second transformative wave emerged in the past decade with the advent of deep neural networks, enabled by the commercialisation of the first graphics processing units (Graphics Processing Units (GPUs)) that made training large-scale models feasible. DL algorithms are neural networks that differ from classical Artificial Neural Networks (ANNs) primarily in their depth and scale. While traditional ANNs typically consist of just one or two hidden layers and a relatively small number of parameters, DL models are composed of dozens or even hundreds of layers stacked in sequence. By decomposing the data into multiple layers, each with different levels of abstraction, the algorithm is able to automatically learn complex features representing the data (Christin et al., 2019) [3]. While ML algorithms often rely on simpler model structures and perform well on structured datasets with engineered features, DL excels in processing high-dimensional and unstructured data (Jordan and Mitchell, 2015). This DL revolution led to landmark achievements such as mastering the game of chess (Silver et al., 2018), advancing Natural Language Processing (NLP) (Radford et al., 2019), enhancing image recognition and segmentation (Oquab et al., 2023) and revolutionizing protein structure prediction (Senior et al., 2020).

1.3.2 Deep Learning architectures

DL includes a variety of model architectures, each tailored to handle different types of data and representation needs. Among the most widely used are Convolutional Neural Networks (CNNs), Recurrent Neural Networks, Variational Autoencoders, Generative Adversarial Networks and Transformers.

In the following part we will focus specifically on CNNs and Transformers, as these architectures are particularly well-suited for CV tasks, which is why they were selected and applied throughout this thesis.

CNNs are primarily used for tasks involving spatial data, particularly images. They were introduced in the 1990s by Yann LeCun in order to solve two technical problems

[2] For example, in order to detect sea cucumbers in images, the algorithm requires certain properties of the holothurians (e.g., shape, colour, size, background) to be explicitly specified in terms of pixel patterns.

[3] This means that we only need to tell a DL algorithm whether a sea cucumber is present in an image and, given enough examples, it will be able to figure out by itself what a sea cucumber looks like.

that were inherent in fully connected neural networks: the high number of parameters and the lack of translation invariance (LeCun et al., 1998). CNNs leverage convolutional layers to detect local patterns such as edges, textures, or object parts. For example, this allows them to analyse passive acoustic data such as whale songs (Allen et al., 2021), predict species distributions in marine environments (Tong et al., 2023; Morand et al., 2024) or detect fishing vessels on satellite images (Talpaert Daudon et al., 2023; Bai et al., 2024).

Transformers, first introduced by Vaswani et al. (2017), marked a major turning point in the way sequential data are processed. Unlike traditional models such as CNNs, transformers rely on self-attention mechanisms to model contextual relationships across all positions in a sequence. This means that, instead of processing data sequentially like Recurrent Neural Networks (RNNs) or focusing on local patterns like CNNs, transformers can evaluate the importance of each element in the input relative to every other element simultaneously, allowing them to capture global information more effectively (Khan et al., 2022; Chinesta et al., 2025). Initially developed for NLP (Wolf et al., 2020), transformers have become the backbone of many state-of-the-art models and have achieved groundbreaking results in tasks like language modelling (Roy et al., 2021), machine translation (Fan et al., 2021) and summarization (Aswani et al., 2024). Their versatility has also led to successful applications beyond NLP. Attention-based models have shown strong performance in audio applications (Chen et al., 2021b), physical systems modelling (Geneva and Zabaras, 2022) and molecular property prediction (Chen et al., 2021a). More recently, the introduction of Vision Transformer (ViT) has extended the transformer paradigm to image analysis, with models demonstrating competitive or superior accuracy to traditional CNNs in image classification benchmarks (Wang et al., 2025). This shift signals a growing trend in ecology, where transformer-based architectures are redefining how visual features are extracted and understood (Picek et al., 2022; Leblanc et al., 2024; Zheng et al., 2024).

1.3.3 Computer vision tasks

CV is a field closely linked to AI, focused on enabling machines to interpret and understand visual information from the world around them. While early CV relied on rule-based and mathematical approaches, modern advances are largely driven by ML and DL techniques, as shown in Figure 1.8.

With the advent of DL models, the field has experienced rapid progress. In particular, the introduction of CNNs marked a major turning point, leading to substantial improvements in performance on key benchmarks such as the *ImageNet Large Scale Visual Recognition Challenge* (Russakovsky et al., 2015), as demonstrated by Guo et al. (2016); Chai et al. (2021). The subsequent development of transformers further expanded the capabilities of CV models, allowing them to consider long-range dependencies and contextual interactions between distant regions of an image (Khan et al., 2022).

Typical tasks in CV include image classification, object detection, segmentation, object tracking, 3D reconstruction, image super-resolution and image generation. In Figure 1.9, we provide an overview of the two major tasks related to this thesis: image

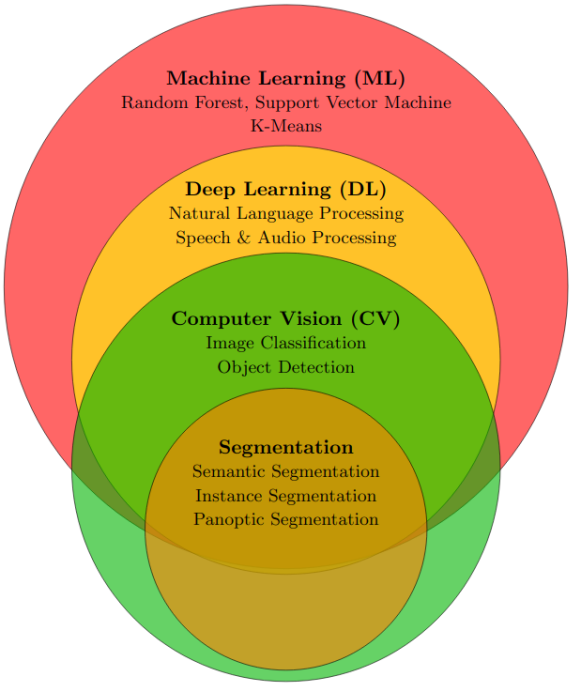


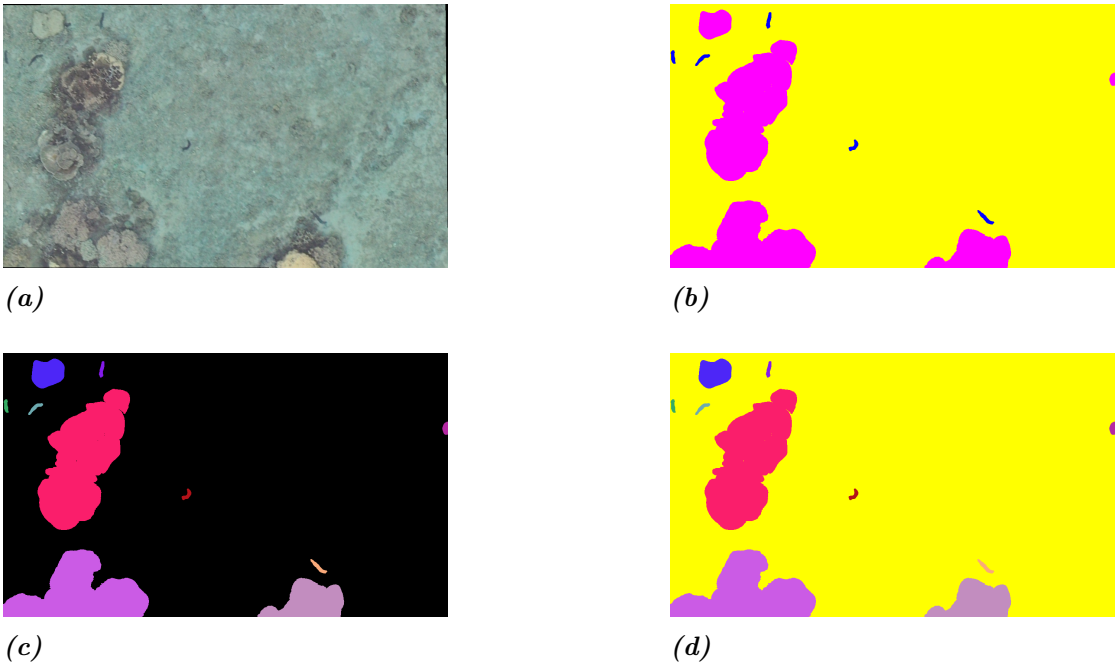
Figure 1.8: *Venn diagram illustrating the relationship between ML, DL and CV. DL is shown as a subset of ML and CV overlaps with both fields. Example tasks are provided to show how each field contributes to real-world AI systems.*

classification and segmentation.

Image Classification is the task of assigning a label to an image based on its content. Depending on the number of labels assigned to each image, image classification can be divided into two categories: mono label classification (where each image is assigned a single label) and multi label classification (where images can be assigned multiple labels). For example Figure 1.9a can be labelled using mono-label classification as an image of a coral reef (or by choosing arbitrarily one of the coral morphotypes present in the image) and in multi-label classification as an image representing *Coral*, *Sea cucumber* and *Sand* classes.

Image Segmentation is the task of partitioning an image into multiple segments or masks (group of pixels) based on the content of the image. This task can be divided into three categories: semantic segmentation, instance segmentation and panoptic segmentation.

- **Semantic segmentation** assigns a class label to each pixel in the image, creating a bi-univocal correspondence between pixels and labels. For example, in Figure 1.9b, an aerial image of a lagoon in Réunion Island is segmented into *Coral*, *Sea cucumber* and *Sand* classes.



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

Figure 1.9: *Illustration of the three main types of image segmentation tasks: semantic segmentation, instance segmentation and panoptic segmentation. (a)Aerial image of the Trou d'eau lagoon in Réunion Island. The area measures* $63\,\mathrm{m}^2$ *and is representative of the diverse coral morphotypes and habitats, objects of this study. (b)Semantic segmentation annotation of the image. Image pixels are divided into three classes: Coral (pink), Sea cucumber (blue) and Sand (yellow). (c)Instance segmentation annotation of the image. The image is divided into individual instances of the same class, with each object assigned a unique identifier/colour. The background (Sand) is not annotated since it is considered as "stuff". (d)Panoptic segmentation annotation of the image. The image is divided into "stuff" and "things", with each object assigned a unique identifier/colour.*

- **Instance segmentation** goes a step further by distinguishing between individual instances of the same class, assigning a unique identifier to each object. However, uncountable classes (e.g. *Sand*) are not annotated. This is illustrated in Figure 1.9c, where each *Coral* colony and each , *Sea cucumber* individual are assigned a unique colour and the background (*Sand*) is not annotated.

- **Panoptic segmentation** combines the two previous tasks, providing a unified framework for segmenting both "stuff" and "things" in an image. In Figure 1.9d, the image is divided into "stuff" (*Sand*) and "things" (instances of *Coral* colonies and *Sea cucumber* individuals), with each object assigned a unique colour.

In this framework, "stuff" refers to amorphous, uncountable regions that lack clear boundaries like *Sand*, *Rubble* or *Seagrass*, which are typically represented only in semantic or panoptic segmentation tasks. In contrast, "things" are discrete, countable objects

with well-defined boundaries like *Coral* colonies or *Sea cucumber*. Individuals of classes that falls under the category of things can be distinguished from each other, in instance or panoptic segmentation tasks. In this thesis we will focus on two major tasks: multi label image classification and semantic segmentation. Multi-label classification was selected for underwater imagery due to the relative simplicity of generating image-level annotations. Semantic segmentation, on the other hand, was included because recent techniques (discussed in Section 1.3.4) make it possible to generate dense annotations from simpler forms of supervision, reducing the need for fully manual pixel-level labelling.

1.3.4 Trends in deep learning and computer vision

Over the past decade, CV has undergone a major and ongoing transformation, evolving beyond traditional supervised learning toward more scalable, flexible and data-efficient paradigms (Zhou et al., 2024). Central to this shift are three key techniques that have enabled models to generalize better and adapt more efficiently to new tasks: transfer learning, fine-tuning and knowledge distillation.

Transfer learning refers to the practice of leveraging knowledge learned from one task or dataset to improve performance on another task. In CV, this usually involves initializing a model with pretrained weights from a large-scale dataset and adapting it to a more specific target dataset. This approach is especially useful in fields like marine ecology, where labelled data are limited or expensive to obtain.

Fine-tuning builds on transfer learning by updating part or all of the model's parameters using the new dataset. Depending on the task, the volume of labelled data and the computational budget, researchers may fine-tune only the final layers (e.g., classification heads) or the entire network. This process allows models to adapt to domain-specific features, such as those found in underwater imagery.

Knowledge distillation is another powerful method, in which a student model learns to replicate the outputs of a more complex teacher model. Traditionally used to compress large models into smaller and more efficient ones, distillation can also be applied across domains. In this work, for example, we will use distillation to transfer ecological knowledge from high-resolution underwater imagery (teacher) to lower-resolution aerial imagery (student), rather than for compression. This enables broader reef monitoring without requiring additional annotation (see Chapter 4 and Chapter 5).

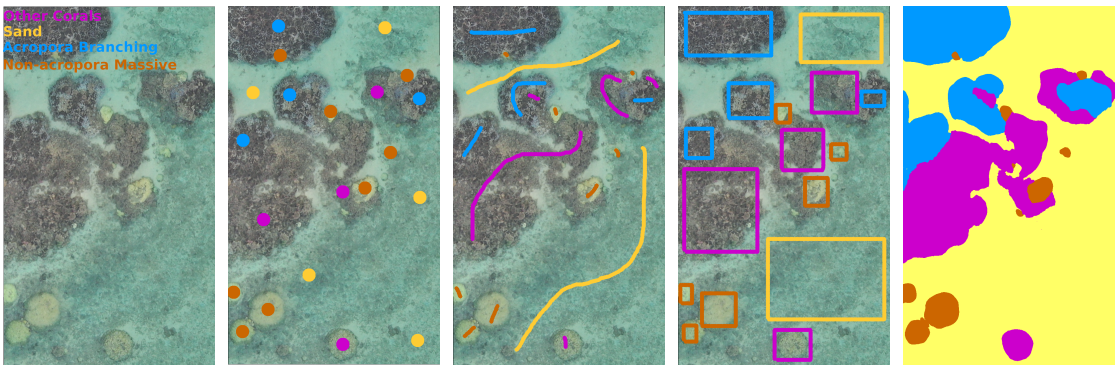
These techniques have become increasingly effective thanks to the emergence of foundational models: large-scale, pretrained models capable of generalizing across a wide range of tasks with minimal supervision (Bommasani et al., 2021). Originally developed for NLP, Foundation Models are now redefining visual learning. Trained on massive and diverse datasets using self-supervised methods, vision foundation models like DinoV2 (Oquab et al., 2023) and CLIP (Radford et al., 2021) provide rich visual representations that can be transferred or fine-tuned for various applications, including ecological monitoring. Their broad generalization ability makes them particularly useful when labelled data is scarce, as they significantly reduce the need for task-specific training. One of the most significant recent developments in this field is Segment Anything Model (SAM): a vision foundation model designed for general-purpose image segmentation (Kirillov

et al., 2023). SAM supports zero-shot segmentation, meaning it can generate masks in new images without task-specific training. Moreover, by accepting simple prompts (such as points, boxes or free-form masks) it can segment objects with a human in the loop. This enables the creation of efficient semi-automated annotation pipelines, reducing both the time and cost required to build labelled datasets.

Complementing these advances, Weakly Supervised Semantic Segmentation (WSSS) offers another promising strategy for minimizing annotation effort (Huang et al., 2023; Wang et al., 2023). In WSSS, models are trained using incomplete, inexact or less expensive forms of annotation rather than requiring pixel-perfect ground truth masks. The objective is to learn to segment objects in images using weaker forms of supervision thus achieving a gain in the annotation process, while still producing high-quality segmentation masks. Several types of weak supervision are commonly used (Zhao et al., 2024). To illustrate these different strategies, we present Figure 1.10, which shows examples of:

- image-level labels, where only the presence of classes is indicated for the entire image, but no information about their boundaries is provided (see Figure 1.10a)
- point-level supervision, where only a few pixels are labelled (see Figure 1.10b)
- scribble annotations, which provide sparse, user-drawn lines (see Figure 1.10c)
- bounding boxes, which outline the objects of interest with a rectangle (see Figure 1.10d)

All these weak forms of supervision can be used to train a segmentation model that learns to predict pixel-level masks (see Figure 1.10e).



(a) *Image-level annotation* ***(b)*** *Point annotation* ***(c)*** *Scribble annotation* ***(d)*** *Bounding box annotation* ***(e)*** *Pixel-level mask*

Figure 1.10: *Illustration of different types of annotation in weakly supervised semantic segmentation.*

When combined with models like SAM or used as pre-training strategies to generate pseudo-labels, WSSS can be coupled with self-knowledge distillation techniques to im-

prove the performance of segmentation models in a training loop where the model learns from its own predictions (Xu et al., 2023).

1.3.5 The cost of understanding: interpretability in high-performance models

The rapid rise and widespread adoption of DL methods can be attributed to their ability to automatically extract relevant features from complex, high-dimensional data. In ecology, research conducted across various levels (from individuals to entire ecosystems) often generates rich, high-dimensional datasets, making DL particularly well-suited for analysis in this field. However, this advantage comes with a trade-off. While DL models often outperform traditional methods in terms of performance, their ability to automatically learn features directly from the data makes them inherently less interpretable as shown in Figure 1.11. Indeed, the strong empirical performance of DL models is largely attributed to the combination of powerful learning algorithms and their vast parametric space. This space, often involving hundreds of layers and millions of parameters, contributes to the perception of these neural networks as highly complex black-box models (Castelvecchi, 2016).

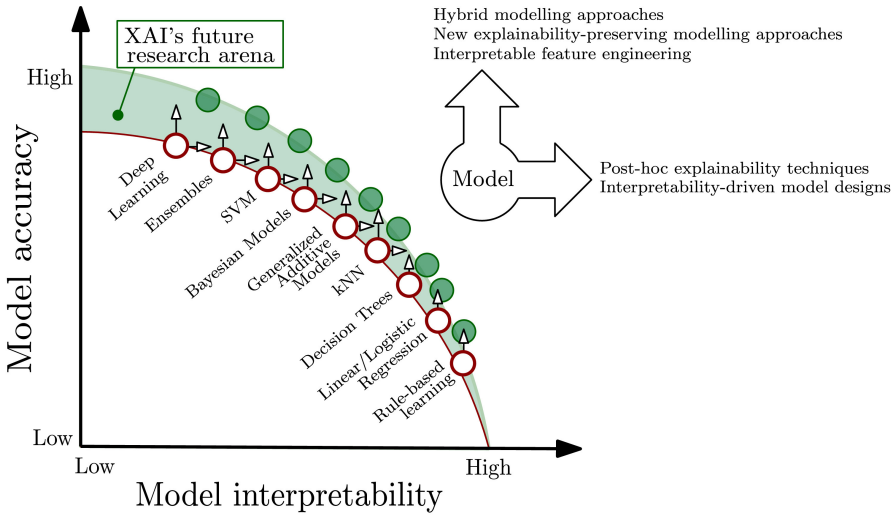


Figure 1.11: *The trade-off between model interpretability and performance highlights a key challenge in deep learning and it is within this gap (green highlighted section) that the potential of Explainable Artificial Intelligence (XAI) techniques and tools lies. These methods aim to improve the transparency of complex models without compromising their predictive power. Figure extracted from Arrieta et al. (2020).*

As DL models become more common in everyday applications, there is a growing demand for transparency from stakeholders concerned with how these systems make decisions (Preece et al., 2018). These models offer limited insight into how decisions are made, especially when compared to more interpretable classical ML approaches (Arrieta

et al., 2020). This lack of interpretability can slow down or even block the adoption of DL in ecological research, where transparency and explainability are essential for understanding complex ecological processes and informing conservation strategies (Ryo et al., 2021).

In order to address these challenges, the field of XAI has emerged as a promising solution (Longo et al., 2024).

XAI comprises a range of techniques and tools that are designed to provide greater understanding and explanations of the functioning of complex models. XAI has seen rapid growth in recent years, evolving from a niche area within the broader field of AI (Swartout et al., 1991) into an active developing research domain. What was once a specialized topic has now attracted widespread attention, with a growing body of theoretical advancements, empirical studies and reviews being published each year (Schwalbe and Finzel, 2024). The main goal of these methods is to explain how inputs are turned into outputs, identify the most influential features in a prediction and detect any possible biases or uncertainties in how the model works. In ecology, XAI helps researchers not only check and trust the model's results but also uncover previously unknown patterns and ecological relationships (He et al., 2022). By improving the balance between model accuracy and interpretability, XAI supports the responsible and practical use of DL in ecological research and decision-making.

During this thesis, this field was not explored, even though XAI could provide valuable insights into which parts of an image (whether in multilabel classification or semantic segmentation tasks) are most informative for the model's predictions, thereby helping to guide and optimize future data collection efforts.

1.4 AI in marine vision

The integration of CV in post-processing of marine imagery has been a game-changer for ecological monitoring (Salman, 2023), enabling the analysis of large datasets with unprecedented speed and accuracy. As technologies such as ASVs, UAVs, satellite imagery or citizen science projects enable the collection of large volumes of data, it becomes unfeasible to annotate and analyse them manually. AI offers a solution by processing and interpreting these data with unmatched speed, precision and reproducibility.

Several recent studies illustrate this potential across multiple data scales. Villon et al. (2018) demonstrated that CNNs can outperform human experts in identifying reef fish species from underwater imagery. Sauder et al. (2024) introduced a scalable framework that combines semantic segmentation with SfM photogrammetry to generate accurate 3D semantic maps of coral reefs. Such methods enable the creation of ecologically meaningful habitat maps directly from underwater footage, with applications for conservation planning and biodiversity monitoring.

AI also plays a key role in processing aerial and satellite imagery. Øie Kvile et al. (2024) and Doukari and Topouzelis (2022) used aerial drone images and ML to classify broad habitat types such as sand, coral and seagrass, demonstrating how remote sensing platforms can support shallow waters reef mapping. At larger scales, Franceschini et al.

(2023) employed high-resolution multispectral satellite imagery to detect coral reef halos (circular sand formations around coral reef patches) while Li and Asner (2023) mapped benthic complexity using Sentinel-2 data across diverse coral reef environments.

However, a major bottleneck remains the annotation process, which is time-consuming and often requires expert knowledge. In our case, annotating 14,492 underwater images using the multilabel classification convention described in Section 1.3.3 required over 1,200 hours of work and the involvement of five marine ecologists. This corresponds to about 5 minutes of expert time per image. In contrast, segmentation tasks (which require labelling each pixel) are significantly more demanding. Based on our experience with an expert benthologist, annotating a single underwater image for segmentation typically takes around 19 minutes, with a range of 6 to 160 objects per image and an average of about 55. This high time cost underscores the need for more efficient annotation workflows.

To address this, Raine et al. (2024) introduced a human-in-the-loop segmentation framework that combines expert point annotations with semi-automated tools to accelerate the generation of segmentation masks. Similarly, the use of WSSS techniques has been shown to reduce annotation costs while still producing high-quality segmentation masks (Sun et al., 2023). Such approaches are promising steps towards reducing the time and effort spent on annotations and facilitating the development of ecological monitoring models.

In this thesis, we build on these ideas by applying them to coral reef imagery collected from autonomous platforms. To our knowledge, WSSS approaches have not yet been applied in multi-scale contexts, where cross-modal knowledge transfer (e.g., from ASV-based underwater classification to UAV-based aerial segmentation) can be used to improve the quality and scalability of segmentation models. We demonstrate how predictions from underwater image classification can be used to train segmentation models on aerial imagery, enabling scalable reef mapping without the need for manually annotated aerial ground truth. This cross-domain supervision framework bridges the gap between the underwater and aerial domains, creating new possibilities for large-scale habitat mapping using different data sources.

1.5 Aims of the thesis and thesis structure

Mapping marine biodiversity is a multifaceted challenge requiring scalable and cost-effective data collection strategies, along with advanced analytical methods to process huge volumes of data.

This PhD research is positioned at the intersection of marine ecology, AI and remote sensing with the goal of developing a multi-scale mapping framework for tropical reef ecosystems in shallow waters of the Indian Ocean. It aims to explore how modern data collection platforms can be combined with new AI techniques to support scalable biodiversity monitoring from underwater environments to airborne and satellite levels. Contrary to existing approaches that rely either on expert-driven annotation or single-platform data, this thesis introduces a cross-platform framework that scales from underwater imagery to aerial data as shown in Figure 1.5. The originality of our approach lies in combining weak supervision, citizen science imagery and georeferenced acquisition pipelines. The goal is to evaluate how these tools, individually and in combination, can contribute to more systematic, spatially extensive and temporally frequent ecological assessments for coral reef ecosystems. It is worth noting that the developed framework is model agnostic, meaning that it can integrate evolving AI models without requiring changes to the overall pipeline. It provides a simple, scalable tool for generating habitat maps by applying these models to georeferenced imagery. By embedding AI directly into the spatial process, it helps translate ecological predictions into practical and shareable maps, useful for both scientists and local decision-makers.

This manuscript follows a logical progression from localized, well-annotated data (underwater imagery) to broader-scale monitoring using low-altitude and high-altitude aerial imagery. Each chapter is based on one or more scientific publications (published or under review); however, to improve the overall readability of the thesis, update key information and expand certain sections, all chapters have been revised and harmonized. Every chapter begins with a brief introductory page outlining its content, followed by a dedicated *Context and objectives* section to guide the reader. To maintain a coherent narrative, each chapter concludes with a *Conclusion and continuity* section, which both summarizes the main findings and establishes a clear link with the subsequent chapters, helping the reader follow the "fil rouge" of the thesis.

The chapters are structured to reflect this multi-scale and methodological evolution:

- Chapter 2 introduces *Seatizen Atlas*, a large-scale, collaborative dataset comprising over 1.6 million geolocated underwater and aerial images collected across the Southwest Indian Ocean as well as 14,492 annotated images that could be used for AI training. This contribution goes beyond existing datasets by offering a FAIR-structured, multi-source collection that supports both AI training and ecological monitoring across spatial scales. It describes diverse acquisition strategies employed throughout the thesis, including simulated citizen science methods and scientific platforms such as ASVs and UAVs. Each data collection event and corresponding metadata are openly accessible, has been assigned a specific Digital Object Identifier (DOI) through Zenodo data repository and is also referenced

within a global meta-dataset for simplified exploration. The data was structured according to FAIR principles to encourage contributions from both scientists and citizens, promoting the development of a broader, collaborative resource.

Chapter 3 presents the methodological framework developed for the classification of underwater imagery, which was evaluated during the 2022 Indian Ocean Expedition. The data collected with low-cost acquisition platforms like ASV are processed through a workflow that applies geolocation and metadata synchronisation techniques, enabling structured and scalable marine data collection. A set of 14,492 images was manually annotated using the multilabel image classification technique and used to fine-tune the vision transformer model DinoV2. The resulting model, named *DinoVdeau*, predicts the presence of multiple benthic classes (coral morphotypes, seagrass and substrates) within each image. Predictions were validated using independent reference data acquired at Aldabra Atoll by the Seychelles Islands Foundation in 2022 and by satellite surveys conducted in 2009, demonstrating the model's potential for long-term and reproducible monitoring. This chapter highlights how the integration of affordable field equipment with AI-based analysis can support cost-effective conservation strategies even in remote islands.

Chapter 4 presents a novel multi-scale DL framework for coral reef monitoring that combines underwater imagery with broader-scale aerial orthophotos. We address the challenge of mapping coral habitats over large areas without relying on manually annotated aerial images by introducing a novel cross-domain knowledge distillation approach, where fine-grained ecological predictions from an underwater model are transferred to a second model operating on aerial drone tiles. By aligning imagery across scales and weighting *DinoVdeau* predictions based on the spatial overlap between underwater and aerial images, this method enables accurate large-area habitat classification without the need for aerial annotations. The chapter discusses the full processing pipeline (from data acquisition and processing to model training and evaluation) demonstrating how the integration of multi-source data can support scalable, cost-efficient and high-resolution coral reef assessments.

Chapter 5 extends the previous chapter by converting underwater pointwise predictions into dense segmentation labels for aerial imagery. This work advances the field by demonstrating that semantic segmentation models can be trained on aerial data without any manually annotated masks, relying instead on geolocated predictions from underwater models as weak supervision. This is achieved by spatially interpolating the underwater predictions to generate weak supervision masks that are then aligned with aerial orthophotos. These pseudo-labels are used to fine-tune the full semantic segmentation model *SegFormer* trained to recognize five benthic classes, including coral morphotypes and habitat types. A self-distillation step is then applied to refine the predictions and improve segmentation accuracy.

The results demonstrate that the model can recover ecologically meaningful spatial patterns, confirming the feasibility of scaling coral reef monitoring to larger areas with minimal manual annotation, while also upgrading prediction quality by transitioning from multilabel classification to the more complex task of semantic segmentation. Finally, the chapter discusses how the pipeline can be incrementally improved through the integration of new aerial annotations, illustrated by a case study on the detection of *Sea cucumber*.

Chapter 6 develops from upon the previous UAV-based segmentation work by exploring the upscaling of ecological annotations to national high-altitude aerial imagery from the IGN *BD Ortho* program. This chapter investigates whether high-resolution predictions from UAV surveys can be transferred to coarser-resolution aerial images to enable large-scale coral reef mapping. A semantic segmentation model is trained using weak supervision from UAV-derived predictions aligned with IGN tiles acquired in 2017 and 2022. Despite the presence of temporal noise in the training data, the model achieves strong performance and successfully distinguishes between key benthic classes, including coral morphotypes and historical seagrass meadows. The method is further validated on independent test zones using both UAV-derived and human-annotated references. This chapter demonstrates the feasibility of bridging local UAV observations with broader aerial imagery and highlights the potential for integrating very high-resolution satellite data for regional-scale reef monitoring.

This sequence of contributions builds toward a scalable, low-cost and multi-platform framework for coral reef monitoring, emphasizing both methodological innovation and ecological applicability.

Bibliography

Ackleson, S. G., Smith, J. P., Rodriguez, L. M., Moses, W. J., and Russell, B. J. (2017). Autonomous coral reef survey in support of remote sensing. *Frontiers in Marine Science*, 4:325.

Allen, A. N., Harvey, M., Harrell, L., Jansen, A., Merkens, K. P., Wall, C. C., Cattiau, J., and Oleson, E. M. (2021). A convolutional neural network for automated detection of humpback whale song in a diverse, long-term passive acoustic dataset. *Frontiers in Marine Science*, 8:607321.

Amone-Mabuto, M., Bandeira, S., Hollander, J., Hume, D., Campira, J., and Adams, J. (2024). Monitoring seagrass meadows in maputo bay using integrated remote sensing techniques and machine learning. *Regional Studies in Marine Science*, 79:103816.

Anton, V., Germishuys, J., Bergström, P., Lindegarth, M., and Obst, M. (2021). An open-source, citizen science and machine learning approach to analyse subsea movies. *Biodiversity data journal*, 9:e60548.

Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., García, S., Gil-López, S., Molina, D., Benjamins, R., et al. (2020). Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. *Information fusion*, 58:82–115.

Asner, G. P., Vaughn, N. R., Heckler, J., Knapp, D. E., Balzotti, C., Shafron, E., Martin, R. E., Neilson, B. J., and Gove, J. M. (2020). Large-scale mapping of live corals to guide reef conservation. *Proceedings of the National Academy of Sciences*, 117(52):33711–33718.

Aswani, S., Choudhary, K., Shetty, S., and Nur, N. (2024). Automatic text summarization of scientific articles using transformers—a brief review. *Journal of Autonomous Intelligence*, 7(5).

B. Lyons, M., M. Roelfsema, C., V. Kennedy, E., M. Kovacs, E., Borrego-Acevedo, R., Markey, K., Roe, M., M. Yuwono, D., L. Harris, D., R. Phinn, S., et al. (2020). Mapping the world's coral reefs using a global multiscale earth observation framework. *Remote Sensing in Ecology and Conservation*, 6(4):557–568.

Bai, Y., Li, S., Ju, R.-Y., Yang, Z., Yu, J., and Chiang, J.-S. (2024). Fad-sar: A novel fishing activity detection system via synthetic aperture radar images based on deep learning method. In *2024 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, pages 1–5. IEEE.

Basyuni, M., Mubaraq, A., Amelia, R., Wirasatriya, A., Iryanthony, S. B., Slamet, B., Al Mustaniroh, S. S., Pradisty, N. A., Sidik, F., Hanintyo, R., et al. (2025). Mangrove aboveground biomass estimation using uav imagery and a constructed height model in budeng–perancak, bali, indonesia. *Ecological Informatics*, page 103037.

Benko, A. and Lányi, C. S. (2009). History of artificial intelligence. In *Encyclopedia of Information Science and Technology, Second Edition*, pages 1759–1762. IGI global.

Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., Bernstein, M. S., Bohg, J., Bosselut, A., Brunskill, E., et al. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.

Bonnet, P., Joly, A., Faton, J.-M., Brown, S., Kimiti, D., Deneu, B., Servajean, M., Affouard, A., Lombardo, J.-C., Mary, L., et al. (2020). How citizen scientists contribute to monitor protected areas thanks to automatic plant identification tools. *Ecological Solutions and Evidence*, 1(2):e12023.

Brignone, L., Opderbecke, J., Bouhier, M. E., Marfia, C., Piasco, R., Gobillot, N., Magnier, C., and Arnaubec, A. (2023). Validation at sea for uly x, the latest deep diving auv for oceanographic research. In *OCEANS 2023-Limerick*, pages 1–6. IEEE.

Broudic, L., Pinault, M., Claud, R., Bajjouk, T., Rungassamy, T., Nikolic, N., Crochelet, E., Mazé, C., and Bergerot, B. (2025). Mapping benthic biodiversity indicators of coral reefs using spatial interpolation. *Coral Reefs*.

Buddemeier, R. W. (2004). *Coral reefs & global climate change: potential contributions of climate change to stresses on coral reef ecosystems*, volume 10. Pew Center on Global Climate Change.

Burke, L., Reytar, K., Spalding, M., and Perry, A. (2011). *Reefs at risk revisited*. Washington, DC: World Resources Institute (WRI).

Caldwell, Z. R., Zgliczynski, B. J., Williams, G. J., and Sandin, S. A. (2016). Reef fish survey techniques: assessing the potential for standardizing methodologies. *PloS one*, 11(4):e0153066.

Cappo, M., Harvey, E., and Shortis, M. (2006). Counting and measuring fish with baited video techniques-an overview. In *Australian Society for Fish Biology Workshop Proceedings*, volume 1, pages 101–114. Australian Society for fish biology Tasmania.

Cardenas, J. A., Samadikhoshkho, Z., Rehman, A. U., Valle-Pérez, A. U., de León, E. H.-P., Hauser, C. A., Feron, E. M., and Ahmad, R. (2024). A systematic review of robotic efficacy in coral reef monitoring techniques. *Marine Pollution Bulletin*, 202:116273.

Castelvecchi, D. (2016). Can we open the black box of ai? *Nature News*, 538(7623):20.

Chai, J., Zeng, H., Li, A., and Ngai, E. W. (2021). Deep learning in computer vision: A critical review of emerging techniques and application scenarios. *Machine Learning with Applications*, 6:100134.

Chaloux, N., Phillips, B. T., Gruber, D. F., Schelly, R. C., and Sparks, J. S. (2021). A novel fish sampling system for rovs. *Deep Sea Research Part I: Oceanographic Research Papers*, 167:103428.

Chapin, F. S., Walker, B. H., Hobbs, R. J., Hooper, D. U., Lawton, J. H., Sala, O. E., and Tilman, D. (1997). Biotic control over the functioning of ecosystems. *Science*, 277(5325):500–504.

Chen, D., Gao, K., Nguyen, D. D., Chen, X., Jiang, Y., Wei, G.-W., and Pan, F. (2021a). Algebraic graph-assisted bidirectional transformers for molecular property prediction. *Nature communications*, 12(1):3521.

Chen, X., Wu, Y., Wang, Z., Liu, S., and Li, J. (2021b). Developing real-time streaming transformer transducer for speech recognition on large-scale dataset. In *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 5904–5908. IEEE.

Chevrier, T., Cowart, D. A., Nieblas, A.-E., Charrier, G., Bernard, S., Evano, H., Brisset, B., Chanut, J., and Bonhommeau, S. (2025). Population structure of the swordfish, xiphias gladius, across the indian ocean using next-generation sequencing. *ICES Journal of Marine Science*, 82(5):fsae179.

Chinesta, F., Cueto, E., Champaney, V., Ghnatios, C., Ammar, A., Hascoët, N., González, D., Alfaro, I., Di Lorenzo, D., Pasquale, A., et al. (2025). *A Gentle Introduction to Data, Learning, and Model Order Reduction: Techniques and Twinning Methodologies*. Springer Nature.

Christin, S., Hervet, É., and Lecomte, N. (2019). Applications for deep learning in ecology. *Methods in Ecology and Evolution*, 10(10):1632–1644.

Cinner, J. E., Maire, E., Huchery, C., MacNeil, M. A., Graham, N. A., Mora, C., McClanahan, T. R., Barnes, M. L., Kittinger, J. N., Hicks, C. C., et al. (2018). Gravity of human impacts mediates coral reef conservation gains. *Proceedings of the National Academy of Sciences*, 115(27):E6116–E6125.

Colloca, F., Scarcella, G., and Libralato, S. (2017). Recent trends and impacts of fisheries exploitation on mediterranean stocks and ecosystems. *Frontiers in Marine Science*, 4:244.

Contini, M., Illien, V., Bonhommeau, S., Barde, J., and Joly, A. (2025). Underwater images collected by scuba diving in ifaty, madagascar - 2023-05-07.

Costanza, R., De Groot, R., Sutton, P., Van der Ploeg, S., Anderson, S. J., Kubiszewski, I., Farber, S., and Turner, R. K. (2014). Changes in the global value of ecosystem services. *Global environmental change*, 26:152–158.

Davis, L. S., Zhu, L., and Finkler, W. (2023). Citizen science: Is it good science? *Sustainability*, 15(5):4577.

de Camargo Barbosa, K. V., Develey, P. F., Ribeiro, M. C., and Jahn, A. E. (2021). The contribution of citizen science to research on migratory and urban birds in brazil. *Ornithology Research*, 29:1–11.

Dickens, L. C., Goatley, C. H., Tanner, J. K., and Bellwood, D. R. (2011). Quantifying relative diver effects in underwater visual censuses. *PloS one*, 6(4):e18965.

Dobson, T., Lenchine, V., and Bainbridge, S. (2024). A review on the interactions between engineering and marine life: key information for engineering professionals. *Journal of Ocean Engineering and Marine Energy*, 10(2):449–459.

Doukari, M. and Topouzelis, K. (2022). Overcoming the uas limitations in the coastal environment for accurate habitat mapping. *Remote Sensing Applications: Society and Environment*, 26:100726.

Drap, P. (2012). Underwater photogrammetry for archaeology. *Special applications of photogrammetry*, 114.

Duarte, C. M., Agusti, S., Barbier, E., Britten, G. L., Castilla, J. C., Gattuso, J.-P., Fulweiler, R. W., Hughes, T. P., Knowlton, N., Lovelock, C. E., et al. (2020). Rebuilding marine life. *Nature*, 580(7801):39–51.

Duarte, C. M., Holmer, M., Olsen, Y., Soto, D., Marbà, N., Guiu, J., Black, K., and Karakassis, I. (2009). Will the oceans help feed humanity? *BioScience*, 59(11):967–976.

Edgar, G. J., Cooper, A., Baker, S. C., Barker, W., Barrett, N. S., Becerro, M. A., Bates, A. E., Brock, D., Ceccarelli, D. M., Clausius, E., Davey, M., Davis, T. R., Day, P. B., Green, A., Griffiths, S. R., Hicks, J., Hinojosa, I. A., Jones, B. K., Kininmonth, S., Larkin, M. F., Lazzari, N., Lefcheck, J. S., Ling, S. D., Mooney, P., Oh, E., Pérez-Matus, A., Pocklington, J. B., Riera, R., Sanabria-Fernandez, J. A., Seroussi, Y., Shaw, I., Shields, D., Shields, J., Smith, M., Soler, G. A., Stuart-Smith, J., Turnbull, J., and Stuart-Smith, R. D. (2020). Establishing the ecological basis for conservation of shallow marine life using reef life survey. *Biological Conservation*, 252:108855.

Elise, S., Bailly, A., Urbina-Barreto, I., Mou-Tham, G., Chiroleu, F., Vigliola, L., Robbins, W. D., and Bruggemann, J. H. (2019a). An optimised passive acoustic sampling scheme to discriminate among coral reefs' ecological states. *Ecological Indicators*, 107:105627.

Elise, S., Urbina-Barreto, I., Pinel, R., Mahamadaly, V., Bureau, S., Penin, L., Adjeroud, M., Kulbicki, M., and Bruggemann, J. H. (2019b). Assessing key ecosystem functions through soundscapes: A new perspective from coral reefs. *Ecological Indicators*, 107:105623.

Emslie, M. J., Bray, P., Cheal, A. J., Johns, K. A., Osborne, K., Sinclair-Taylor, T., and Thompson, C. A. (2020). Decades of monitoring have informed the stewardship and ecological understanding of australia's great barrier reef. *Biological conservation*, 252:108854.

Espitalier, V., Lombardo, J.-C., Goëau, H., Botella, C., Høye, T. T., Dyrmann, M., Bonnet, P., and Joly, A. (2025). Adapting a global plant identification model to detect invasive alien plant species in high-resolution road side images. *Ecological Informatics*, page 103129.

Fan, Z., Gong, Y., Liu, D., Wei, Z., Wang, S., Jiao, J., Duan, N., Zhang, R., and Huang, X.-J. (2021). Mask attention networks: Rethinking and strengthen transformer. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 1692–1701.

Ferrari, R., Figueira, W. F., Pratchett, M. S., Boube, T., Adam, A., Kobelkowsky-Vidrio, T., Doo, S. S., Atwood, T. B., and Byrne, M. (2017). 3d photogrammetry quantifies growth and external erosion of individual coral colonies and skeletons. *Scientific reports*, 7(1):16737.

Ferrario, F., Beck, M., Storlazzi, C., Micheli, F., Shepard, C., and Airoldi, L. (2014). The effectiveness of coral reefs for coastal hazard risk reduction and adaptation. nat commun.

Figuerola, B., Hancock, A. M., Bax, N., Cummings, V. J., Downey, R., Griffiths, H. J., Smith, J., and Stark, J. S. (2021). A review and meta-analysis of potential impacts of ocean acidification on marine calcifiers from the southern ocean. *Frontiers in Marine Science*, 8:584445.

Fraisl, D., Hager, G., Bedessem, B., Gold, M., Hsing, P.-Y., Danielsen, F., Hitchcock, C. B., Hulbert, J. M., Piera, J., Spiers, H., et al. (2022). Citizen science in environmental and ecological sciences. *Nature reviews methods primers*, 2(1):64.

Franceschini, S., Meier, A. C., Suan, A., Stokes, K., Roy, S., and Madin, E. M. (2023). A deep learning model for measuring coral reef halos globally from multispectral satellite imagery. *Remote Sensing of Environment*, 292:113584.

Geneva, N. and Zabaras, N. (2022). Transformers for modeling physical systems. *Neural Networks*, 146:272–289.

Gogendeau, P., Bonhommeau, S., Fourati, H., Julien, M., Contini, M., Chevrier, T., Nieblas, A. E., and Bernard, S. (2025). An autonomous surface vehicle for acoustic tracking, bathymetric and photogrammetric surveys. *Ocean Engineering*, 331:121201.

Graham, N. A., Chabanet, P., Evans, R. D., Jennings, S., Letourneur, Y., Aaron MacNeil, M., McClanahan, T. R., Öhman, M. C., Polunin, N. V., and Wilson, S. K. (2011). Extinction vulnerability of coral reef fishes. *Ecology letters*, 14(4):341–348.

Gréboval, D., Officer, S. F. P., et al. (2002). Report and documentation of the international workshop on factors of unsustainability and overexploitation in fisheries (bangkok, thailand, 4-8 february 2002). *FAO Fisheries Report*.

Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., and Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, 187:27–48.

Harvey, E., Shortis, M., Stadler, M., and Cappo, M. (2002). A comparison of the accuracy and precision of measurements from single and stereo-video systems. *Marine Technology Society Journal*, 36(2):38–49.

He, B., Zhao, Y., and Mao, W. (2022). Explainable artificial intelligence reveals environmental constraints in seagrass distribution. *Ecological Indicators*, 144:109523.

Hedley, J. D., Roelfsema, C. M., Chollett, I., Harborne, A. R., Heron, S. F., J. Weeks, S., Skirving, W. J., Strong, A. E., Eakin, C. M., Christensen, T. R., et al. (2016). Remote sensing of coral reefs for monitoring and management: a review. *Remote Sensing*, 8(2):118.

Hoegh-Guldberg, O. and Bruno, J. F. (2010). The impact of climate change on the world's marine ecosystems. *Science*, 328(5985):1523–1528.

Hoegh-Guldberg, O., Mumby, P. J., Hooten, A. J., Steneck, R. S., Greenfield, P., Gomez, E., Harvell, C. D., Sale, P. F., Edwards, A. J., Caldeira, K., et al. (2007). Coral reefs under rapid climate change and ocean acidification. *science*, 318(5857):1737–1742.

Huang, L., Jiang, B., Lv, S., Liu, Y., and Fu, Y. (2023). Deep-learning-based semantic segmentation of remote sensing images: A survey. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:8370–8396.

Inaba, K. and Hall-Spencer, J. (2020). *Japanese Marine Life A Practical Training Guide in Marine Biology*. Springer.

Johnson-Roberson, M., Pizarro, O., Williams, S. B., and Mahon, I. (2010). Generation and visualization of large-scale three-dimensional reconstructions from underwater robotic surveys. *Journal of Field Robotics*, 27(1):21–51.

Jordan, M. I. and Mitchell, T. M. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349(6245):255–260.

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., and Shah, M. (2022). Transformers in vision: A survey. *ACM computing surveys (CSUR)*, 54(10s):1–41.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., et al. (2023). Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026.

Kullenberg, C. and Kasperowski, D. (2016). What is citizen science?–a scientometric meta-analysis. *PloS one*, 11(1):e0147152.

Lamperti, L., François, O., Mouillot, D., Mathon, L., Sanchez, T., Albouy, C., Pellissier, L., and Manel, S. (2024). A spatial matrix factorization method to characterize ecological assemblages as a mixture of unobserved sources: An application to fish edna surveys. *Methods in Ecology and Evolution*.

Lamperti, L., Sanchez, T., Si Moussi, S., Mouillot, D., Albouy, C., Flück, B., Bruno, M., Valentini, A., Pellissier, L., and Manel, S. (2023). New deep learning-based methods for visualizing ecosystem properties using environmental dna metabarcoding data. *Molecular Ecology Resources*, 23(8):1946–1958.

Leblanc, C., Bonnet, P., Servajean, M., Chytrỳ, M., Aćić, S., Argagnon, O., Bergamini, A., Biurrun, I., Bonari, G., Campos, J. A., et al. (2024). A deep-learning framework for enhancing habitat identification based on species composition. *Applied Vegetation Science*, 27(3):e12802.

Leblanc, C., Joly, A., Lorieul, T., Servajean, M., and Bonnet, P. (2022). Species distribution modeling based on aerial images and environmental features with convolutional neural networks. In *CLEF 2022 Working Notes-23rd Conference and Labs of the Evaluation Forum*, volume 3180, pages 2123–2150.

Leblanc, C., Picek, L., Palard, R., Deneu, B., Servajean, M., Bonnet, P., and Joly, A. (2025). Mapping biodiversity at very-high resolution in europe. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 2349–2358.

LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. (1998). Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324.

Li, B., Hou, J., Li, D., Yang, D., Han, H., Bi, X., Wang, X., Hinkelmann, R., and Xia, J. (2021). Application of lidar uav for high-resolution flood modelling. *Water Resources Management*, 35:1433–1447.

Li, J. and Asner, G. P. (2023). Global analysis of benthic complexity in shallow coral reefs. *Environmental Research Letters*, 18(2):024038.

Liu, X., Ho, L., Bruneel, S., and Goethals, P. (2024). Applications of unmanned vehicle systems for multi-spatial scale monitoring and management of aquatic ecosystems: A review. *Ecological Informatics*, page 102926.

Longo, L., Brcic, M., Cabitza, F., Choi, J., Confalonieri, R., Del Ser, J., Guidotti, R., Hayashi, Y., Herrera, F., Holzinger, A., et al. (2024). Explainable artificial intelligence (xai) 2.0: A manifesto of open challenges and interdisciplinary research directions. *Information Fusion*, 106:102301.

Lotze, H. K., Tittensor, D. P., Bryndum-Buchholz, A., Eddy, T. D., Cheung, W. W., Galbraith, E. D., Barange, M., Barrier, N., Bianchi, D., Blanchard, J. L., et al. (2019). Global ensemble projections reveal trophic amplification of ocean biomass declines with climate change. *Proceedings of the National Academy of Sciences*, 116(26):12907–12912.

Lubczonek, J., Kazimierski, W., Zaniewicz, G., and Lacka, M. (2021). Methodology for combining data acquired by unmanned surface and aerial vehicles to create digital bathymetric models in shallow and ultra-shallow waters. *Remote Sensing*, 14(1):105.

Macreadie, P. I., McLean, D. L., Thomson, P. G., Partridge, J. C., Jones, D. O., Gates, A. R., Benfield, M. C., Collin, S. P., Booth, D. J., Smith, L. L., et al. (2018). Eyes in the sea: unlocking the mysteries of the ocean using industrial, remotely operated vehicles (rovs). *Science of the Total Environment*, 634:1077–1091.

Mandeville, C. P., Nilsen, E. B., Herfindal, I., and Finstad, A. G. (2023). Participatory monitoring drives biodiversity knowledge in global protected areas. *Communications Earth & Environment*, 4(1):240.

Matabos, M., Cottais, P., Leroux, R., Cenatiempo, Y., Gasne-Destaville, C., Roullet, N., Sarrazin, J., Tourolle, J., and Borremans, C. (2025). Deep sea spy: An online citizen science annotation platform for science and ocean literacy. *Ecological Informatics*, page 103065.

McHale, M. E. and Sheehan, K. L. (2024). Bioaccumulation, transfer, and impacts of microplastics in aquatic food chains. *Journal of Environmental Exposure Assessment*, 3(3):N–A.

McKinley, G. A., Fay, A. R., Eddebbar, Y. A., Gloege, L., and Lovenduski, N. S. (2020). External forcing explains recent decadal variability of the ocean carbon sink. *Agu Advances*, 1(2):e2019AV000149.

Mellin, C., Hicks, C., Fordham, D., Golden, C., Kjellevold, M., MacNeil, M., Maire, E., Mangubhai, S., Mouillot, D., Nash, K., et al. (2022). Safeguarding nutrients from coral reefs under climate change. nat eco evo 6: 1808–1817.

Million, W. C., O'Donnell, S., Bartels, E., and Kenkel, C. D. (2021). Colony-level 3d photogrammetry reveals that total linear extension and initial growth do not scale with complex morphological growth in the branching coral, acropora cervicornis. *Frontiers in Marine Science*, 8:646475.

Morand, G., Joly, A., Rouyer, T., Lorieul, T., and Barde, J. (2024). Predicting species distributions in the open ocean with convolutional neural networks. *Peer Community Journal*, 4.

Mouquet, P., Alexandre, C., Rasolomamonjy, J., Rosa, J., Catry, T., Révillion, C., Rakotondraompiana, S., and Pennober, G. (2020). Sentinel-1 and sentinel-2 time series processing chains for cyclone impact monitoring in south west indian ocean. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 43:1593–1599.

Murphy, H. M. and Jenkins, G. P. (2010). Observational methods used in marine spatial monitoring of fishes and associated habitats: a review. *Marine and Freshwater Research*, 61(2):236–252.

Naselli-Flores, L. and Padisák, J. (2023). Ecosystem services provided by marine and freshwater phytoplankton. *Hydrobiologia*, 850(12):2691–2706.

Nieto, A., Ralph, G. M., Comeros-Raynal, M., Heessen, H., and Rijnsdorp, A. (2015). *European Red List of marine fishes*. Publications Office of the European Union.

Nomenisoa, A. L. D., Todinanahary, G., Edwin, H. Z., Razakarisoa, T., Israel, J. B., Raseta, S., Jaonalison, H., Mahafina, J., and Eeckhaut, I. (2024). Remote sensing of coral reef habitats in madagascar using sentinel-2 satellite images. *Western Indian Ocean Journal of Marine Science*, 23(2):41–56.

Olden, J. D., Lawler, J. J., and Poff, N. L. (2008). Machine learning methods without tears: a primer for ecologists. *The Quarterly review of biology*, 83(2):171–193.

OpenDroneMap Development Team (2025). OpenDroneMap: Open source toolkit for aerial image processing. Consulté le 12 mars 2025.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al. (2023). Dinov2: Learning robust visual features without supervision. *Transactions on Machine Learning Research Journal*, pages 1–31.

Pauly, D., Watson, R., and Alder, J. (2005). Global trends in world fisheries: impacts on marine ecosystems and food security. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 360(1453):5–12.

Peck, C. J., Langedock, K., Boone, W., Fourie, F., Moulaert, I., Semeraro, A., Sterckx, T., Geldhof, R., Groenendaal, B., and Ponsoni, L. (2024). The use of autonomous underwater vehicles for monitoring aquaculture setups in a high-energy shallow water environment: case study belgian north sea. *Frontiers in Marine Science*, 11:1386267.

Pelletier, D., Leleu, K., Mou-Tham, G., Guillemot, N., and Chabanet, P. (2011). Comparison of visual census and high definition video transects for monitoring coral reef fish assemblages. *Fisheries Research*, 107(1-3):84–93.

Pelletier, D., Roos, D., Bouchoucha, M., Schohn, T., Roman, W., Gonson, C., Bockel, T., Carpentier, L., Preuss, B., Powell, A., et al. (2021). A standardized workflow based on the staviro unbaited underwater video system for monitoring fish and habitat essential biodiversity variables in coastal areas. *Frontiers in Marine Science*, 8:689280.

Picek, L., Šulc, M., Matas, J., Heilmann-Clausen, J., Jeppesen, T. S., and Lind, E. (2022). Automatic fungi recognition: deep learning meets mycology. *Sensors*, 22(2):633.

Pichler, M. and Hartig, F. (2023). Machine learning and deep learning—a review for ecologists. *Methods in Ecology and Evolution*, 14(4):994–1016.

Pinsky, M. L., Clark, R. D., and Bos, J. T. (2023). Coral reef population genomics in an age of global change. *Annual Review of Genetics*, 57(1):87–115.

Pollio, J. (2006). Applications of underwater photogrammetry (reprinted from marine technology society journal, vol 3, 1969). *Marine Technology Society Journal*, 40:34–51.

Preece, A., Harborne, D., Braines, D., Tomsett, R., and Chakraborty, S. (2018). Stakeholders in explainable ai. *arXiv preprint arXiv:1810.00184*.

Qu, L., Song, X., Zhang, M., Wang, J., Wen, R., and Wang, S. (2024). Gan-based defogging and multiscale fusion approach for uav-based seagrass bed imagery semantic segmentation in challenging marine environments. In *International Conference of Pioneering Computer Scientists, Engineers and Educators*, pages 55–72. Springer.

Quaade, S., Vallebueno, A., Alcabes, O. D., Rodolfa, K. T., and Ho, D. E. (2024). Remote sensing and computer vision for marine aquaculture. *Science Advances*, 10(42):eadn4944.

Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., et al. (2021). Learning transferable visual models from natural language supervision. In *International conference on machine learning*, pages 8748–8763. PmLR.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., Sutskever, I., et al. (2019). Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.

Raine, S., Marchant, R., Kusy, B., Maire, F., Sunderhauf, N., and Fischer, T. (2024). Human-in-the-loop segmentation of multi-species coral imagery. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2723–2732.

Reaka-Kudla, M. L. et al. (1997). The global biodiversity of coral reefs: a comparison with rain forests. *Biodiversity II: Understanding and protecting our biological resources*, 2:551.

Roberts, C. J., Vergés, A., Callaghan, C. T., and Poore, A. G. (2022). Many cameras make light work: opportunistic photographs of rare species in inaturalist complement structured surveys of reef fish to better understand species richness. *Biodiversity and Conservation*, 31(4):1407–1425.

Roberts, C. M., McClean, C. J., Veron, J. E. N., Hawkins, J. P., Allen, G. R., McAllister, D. E., Mittermeier, C. G., Schueler, F. W., Spalding, M., Wells, F., Vynne, C., and Werner, T. B. (2002). Marine biodiversity hotspots and conservation priorities for tropical reefs. *Science*, 295(5558):1280–1284.

Robinson, J. P., Williams, I. D., Edwards, A. M., McPherson, J., Yeager, L., Vigliola, L., Brainard, R. E., and Baum, J. K. (2017). Fishing degrades size structure of coral reef fish communities. *Global change biology*, 23(3):1009–1022.

Rocha, R., Azevedo, F., Oliveira, U., Cardoso, M., Clerier, P., Fortes, R., Lopes-Filho, E., Lorini, M., Miranda, L., Moura, R., et al. (2024). West atlantic coastal marine biodiversity: the contribution of the platform inaturalist. *Aquatic Ecology*, 58(1):57–71.

Rogers, A., Harborne, A. R., Brown, C. J., Bozec, Y.-M., Castro, C., Chollett, I., Hock, K., Knowland, C. A., Marshell, A., Ortiz, J. C., et al. (2015). Anticipative management for coral reef ecosystem services in the 21st century. *Global Change Biology*, 21(2):504–514.

Román, A., Tovar-Sánchez, A., Olivé, I., and Navarro, G. (2021). Using a uav-mounted multispectral camera for the monitoring of marine macrophytes. *Frontiers in Marine Science*, 8:722698.

Roy, A., Saffar, M., Vaswani, A., and Grangier, D. (2021). Efficient content-based sparse attention with routing transformers. *Transactions of the Association for Computational Linguistics*, 9:53–68.

Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al. (2015). Imagenet large scale visual recognition challenge. *International journal of computer vision*, 115:211–252.

Russell, S. J. and Norvig, P. (2016). *Artificial intelligence: a modern approach*. pearson.

Ryo, M., Angelov, B., Mammola, S., Kass, J. M., Benito, B. M., and Hartig, F. (2021). Explainable artificial intelligence enhances the ecological interpretability of black-box species distribution models. *Ecography*, 44(2):199–205.

Salman, A. (2023). Application of machine learning in oceanography and marine sciences.

Sauder, J., Banc-Prandi, G., Meibom, A., and Tuia, D. (2024). Scalable semantic 3d mapping of coral reefs with deep learning. *Methods in Ecology and Evolution*, 15(5):916–934.

Schwalbe, G. and Finzel, B. (2024). A comprehensive taxonomy for explainable artificial intelligence: a systematic survey of surveys on methods and concepts. *Data Mining and Knowledge Discovery*, 38(5):3043–3101.

Senior, A. W., Evans, R., Jumper, J., Kirkpatrick, J., Sifre, L., Green, T., Qin, C., Žídek, A., Nelson, A. W., Bridgland, A., et al. (2020). Improved protein structure prediction using potentials from deep learning. *Nature*, 577(7792):706–710.

Shmueli, G. (2010). To explain or to predict? *Statistical Science*.

Silver, D., Hubert, T., Schrittwieser, J., Antonoglou, I., Lai, M., Guez, A., Lanctot, M., Sifre, L., Kumaran, D., Graepel, T., et al. (2018). A general reinforcement learning algorithm that masters chess, shogi, and go through self-play. *Science*, 362(6419):1140–1144.

Sjollema, S. B., Redondo-Hasselerharm, P., Leslie, H. A., Kraak, M. H., and Vethaak, A. D. (2016). Do plastic particles affect microalgal photosynthesis and growth? *Aquatic toxicology*, 170:259–261.

Slocum, R. K., Wright, W., Parrish, C., Costa, B., Sharr, M., and Battista, T. (2019). Guidelines for bathymetric mapping and orthoimage generation using suas and sfm, an approach for conducting nearshore coastal mapping.

Snelgrove, P. V. R. (1999). Getting to the bottom of marine biodiversity: Sedimentary habitats: Ocean bottoms are the most widespread habitat on earth and support high biodiversity and key ecosystem services. *BioScience*, 49(2):129–138.

Stuart-Smith, R. D., Edgar, G. J., Barrett, N. S., Bates, A. E., Baker, S. C., Bax, N. J., Becerro, M. A., Berkhout, J., Blanchard, J. L., Brock, D. J., et al. (2017). Assessing national biodiversity trends for rocky and coral reefs through the integration of citizen science and scientific monitoring programs. *Bioscience*, 67(2):134–146.

Su, G., Logez, M., Xu, J., Tao, S., Villéger, S., and Brosse, S. (2021). Human impacts on global freshwater fish biodiversity. *Science*, 371(6531):835–838.

Sun, W., Liu, Z., Zhang, Y., Zhong, Y., and Barnes, N. (2023). An alternative to wsss? an empirical study of the segment anything model (sam) on weakly-supervised semantic segmentation problems. *arXiv preprint arXiv:2305.01586*.

Swartout, W., Paris, C., and Moore, J. (1991). Explanations in knowledge systems: Design for explainable expert systems. *IEEE Expert*, 6(3):58–64.

Talpaert Daudon, J., Contini, M., Urbina-Barreto, I., Elliott, B., Guilhaumon, F., Joly, A., Bonhommeau, S., and Barde, J. (2023). Geoai for marine ecosystem monitoring: A complete workflow to generate maps from ai model predictions. In *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, volume 48, pages 223–230. Copernicus GmbH.

Tong, R., Yesson, C., Yu, J., Luo, Y., and Zhang, L. (2023). Key factors for species distribution modeling in benthic marine environments. *Frontiers in Marine Science*, 10:1222382.

Urbina-Barreto, I., Chiroleu, F., Pinel, R., Fréchon, L., Mahamadaly, V., Elise, S., Kulbicki, M., Quod, J.-P., Dutrieux, E., Garnier, R., et al. (2021). Quantifying the shelter capacity of coral reefs using photogrammetric 3d modeling: From colonies to reefscapes. *Ecological Indicators*, 121:107151.

Valdazo, J., Ferrer, N., Vega, C., Martín, J., Luque, Á., and Bergasa, O. (2024). Mapping marine habitats in a shallow beach-reef environment combining direct methods and hyperspectral remote sensing. *Ocean & Coastal Management*, 255:107231.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Villon, S., Mouillot, D., Chaumont, M., Darling, E. S., Subsol, G., Claverie, T., and Villéger, S. (2018). A deep learning method for accurate and fast identification of coral reef fishes in underwater images. *Ecological informatics*, 48:238–244.

Vohland, K., Land-Zandstra, A., Ceccaroni, L., Lemmens, R., Perelló, J., Ponti, M., Samson, R., and Wagenknecht, K. (2021). *The science of citizen science*. Springer Nature.

Wang, J., Wang, Y., Liu, L., Yin, H., Ye, N., and Xu, C. (2023). Weakly supervised forest fire segmentation in uav imagery based on foreground-aware pooling and context-aware loss. *Remote Sensing*, 15(14):3606.

Wang, M., Lian, S., Xiong, X., Yang, J., Chen, C., and Shi, X. (2024). Assessing the consistency and reliability of the line intercept transect method in coral cover estimation using structure from motion photogrammetry techniques. *Geocarto International*, 39(1):2322065.

Wang, Y., Deng, Y., Zheng, Y., Chattopadhyay, P., and Wang, L. (2025). Vision transformers for image classification: A comparative survey. *Technologies*, 13(1):32.

Wehn, U., Bilbao Erezkano, A., Somerwill, L., Linders, T., Maso, J., Parkinson, S., Semasingha, C., and Woods, S. (2025). Past and present marine citizen science around the globe: A cumulative inventory of initiatives and data produced. *Ambio*, pages 1–16.

Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., and Reynolds, J. M. (2012). 'structure-from-motion'photogrammetry: A low-cost, effective tool for geoscience applications. *Geomorphology*, 179:300–314.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., et al. (2020). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pages 38–45.

Xu, R., Wang, C., Sun, J., Xu, S., Meng, W., and Zhang, X. (2023). Self correspondence distillation for end-to-end weakly-supervised semantic segmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 37, pages 3045–3053.

Yang, Z., Yu, X., Dedman, S., Rosso, M., Zhu, J., Yang, J., Xia, Y., Tian, Y., Zhang, G., and Wang, J. (2022). Uav remote sensing applications in marine monitoring: Knowledge visualization and review. *Science of The Total Environment*, 838:155939.

Zhao, S., Cui, P., Shen, J., and Liu, H. (2024). Local saliency consistency-based label inference for weakly supervised salient object detection using scribble annotations. *CAAI Transactions on Intelligence Technology*, 9(1):239–249.

Zheng, Z., Liang, H., Hua, B.-S., Wong, Y. H., Ang, P., Chui, A. P. Y., and Yeung, S.-K. (2024). Coralscop: segment any coral image on this planet. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 28170–28180.

Zhou, T., Zhang, F., Chang, B., Wang, W., Yuan, Y., Konukoglu, E., and Cremers, D. (2024). Image segmentation in foundation model era: A survey. *CoRR*.

Øie Kvile, K., Gundersen, H., Poulsen, R. N., Sample, J. E., Salberg, A.-B., Ghareeb, M. E., Buls, T., Bekkby, T., and Hancke, K. (2024). Drone and ground-truth data collection, image annotation and machine learning: A protocol for coastal habitat mapping and classification. *MethodsX*, 13:102935.

Chapter 2

Seatizen atlas: a collaborative dataset of underwater and aerial marine imagery



Figure 2.1: *Example of a data collection session with two Autonomous Surface Vehicles (ASVs) in the Aldabra Atoll (Seychelles).*
Image credit: Mervyn Ravitchandirane, November 2022.

This chapter is an adapted version of the article originally published in *Scientific Data* (Contini et al., 2025). This article is a data paper describing all data and associated metadata (videos, Global Navigation Satellite System (GNSS) files, annotations, etc.) collected and used during this thesis, published in an open-access format at Seatizen Atlas. While the core contributions remain unchanged, the dataset has been updated to include new data collected in Réunion Island, as well as significant additions from a team of researchers in Madagascar who contributed 68 sessions of underwater annotated images, all curated according to the original dataset's standards.

Since its publication on 25 September 2024, Seatizen Atlas has been downloaded 1,167 times as of 21 August 2025.

Context and objectives

Recent advances in both citizen science and autonomous sensing technologies are transforming the way environmental data is collected at large-scales and high resolutions. While citizen-contributed data offers broad spatial coverage with heterogeneous data structures and quality levels and helps raise awareness among local communities about ecological challenges, autonomous platforms such as ASVs and Unmanned Aerial Vehicles (UAVs) enable systematic and high-precision image acquisition in areas that are otherwise difficult or costly to monitor (as illustrated in Figure 2.1 where a data collection event is shown in the Aldabra Atoll (Seychelles) with two ASVs). Here we describe a dataset made of both underwater and aerial imagery collected mainly in shallow tropical coastal areas by using various low cost platforms operated either by citizens or researchers. This dataset is regularly updated and contains >1.6 M images from the Southwest Indian Ocean. Most of images are geolocated and some are annotated with 51 distinct classes (e.g. fauna and habitats) to train Artificial Intelligence (AI) models. The quality of these photos taken by action cameras along the trajectories of different platforms, is highly heterogeneous (due to varying speed, depth, turbidity and perspectives) and well reflects the challenges of underwater image recognition. Data discovery and access rely on Digital Object Identifier (DOI) assignment while data interoperability and reuse is ensured by complying with widely used community standards. The Findable Accessible Interoperable Reusable (FAIR) data generation and publication steps are automated by a workflow which is also open-source to ease contributions from anyone collecting pictures.

In particular, my contributions to the dataset included the collection of underwater images using various platforms (paddleboard, snorkelling and ASVs) across multiple regions (Réunion Island, Madagascar, Seychelles and Mauritius). I also contributed to the design of the workflow that converts raw data (videos and GPS logs) into processed images with embedded metadata, helped structure the generic repository on Zenodo and participated in the image annotation process.

2.1 Introduction

The methods and tools used for the collection, processing and sharing of data are essential for building reproducible ecological monitoring systems. While much attention in AI is often given to algorithmic innovations (such as model design, hyperparameter tuning and training strategies), the structure, diversity and accessibility of the underlying data are equally important for achieving robust model performance. Thus, high-quality and well-curated datasets often have a lasting scientific impact that outlives specific models, as they enable future methods to be benchmarked and re-evaluated (Koch et al., 2021). Compared to terrestrial image recognition, underwater image recognition is a relatively recent field of research and there are a number of difficulties that come into play when one wishes to classify underwater images. First of all, the non-uniformity of the light leads to numerous shadows and plays of light that can easily be confused by the neural network with marine species. The turbidity of water and, in the case of images taken near a coral reef, background confusion add noise that makes the task more difficult. In addition, depending on the depth at which the photo was taken (with the same light conditions), the colours of the image may vary considerably because water absorbs different wavelengths of light to different degrees (the first to be absorbed is red, followed by orange and yellow). An additional image preprocessing step is therefore strongly recommended in order to reduce colour distortion (Galdran et al., 2015; Wang et al., 2019). The similarity in shape, colour and texture of different species of corals (inter-class similarity) adds another challenge in accurate species classification.

To address these challenges, we introduce a dataset that combines data collected through a participatory science project with a large volume of data collected through scientific platforms. This dataset is unique both in the types of data-collecting devices and the geographical locations covered. It integrates images captured using a variety of platforms, including kitesurfs, paddleboards, snorkelling masks, scientific divers, ASVs and UAVs. Furthermore, the dataset covers a broad geographical area, featuring data collected from various regions in the South-West Indian Ocean, such as Réunion Island, Seychelles, Madagascar and Mauritius. To maximize its impact and ensure that it can be widely distributed and benefits the broader scientific community, we have adhered closely to the FAIR data principles:

- Findable: Each data collection event is assigned a permanent DOI along with metadata elements ensuring easy discovery and retrieval by both researchers and automated indexing systems.
- Accessible: All data is freely available via open-access repositories. Additionally, we developed an interactive web application[1] that enables users to explore the data spatially, filter datasets by date or platform and initiate direct downloads.
- Interoperable: the data complies with standards widely recognized by the scientific community, including Darwin Core for biodiversity, Global Coral Reef Monitoring Network (GCRMN) standards for ecological observations made by divers and

[1] https://seatizenmonitoring.ifremer.re/home Credit: Victor Illien

Exchangeable Image File Format (EXIF)/Extensible Metadata Platform (XMP) standards for camera metadata. Geospatial data follows Open Geospatial Consortium (OGC) standards (e.g., GeoPackage), ensuring compatibility with Geographic Information System (GIS) software and other spatial data infrastructures. Additionally, segmentation annotations follow the Common Objects in Context (COCO) format, widely used in computer vision tasks, facilitating seamless integration with other ecological databases.

- Reusable: All raw data, processing scripts and derived datasets are openly available under the `CC BY 4.0` license via Zenodo and GitHub. Extensive documentation (usage metadata) accompanies these resources, ensuring transparency and facilitating future use. Additionally, all trained models are accessible on Huggingface, enabling straightforward reproduction of the presented results.

The dataset currently contains 2,012,020 images, with 2.4% of these being citizen images. To analyse platform-level variation, we randomly sampled 4,000 images per data collection platform and extracted image features using the DinoV2 model (Oquab et al., 2023). These features were then projected into a two-dimensional space using the t-SNE technique, allowing us to visualize how the model distinguishes between platforms based on the input image (Deneu et al., 2022). As shown in Figure 2.2, despite their relatively small proportion, citizen images contribute substantially to the dataset's diversity, with paddle-based images (orange) forming a cluster set apart from ASV images (blue). It is also noteworthy that UAV aerial images (red) are clearly separated from the rest of the underwater images, as expected.

All images are accurately described with rich metadata and 14,492, of which 97.5% originate from the citizen science project *Seatizen*, have been manually annotated with information about the habitat and species observed following the GCRMN and TDWG / Global Biodiversity Information Facility (GBIF) standards. Figure 2.3 shows sample images collected with different platforms and Table 2.1 and Table 2.2 show the distribution of images across different countries and years.

Table 2.1: *Number of images per territory (ISO country)*

CountryCode	Number of Images
REU	1,799,874
SYC	133,965
MUS	24,370
MDG	53,811

Table 2.2: *Number of images per year*

Year	Number of Images
2015	3,487
2021	47,357
2022	158,335
2023	1,159,948
2024	328,855
2025	314,038

Whereas in previous works cameras were stationary (Villon et al., 2018; Jalal et al., 2020) or rotating around a fixed point (Pelletier et al., 2016), here, cameras are mounted on moving platforms. This increases the variability of the type of images collected which, depending on the speed of the user, may be more or less blurred. Moreover, the diversity of platforms and observer points-of-view (see Figure 2.3) introduce diversity among

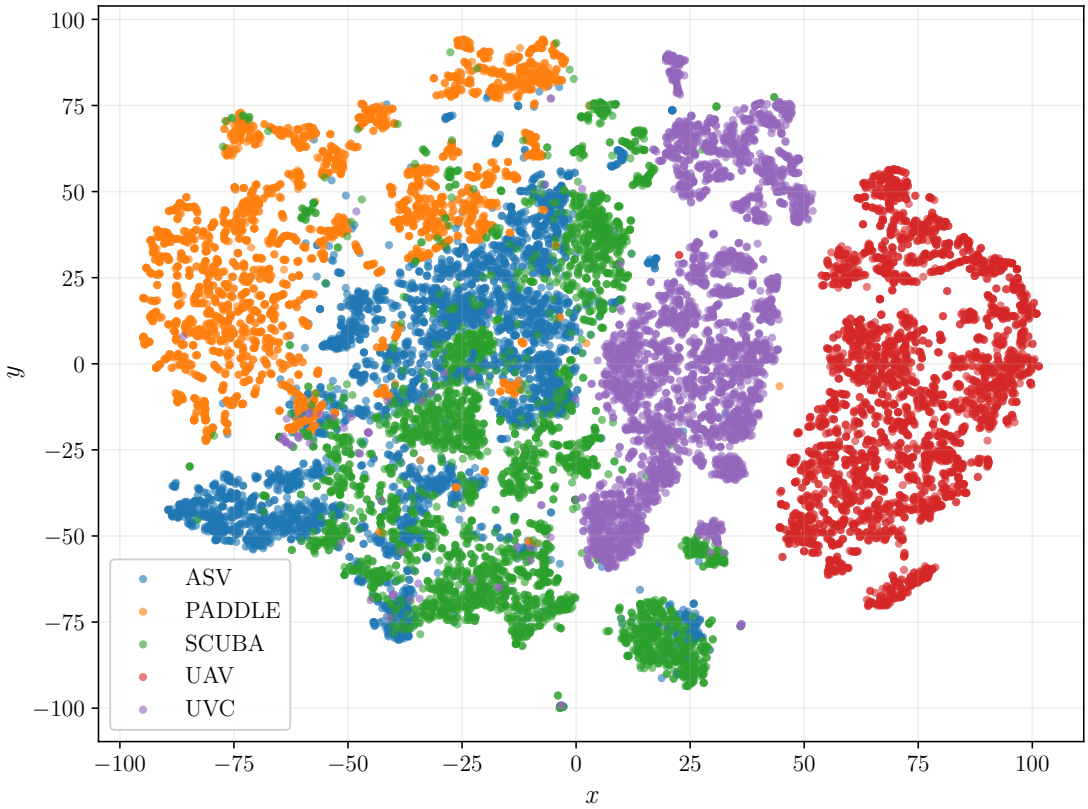
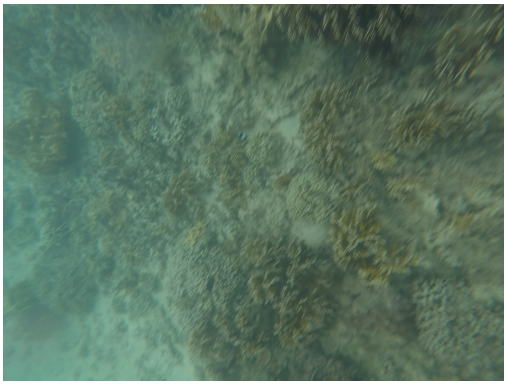


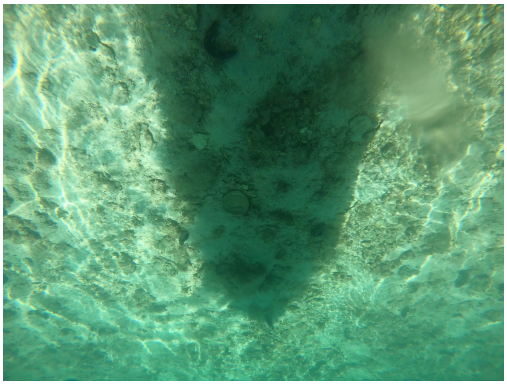
Figure 2.2: *t-SNE visualization of image embeddings extracted with DinoV2, coloured by acquisition platform (PADDLE, Underwater Visual Census (UVC), ASV, SCUBA, UAV). t-SNE is an unsupervised non-linear dimensionality reduction technique.*

images of the same marine species (intra-class dissimilarity). Indeed, pictures taken from boards are always taken from above, whereas pictures taken while snorkelling may be taken from the side.

All images provide geolocation data, which significantly enhances the dataset. However, the precision of the geolocation can greatly vary depending on the platform used to collect data (please refer to Section 2.2.5). For annotated images, this precision makes it possible to directly associate the occurrence of a taxon with latitude and longitude coordinates. These data can typically be shared as part of open international platforms such as the GBIF, which we intend to do in the near future. In the case of unannotated images, the visible taxa can be inferred through the use of artificial intelligence techniques. This data being very dense and precisely geolocated, it can then be used to construct species distribution maps at very high resolution and high survey frequencies. To this end, each underwater image in the dataset has a prediction of presence or absence for classes described in Section 2.2.6, made using the computer vision model described in Section 2.3.1.



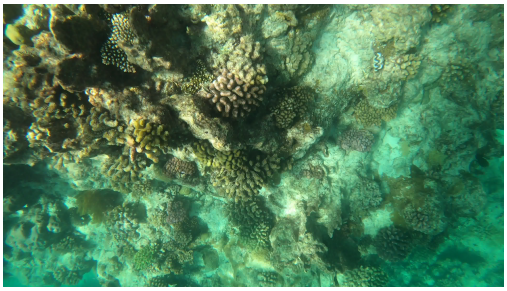
(a) *Kitesurf*



(b) *Paddleboard*



(c) *Scuba diving*



(d) *Autonomous Surface Vehicle*

Figure 2.3: *Examples of underwater images captured from four different types of platforms*

Differing from many existing public datasets in the marine biology domain such as CoralNet (Chen et al., 2021), DeepFish (Qin et al., 2016) and UIEB (Li et al., 2019), the proposed dataset and benchmark do not only aim to advance and evaluate the performance of a specific computer vision task. Instead, we think that this new dataset opens up many possibilities of use, such as:

- providing historical pictures of marine ecosystems with high-precision location, which enables returning to the same sites to assess their evolution and to establish reference states.
- evaluating existing computer vision models on spatialized images, achieving important ecological results (e.g., species distribution maps) on areas of great interest (marine protected areas, strict natural reserves or simply areas that are difficult to access).
- training/testing AI models on a subset of annotated images (e.g., train from scratch or fine-tune computer vision models on a curated annotated dataset).
- combining individual photos to generate aerial or underwater photogrammetric

models using imagery acquired by scientific platforms with sufficient horizontal and vertical overlap, as well as accurate geolocation and attitude (Roll, Pitch and Yaw) data.

- validating remote sensing models with geolocated field data, enabling rapid photo interpretation.
- provide a data structure that fosters the good practices for open science and FAIR data management by implementing various community standards (e.g., Darwin Core for biodiversity, GCRMN for scientific diving, EXIF & XMP for cameras, OGC for geospatial data and COCO for computer vision) to disseminate the same dataset.

Images are described by rich metadata that may vary depending on how they were collected. These metadata elements are a valuable resource that can be used to improve the performance of neural networks, reconstruct photogrammetry models or be used as stand-alone data to produce analysis (e.g., bathymetric maps using depth data, considering that more than 81% of images are associated with a depth measure from an echosounder). Protocols and codes are also described and openly shared for scientists and citizens for the sake of reproducibility and to contribute to building a larger collaborative resource.

2.2 Materials and Methods

2.2.1 Instrumentation

Data were collected from kitesurfs, paddleboards, snorkels, ASVs and UAV `DJI Mavic 2 Pro`. Cameras used to collect underwater images are GoPro HERO5, GoPro HERO7 and GoPro HERO8. The GPS module used to geolocate images varied depending on the platform: it was either a centimetre-resolution *Emlid Reach M2* module or a GPS-enabled watch providing metric-level precision.

2.2.2 Data acquisition system

Data acquisition was carried out using instrumented marine platforms. These platforms can be divided into two groups: citizen and scientific platforms. The first type of platforms (paddleboards, kitesurfs and snorkelling masks) are designed for being used by citizens practising marine sports. These platforms are equipped with a camera and, most of the time, with a GPS module (differential GPS module when possible, allowing the acquisition of geolocated images with centimetre accuracy). Citizen images have been collected by the "Seatizen" team with the objective of creating platforms, standards and a methodology to measure the impact that citizen science can have on scientific projects. The second type of platform consists of scientific systems. Underwater images were collected either by scientific divers or by an ASV equipped with a GPS module, an echosounder and a GoPro camera (Gogendeau et al., 2025). In diver-operated surveys,

one diver followed a compass-guided trajectory, capturing benthic images every 3–5 meters, while a second diver at the surface carried a floating bag with the GPS receiver to provide geolocation (Nomenisoa et al., 2024).



(a) Example of a citizen platform: mask equipped with a camera and a differential GPS



(b) Example of a scientific platform: Autonomous Surface Vehicle

Figure 2.4: *Examples of citizen and scientific instrumented platforms*

Figure 2.4 shows an example of both citizen and scientific platforms. Aerial data acquisition was carried out using a DJI Mavic 2 Pro aerial drone.

2.2.3 Video / camera system

Underwater images in the dataset come from photos (classic or timelapse mode) or videos taken with action cameras such as GoPro cameras. Videos were first split into frames at a fixed time interval. The frame extraction frequency can be adjusted during the post-processing phase. This flexibility allows to extract more frames when identifying interesting moments, such as when a rare species appears briefly, or fewer frames during unimportant parts of the session. The resulting frame resolution is either 1,920 px × 1,080 px or 3,840 px × 2,160 px. LRV files from GoPro can also be used instead of MP4. It results in lower resolution (854 px × 480 px) than MP4 but photogrammetry can still be done with them.

2.2.4 Data synchronization

For most of the platforms, cameras are often immersed in the water. As GPS waves cannot propagate in water, the GPS system (or antenna) has to be placed on the surface of the platform. Some images in the dataset were collected using a differential GPS module to achieve centimetre-level positioning. We use time in order to synchronize positions and images. Before each data acquisition, as differences of several seconds/minutes can be observed between the clocks of the different devices (the GPS receiver clock is not the same as the camera), the user films the time given by a GPS application on his mobile phone with the camera in order to associate the exact satellite time (UTC+0) to a specific frame or image. In the case where the time filmed with the camera follows UTC standards, leap seconds, caused by the difference between UTC and GPS time, must be

taken into account when synchronizing the GPS position with the images. This specific frame can then be used as a starting point to correct the timestamp of all images to infer their accurate location (by using the frequency at which the frames of the video were extracted). The inferred GPS positions are saved in a CSV file, enabling subsequent geospatial analysis of the images. For frames extracted from videos, the GPS position is recorded in the image EXIF metadata thanks to the *Exiftool* Perl library.

2.2.5 Image geolocation

The positional accuracy of images varies depending on the platform used to collect the data, such as paddleboards, scuba diving, ASVs or UAVs. To ensure transparency and allow validation of the position data for each data collection event, when differential GPS was used, the `metadata.csv` file located in the `METADATA` folder (see Section 2.3.1 for more information about a generic folder diagram) contains two columns, `GPSsde` and `GPSsdn`, which provide the standard deviations of the GPS position on the East and North axes respectively. In the `GPS/DEVICE` folder of each data collection event, graphs describing the position accuracy can be found. These values give an estimation of the accuracy of the GPS receiver position.

Citizen platforms

For sessions gathered by citizen scientists using platforms such as paddleboards, kitesurfs or snorkel masks, attitude (Roll, Pitch and Yaw) and depth data are not provided in image mode. When cameras are operated in video mode, attitude data can be extracted from the embedded telemetry, although exploiting them for precise geolocation of underwater remains challenging. Therefore, the positional accuracy of each image is affected by the camera's immersion depth and capture angle, which create a spatial offset between the GPS receiver and the actual image position. In future versions of the dataset, an additional uncertainty can be introduced by the offset between the GPS receiver position and the camera position, that varies in space and time within the same session (e.g., when the GPS receiver is a watch worn by the user and the camera is on the board, during a kitesurf or paddleboard session).

ASV

For sessions conducted with a differential GPS (e.g., ASV sessions) the GPS data is corrected using Post-Processing Kinematic (PPK) techniques, ensuring centimetre-level positional precision. This centimetre-level precision applies to the GPS data itself. However, it does not necessarily indicate the exact footprint of the image on the seabed. Factors such as the platform's attitude can affect the precise location of the image footprint. To account for angles of rotation on the three axes for ASV sessions, each image metadata includes attitude data collected from the ArduPilot system (Gogendeau et al., 2025), along with corresponding echosounder-derived depth data. These additional data

can be used to compute the image footprint on the seabed and further refine the image position. More information about time synchronization and metadata correction for underwater images can be found in Chapter 3.

UAV

The UAV data included in the dataset were acquired without differential GPS, resulting in meter-level positional accuracy. If Ground Control Points (GCPs) were used to refine the positioning of the Structure from Motion (SfM) model, their coordinates are saved in the `METADATA` folder, along with a description of the chosen points.

To assist users in evaluating the positioning accuracy of each SfM model, a detailed photogrammetry report
`000_photogrammetry_report.pdf` automatically generated by OpenDroneMap is provided at the root of each data collection event. Specifically, in the `Processing Summary` section of this file, the geolocation accuracy is estimated by the `GPS errors` field. Further details on this error, including GPS, GCP (when available) and 3D errors, are developed in the `GPS/GCP/3D Errors Details` section of the report.

The effects of water refraction was not considered during SfM processing. More details on the correction of aerial image positions can be found in Chapter 4.

2.2.6 Data annotation

Some images of the dataset have also been annotated to describe their content and enable image recognition. Three different types of annotation have been used in the current dataset:

- the binary convention: images are classified as "useful" or "useless" based on their ecological interest (those annotated useful are considered of interest from an ecological perspective, while those taken out of the water, too blurry or unusable are considered useless). Images where people can be identified are also removed automatically. This type of annotation is not shared because of privacy rights, as they often contain identifiable people.

- the multilabel convention: the annotator indicates all the classes that he/she observes on the image. The classes are derived from the GCRMN standards (see below).

- the instance segmentation convention: the annotation is the same as the multilabel one, but in addition to indicating all the classes present on the image, the annotator must also highlight the contours of each instance for each class.

Annotation classes are based on GCRMN scenario 2 standards, a widely accepted framework for coral reef monitoring (Conand et al., 1999). To apply GCRMN classes for deep learning and computer vision tasks, we modify the description of two classes: *Dead coral* and *Living coral*:

- The *Living coral* category was refined to only include healthy corals that could not be assigned to a specific morphotype due to image limitations (such as distance from the camera or water turbidity). This adjustment prevents all morphotype corals (e.g., *Acropora Branching*, *Acropora Digitate*) from being broadly categorized under *Living coral*, thus ensuring that no overlap occurs between different classes.
- The *Dead coral* category was modified to include all corals that have died, regardless of their original morphotype. By placing all dead corals into a single category, we avoid potential ambiguity and overlap that could arise from attempting to maintain morphotype distinctions for non-living corals.

In order to further improve the interoperability and applicability of our dataset for both AI experts and marine biologists, we have integrated the corresponding taxon for each class in the Darwin Core format promoted by GBIF. This ensures that the classification system is standardised, thus facilitating the work of both communities with the data. Further information can be found in the `seatizen_atlas_db.gpkg` file and in the `gbif_class_mapping.csv` file.

Classes were then completed with classes added by the Seatizen project team (underlined classes). In total, 38 classes were retained for annotation. Classes can be divided into 6 groups:

- Coral
  1. *Acropora Branching*
  2. *Acropora Digitate*
  3. *Acropora Submassive*
  4. *Acropora Tabular*
  5. *Bleached coral*
  6. *Dead coral*
  7. *Gorgonian*
  8. *Living coral*
  9. *Non-acropora Millepora*
  10. *Non-acropora Branching*
  11. *Non-acropora Encrusting*
  12. *Non-acropora Foliose*
  13. *Non-acropora Massive*
  14. *Non-acropora Coral free*
  15. *Non-acropora Submassive*
- Habitat
  1. *Rock*
  2. *Rubble*
  3. *Sand*
- Other Organisms
  1. *Thorny starfish*
  2. *Sea anemone*
  3. *Ascidians*
  4. *Giant clam*
  5. *Fish*
  6. *Other starfish*
  7. *Sea cucumber*
  8. *Sea urchin*
  9. *Sponges*
  10. *Turtle*
- Algae

1. *Algal Assemblage*
2. *Algae Halimeda*
3. *Algae Coralline*
4. *Algae Turf*

- Seagrass
  1. *Syringodium isoetifolium*
  2. *Thalassodendron ciliatum*
- Custom classes
  1. *Blurred*
  2. *Homo Sapiens*
  3. *Human object*
  4. *Trample*
  5. *Useless*
  6. *Waste*

More details about each class can be found in Appendix 1.1.

2.3 Data Records

Data are stored and shared using the Zenodo multi-disciplinary open repository maintained by CERN (European Organization for Nuclear Research). The dataset presented in Contini et al. (2024a) is currently made of 459 deposits which are data collection events (surveys) that have been operated either by an ASV (277), an UAV (85), by scuba diving (27), by paddleboard (2) or by UVC (68). Each survey has been assigned its own DOI and complies with standardized data structure and naming conventions. The main dataset is finally built on top of all surveys as a catalogue referencing all DOIs. The next three subsections describe these three types of data deposits and the structure of the related DOIs.

2.3.1 Session deposit

The first type of data deposit refers to a specific survey which is a data collection event. Two distinct versions may be available in each repository:

- RAW_DATA: a version containing raw data that can be either open or restricted (open under request) if containing images that are not shareable with the community (e.g., human faces or blurry images). Raw data are kept safe to ensure the work can be reproduced or improved if needed.
- PROCESSED_DATA: a version containing the processed data.

Each data survey complies with a file structure and a naming convention which is illustrated in Figure 2.5:

- each data survey has a unique identifier which is created by concatenating the following four elements (separated by the underscore symbol, e.g., `YYYYMMDD_COUNTRYCODE-optionalplace_device_session-number`):

1. the date of the session which complies with the standard ISO 8601 date format YYYYMMDD.
2. the Alpha-3 code of the country where data were collected and, if the user deems it appropriate, an additional tag describing the location where the data collection took place separated by a dash symbol (e.g., if the session took place in Le Morne area in Mauritius then the user may write MUS-LeMorne or MUS-LEMORNE or MUS-Le-Morne).
3. the type of platform used in order to collect data (e.g., scuba diving, kitesurf, paddleboard, surfboard, ASV, UAV).
4. the sequential number of the survey (needed to distinguish different surveys within the same day in the same location with the same platform).

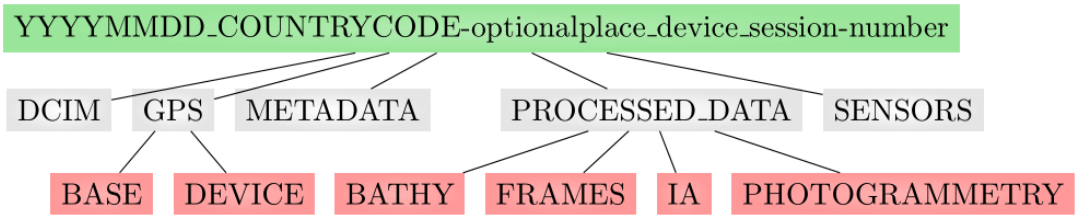


Figure 2.5: *Folder diagram of a data collection session*

Each data deposit is divided into five folders: DCIM, GPS, METADATA, PROCESSED_DATA and SENSORS. Below is a description of each of them:

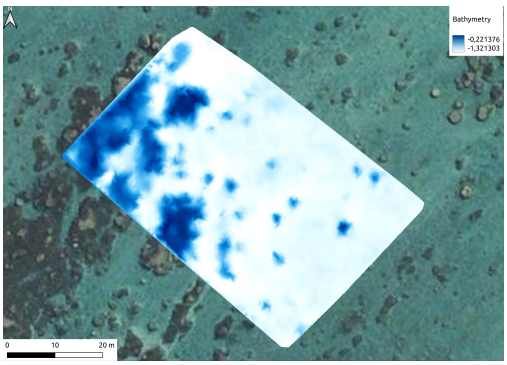
```
YYYYMMDD_COUNTRYCODE-optionalplace_device_session-number
├─ DCIM: raw media files (videos or images)
├─ GPS: any positioning related file
│  ├─ BASE: files coming from a static positioning instrument
│  └─ DEVICE: files coming from the rover
├─ METADATA: general information about the session
├─ PROCESSED_DATA
│  ├─ BATHY: bathymetry processed data
│  ├─ FRAMES: frames extracted from DCIM videos
│  ├─ IA: image recognition predictions
│  └─ PHOTOGRAMMETRY: outputs of the photogrammetric process
└─ SENSORS: files coming from other kinds of sensors
```

- DCIM: folder to store videos and photos depending on the media collected.
- GPS: folder to store any positioning related file. If any kind of correction is possible on files (e.g., Post-Processed Kinematic thanks to rinex data) then the distinction between device data and base data is made: data are then sorted in the BASE folder for files coming from Real Time Kinematic (RTK) station or any static positioning instrument and DEVICE when coming from the device. On the other hand, if only device position data are present and the files cannot be corrected by post-processing techniques (e.g., gpx files), then the distinction between base and device is not necessarily made and the files may be placed directly at the root of the GPS folder.
- METADATA: folder with general information files about the session, EXIF metadata...
- PROCESSED_DATA: contains all the folders needed to store the results of the data processing of the current session.
  - BATHY: destination folder for bathymetry processed data. See for instance Figure 2.6a.
  - FRAMES: destination folder for frames extracted from DCIM videos, enriched with metadata (when present). See for instance Figure 2.6b.
  - IA: destination folder for image recognition predictions. Inference has been made with *DinoVdeau* model. More informations is available in Section 2.4.2. See for instance Figure 2.6c.

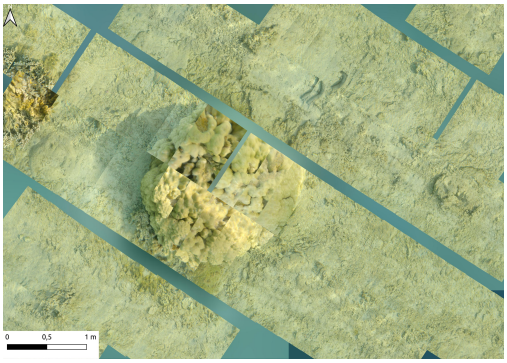
  - PHOTOGRAMMETRY: destination folder for reconstructed models in photogrammetry. See for instance Figure 2.6d.

- SENSORS: folder to store files coming from other sources (bathymetry data from the echosounder, log file from the autopilot, mission plan, etc.)

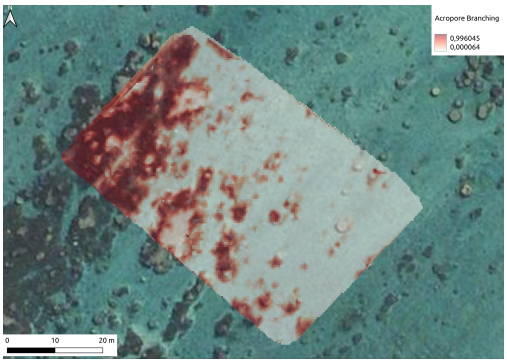
Figure 2.6 illustrates four examples of use of the dataset based on advanced technologies. These include the generation of detailed bathymetric maps, the integration of underwater images with aerial orthophotos thanks to precise georeferencing, the use of Deep Learning (DL) for underwater image classification and the creation of orthophotos using SfM techniques.



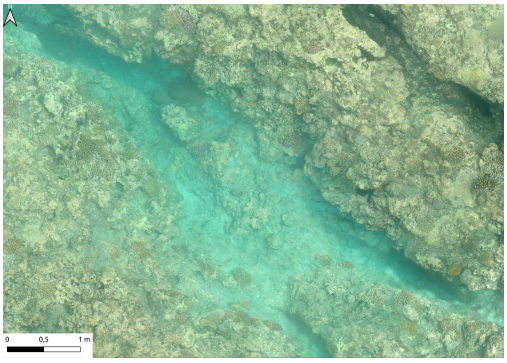
(a) *Example of a bathymetric map obtained from data collected by the ASV echosounder.*



(b) *Example of underwater ASV images geolocated using differential GPS.*



(c) *Example of a distribution map of the presence of Acropora Branching computed with a DinoV2 vision transformer trained on the annotated data of Seatizen Atlas and used to automatically make predictions on unannotated images.*



(d) *Example of an underwater orthophoto obtained from geolocated ASV images, using SfM techniques and the open-source software OpenDroneMap.*

Figure 2.6: *Examples of processed data obtained from ASV and UAV raw data in the west part of Réunion Island*

2.3.2 Global deposit

The second type of data deposit is the main one and is called Seatizen Atlas (Contini et al., 2024a), which provides a description of the dataset as a whole. This repository contains the following files:

- `seatizen_atlas_db.gpkg`: geopackage file that stores extensive geospatial data, allowing for efficient management and analysis of spatial information. It also contains its own layer styles and QGIS project to facilitate data visualization, especially in GIS software. More information about the file structure can be found in the Unified Modeling Language diagram in Figure 2.7.

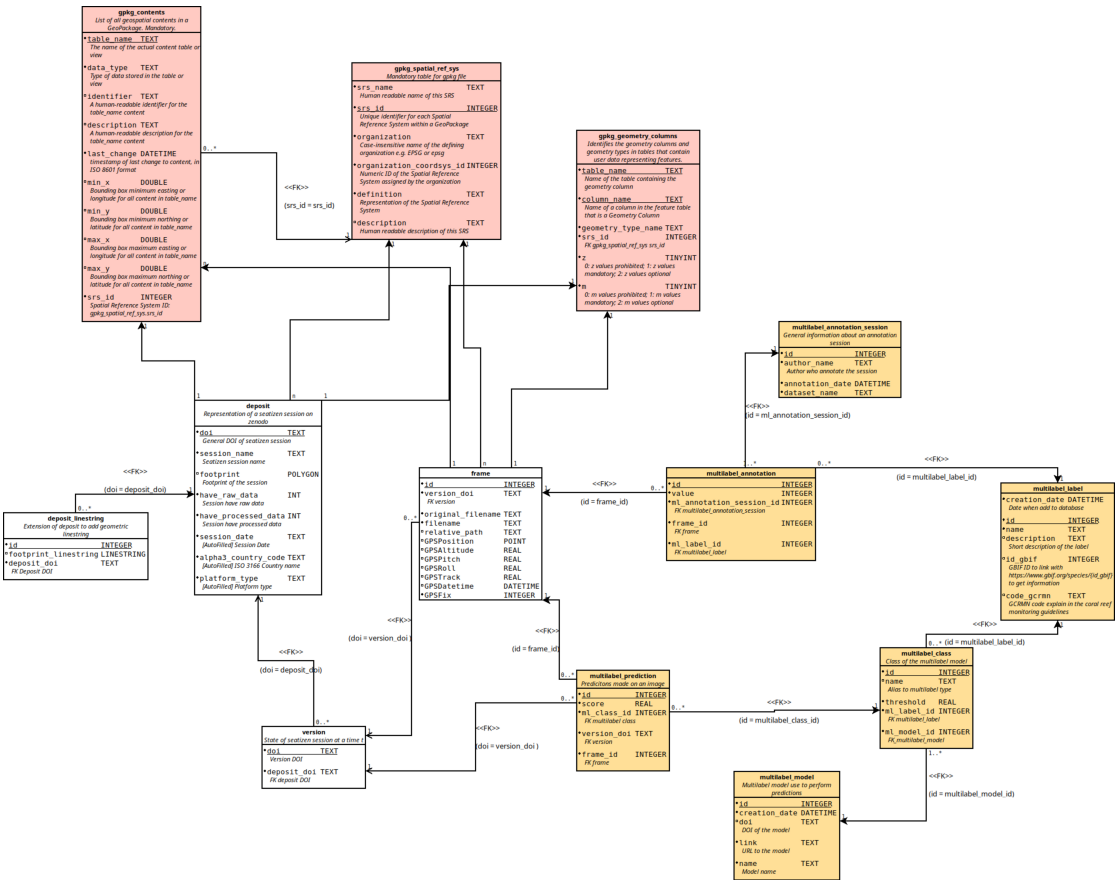


Figure 2.7: *UML diagram of the geopackage file*

- `session_doi.csv`: a CSV file listing all sessions published on Zenodo. This file contains the following columns:
  - session_name: identifies the session.
  - session_doi: indicates the URL of the session.

  - place: indicates the location of the session.
  - date: indicates the date of the session.
  - raw_data: indicates whether the session contains raw data or not.
  - processed_data: indicates whether the session contains processed data.

- `metadata_images.csv`: a CSV file describing all metadata for each image published in open access. This file contains the following columns:
  - OriginalFileName: indicates the original name of the photo.
  - FileName: indicates the name of the photo adapted to the naming convention adopted by the Seatizen team[2].
  - relative_file_path: indicates the path of the image in the deposit.
  - frames_doi: indicates the DOI of the version where the image is located.
  - GPSLatitude: indicates the latitude of the image (if available).
  - GPSLongitude: indicates the longitude of the image (if available).
  - GPSAltitude: indicates the depth of the frame (if available).
  - GPSRoll: indicates the roll of the image (if available).
  - GPSPitch: indicates the pitch of the image (if available).
  - GPSTrack: indicates the track of the image (if available).
  - GPSDatetime: indicates when the frame was taken (if available).
  - GPSFix: indicates GNSS quality levels (if available).

- `metadata_multilabel_predictions.csv`: a CSV file describing all predictions from the latest multilabel model with geolocated data.
  - FileName: indicates the name of the photo adapted to the naming convention adopted by the Seatizen team.
  - frames_doi: indicates the DOI of the version where the image is located.
  - GPSLatitude: indicates the latitude of the image (if available).
  - GPSLongitude: indicates the longitude of the image (if available).
  - GPSAltitude: indicates the depth of the frame (if available).
  - GPSRoll: indicates the roll of the image (if available).
  - GPSPitch: indicates the pitch of the image (if available).
  - GPSTrack: indicates the track of the image (if available).
  - GPSFix: indicates GNSS quality levels (if available).

[2]The naming convention adopted by the Seatizen team is based on the date of the session, the country code, an optional place name, the type of device used to collect data, the session number and the original image name: `YYYYMMDD_COUNTRYCODE-optionalplace_device_session-number_originalimagename`.

  - prediction_doi: refers to a specific AI model prediction on the current image (if available).
  - A column for each class predicted by the AI model.

- `metadata_multilabel_annotation.csv`: a CSV file listing the subset of all the images that are annotated, along with their annotations. This file contains the following columns:
  - FileName: indicates the name of the photo.
  - frame_doi: indicates the DOI of the version where the image is located.
  - relative_file_path: indicates the path of the image in the deposit.
  - annotation_date: indicates the date when the image was annotated.
  - A column for each class with values:
    * 1: if the class is present.
    * 0: if the class is absent.
    * -1: if the class was not annotated.

- `darwincore_multilabel_annotations.zip`: a Darwin Core Archive (DwC-A) file listing the subset of all the images that are annotated, along with their annotations.

These files bring together all key information in an accessible and easy-to-read format, ensuring that key data are easily available and well-organized.

2.3.3 Annotation deposit

In order to facilitate the use of annotated images, a third deposit containing all annotated images along with their corresponding annotation files was created: Seatizen Atlas image dataset (Contini et al., 2024b). This repository contains the following files:

- `20241016_132819_multilabel_annotations.csv`: a CSV file describing all multi-label annotations. This file contains the following columns:
  - FileName: indicates the name of the photo.
  - A column for each class with values:
    * 1: if the class is present.
    * 0: if the class is absent.
    * -1: if the class was not annotated.

- `segmentation_annotations.json`: a JSON file describing all instance segmentation annotations, formatted according to the COCO standard.

- `gbif_class_mapping.csv`: a CSV file mapping the dataset classes to their corresponding GBIF taxa. This file contains the following columns:

  - label: the name of the class in the current dataset.
  - old_label: the original name used in previous versions of the dataset.
  - code_gcrmn: the GCRMN code (Conand et al., 1999) (if available).
  - id_gbif: a link to the corresponding taxon in GBIF (if available).

- `images.zip`: a compressed file containing all annotated images.

This deposit currently includes 14,492 multilabel annotations and 1,200 instance segmentation annotated images.

2.4 Technical Validation

2.4.1 Annotation methodology

Multilabel annotations were made by trained marine scientists: Thomas Chevrier, Léanne Carpentier, Cam Ly Rintz, Julien Barde, Belen de Ana, Laurence Maurel and Pascal Mouquet. In order to standardise the annotations and eliminate annotator-related bias, Victor Russias (biologist working in the field of marine biology of tropical ecosystems) later took over all the annotations to ensure consistency. The annotator asked expert benthologists (Laurence Maurel, Lionel Bigot, Thierry Mulochau) to confirm their annotations.

Instance segmentation annotations were made by Laurence Maurel, expert biologist with 20 years of experience in the field.

2.4.2 Spatial data accuracy

In the current version of the *Seatizen Atlas* dataset, 96.08% of underwater images embed a position collected by a differential GPS receiver Emlid Reach M2. GPS data are saved in RINEX (Receiver Independent Exchange Format) files, which allow a spatial correction using PPK techniques. In order to do so, a base station is needed. In Réunion Island data were corrected using data from the centipede collaborative open-access network (Julien et al., 2022).

For each data collection event involving a differential GPS receiver, the GPSfix, GPSsde and GPSsdn columns recorded in the `metadata.csv` file for each image give an estimate of the accuracy of the positioning of the GPS:

- GPSfix: indicates GNSS quality levels (1, 2 or 5). A value of 1 indicates a high-precision fixed solution of the order of centimetres. A value of 2 indicates a float solution of the order of decimetres. A value of 5 indicates a single-point positioning solution of the order of meters.
- GPSsde: indicates the standard deviation of the GPS position on the East axis.
- GPSsdn: indicates the standard deviation of the GPS position on the North axis.

In Figure 2.8, the percentages of each GNSS quality level are shown for underwater images of the dataset. As can be seen, 78.71% of the images have a GNSS quality level 1, corresponding to a high-precision fixed solution at the centimetre level.

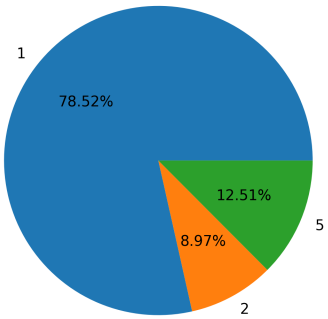


Figure 2.8: *GNSS quality levels for underwater images of the dataset.*

To better understand the accuracy of the GPS position, we compute the mean of the standard deviation of the GPS position on the East and North axes for each image. The mean of the GPSsde and GPSsdn values are 0.065 m and 0.087 m, respectively. However, computing them only on images with a GNSS quality level of 1, the mean values of both GPSsde and GPSsdn are of 5 mm.

Data availability

To facilitate the reuse, sharing and encourage new contributions to the dataset, a GitHub repository has been created at zenodo-tools Github as well as a dedicated Zenodo community. After selecting the sessions you are interested in by filtering on the global repository Seatizen Atlas, you can download them directly by following the instructions in the README file. This repository also aims to allow those who wish to contribute to the growth of the dataset to participate in a simple and intuitive manner. Instructions for uploading new sessions to Zenodo can be found in the repository.

Conclusion and continuity

The dataset presented in this chapter provides a multi-platform collection of underwater and aerial images from shallow tropical coastal ecosystems, enriched with metadata. Through the integration of low-cost citizen-science platforms with scientific ones, we demonstrate the feasibility of generating a large, geolocated dataset that adheres to FAIR principles. This resource is designed to support ecological monitoring and AI-based analyses and is openly shared to encourage reuse by the community.

The annotated subset of this dataset constitutes the training and validation foundation for the *DinoVdeau* model, which is introduced in Chapter 3. The centimetre-level spatial accuracy achieved during ASV surveys, enabled by RTK, PPK and Precise Point Positioning (PPP) geolocation techniques, combined with georeferenced UAV imagery supported by GCPs, provides the backbone for the multi-scale approach developed in Chapters 4 and 5. In this way, the dataset introduced here is not only a stand-alone scientific resource but also the cornerstone of the multi-scale workflow presented throughout this thesis: from data collection and curation, to DL-based mapping for reef ecological assessment.

Bibliography

Chen, Q., Beijbom, O., Chan, S., Bouwmeester, J., and Kriegman, D. (2021). A new deep learning engine for coralnet. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3693–3702.

Conand, C., Chabanet, P., Quod, J.-P., Bigot, L., and de Grissac, A. J. (1999). *Guidelines for coral reef monitoring in the south-west region of the Indian Ocean.* Indian Ocean Commission.

Contini, M., Barde, J., Bonhommeau, S., Illien, V., and Joly, A. (2024a). Seatizen atlas.

Contini, M., Barde, J., Bonhommeau, S., Illien, V., and Joly, A. (2024b). Seatizen atlas image dataset.

Contini, M., Illien, V., Julien, M., Ravitchandirane, M., Russias, V., Lazennec, A., Chevrier, T., Rintz, C. L., Carpentier, L., Gogendeau, P., et al. (2025). Seatizen atlas: a collaborative dataset of underwater and aerial marine imagery. *Scientific Data*, 12(1):67.

Deneu, B., Joly, A., Bonnet, P., Servajean, M., and Munoz, F. (2022). Very high resolution species distribution modeling based on remote sensing imagery: how to capture fine-grained and large-scale vegetation ecology with convolutional neural networks? *Frontiers in plant science*, 13:839279.

Galdran, A., Pardo, D., Picón, A., and Alvarez-Gila, A. (2015). Automatic red-channel underwater image restoration. *Journal of Visual Communication and Image Representation*, 26:132–145.

Gogendeau, P., Bonhommeau, S., Fourati, H., Julien, M., Contini, M., Chevrier, T., Nieblas, A. E., and Bernard, S. (2025). An autonomous surface vehicle for acoustic tracking, bathymetric and photogrammetric surveys. *Ocean Engineering*, 331:121201.

Jalal, A., Salman, A., Mian, A., Shortis, M., and Shafait, F. (2020). Fish detection and species classification in underwater environments using deep learning with temporal information. *Ecological Informatics*, 57:101088.

Julien, A., Stefal, Julien, Moinard, S., Wilthezin, François, and Bazile, R. (2022). jancelin/docs-centipedertk: v2.4.

Koch, B., Denton, E., Hanna, A., and Foster, J. G. (2021). Reduced, reused and recycled: The life of a dataset in machine learning research. *Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)*.

Li, C., Guo, C., Ren, W., Cong, R., Hou, J., Kwong, S., and Tao, D. (2019). An underwater image enhancement benchmark dataset and beyond. *IEEE transactions on image processing*, 29:4376–4389.

Nomenisoa, A. L. D., Todinanahary, G., Edwin, H. Z., Razakarisoa, T., Israel, J. B., Raseta, S., Jaonalison, H., Mahafina, J., and Eeckhaut, I. (2024). Remote sensing of coral reef habitats in madagascar using sentinel-2 satellite images. *Western Indian Ocean Journal of Marine Science*, 23(2):41–56.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al. (2023). Dinov2: Learning robust visual features without supervision. *Transactions on Machine Learning Research Journal*, pages 1–31.

Pelletier, D., Carpentier, L., Roman, W., and Bockel, T. (2016). La vidéo rotative autonome pour l'observation des habitats et de la macrofaune côtiers. guide méthodologique des systèmes staviro et micado. Report, ifremer.

Qin, H., Li, X., Liang, J., Peng, Y., and Zhang, C. (2016). Deepfish: Accurate underwater live fish recognition with a deep architecture. *Neurocomputing*, 187:49–58.

Villon, S., Mouillot, D., Chaumont, M., Darling, E. S., Subsol, G., Claverie, T., and Villéger, S. (2018). A deep learning method for accurate and fast identification of coral reef fishes in underwater images. *Ecological informatics*, 48:238–244.

Wang, K., Hu, Y., Chen, J., Wu, X., Zhao, X., and Li, Y. (2019). Underwater image restoration based on a parallel convolutional neural network. *Remote sensing*, 11(13):1591.

Chapter 3

Underwater marine biodiversity mapping in remote islands

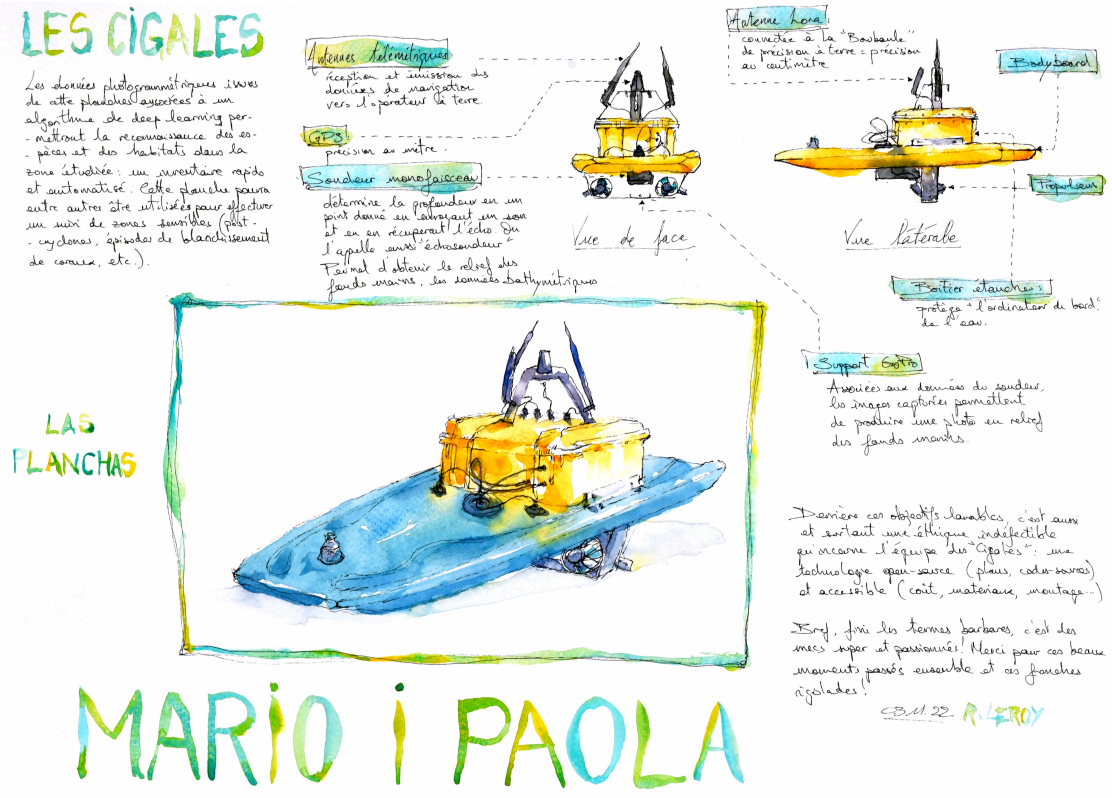


Figure 3.1: *Diagram of the Plancha Autonomous Surface Vehicle (ASV) realized by Rémi Leroy during the Indian Ocean Expedition 2022.*
Image credit: Rémi Leroy, November 2022.

This chapter is an adapted version of an article currently under revision in *Deep Sea Research II*. It describes the methodology we developed for underwater data collection during this thesis and forms the core foundation upon which the subsequent analyses and chapters are built.

Context and objectives

Developing automated tools for monitoring benthic habitats is a major challenge in marine ecology. Standardising methodologies to ensure reproducibility over time and across different users is a necessary step for transparent and long-term ecological assessments. Traditional manual annotation methods are not only expensive in human terms, but also susceptible to observer bias and limited in their spatial coverage.

This chapter presents the design and validation of our monitoring framework integrating Deep Learning (DL) techniques and geolocation methods with timestamp synchronization to embed metadata in corresponding images. A transformer-based model (DinoV2) was fine-tuned on the *Seatizen Atlas image dataset* to identify coral morphotypes, seagrass species and benthic habitats under challenging optical conditions. Developed to manage heterogeneous data collected across diverse tropical marine locations, the resulting *DinoVdeau* model addresses challenges associated with variability in environmental conditions, camera perspectives and geographic differences.

This methodology enables non-invasive, cost-effective monitoring of remote regions and supports the wider adoption of Artificial Intelligence (AI) in marine research through the use of standardized, open data and software tools. The effectiveness of this approach was tested at Réunion Island and during the "2022 Indian Ocean Expedition" at Aldabra Atoll (Seychelles).

3.1 Introduction

In the field of underwater image analysis, Computer Vision (CV) techniques are gaining prominence, including coral classification (Mahmood et al., 2016; Chen et al., 2021), fish detection (Villon et al., 2018; Ben Tamou et al., 2021) and the identification of marine species, such as sea turtles (Adam et al., 2024) or dugongs (Mannocci et al., 2024).

However, classification in underwater images faces challenges that are not present in images taken out of the water (Espinosa et al., 2023), such as uneven lighting (Marques and Albu, 2020), water turbidity, background clutter near coral reefs and colour distortion with depth (Jian et al., 2024). Fortunately, with the advent of DL models such as transformers (Vaswani et al., 2017), model performance on benchmark datasets has significantly improved: starting from Natural Language Processing (NLP) tasks up to audio processing and CV tasks (Khan et al., 2022; Lin et al., 2022). This improvement is due to the introduction of the self-attention mechanism that enables the model to focus on important parts of an image, allowing the model to capture global dependencies and better understand the context in large image datasets. This is reflected in transformers' ability to achieve State Of The Art (SOTA) performance on standard benchmark datasets. In CV for example, Vision Transformer (ViT) dominate rankings on image classification (ImageNet, CIFAR-10), object detection (COCO 2017, PASCAL VOC 2007) and semantic segmentation datasets (Cityscapes, ADE20K), where they have surpassed previous models in metrics like accuracy, mean average precision (mAP) and mean intersection over union (mIoU) (Liu et al., 2021; Xie et al., 2021).

The present study proposes a methodology for the precise and reproducible monitoring of a coastal marine site, with low costs associated with data collection and analysis. While the initial investment in advanced data collection hardware and tools may be higher than in traditional marine monitoring equipment, the overall costs are compensated by the reduction in human work. Automated data collection with ASVs reduces the need for expert human divers. In addition, DL models optimise data annotation, reducing the time spent by experienced biologists in annotating images (Mahmood et al., 2016). Combining advanced mapping techniques with SOTA image classification models, we aim to enhance our understanding of marine biodiversity hotspots. Central to our research is the development of an image classification model based on DinoV2 framework (Oquab et al., 2023), fine-tuned using the open-source Seatizen Atlas image dataset presented in Section 2.3.3.

3.2 Materials and methods

3.2.1 Study sites

Throughout the course of this thesis, the lagoons of Réunion Island served as the primary study site due to their accessibility. A total of 236 data collection sessions were carried out using ASVs, 24 using citizen-operated platforms and 52 with Unmanned Aerial Vehicles (UAVs).

In addition to Réunion, the atoll of Aldabra was selected as a secondary proof-of-concept site as part of the Indian Ocean mission. The data collected there illustrate the ability of the ASV to operate under more constrained conditions (absence of internet connection, lack of fixed GPS base stations and limited site accessibility). A diagram of the *Plancha* ASV realized by the onboard artist Rémi Leroy is shown in Figure 3.1

Réunion Island

Located in the Mascarenhas archipelago, together with the island of Mauritius, as shown in Figure 3.2), Réunion Island is a recent volcanic island (~ 3 million years old) with a total surface area of 2,512 km^2 and a great diversity of habitats and associated species.

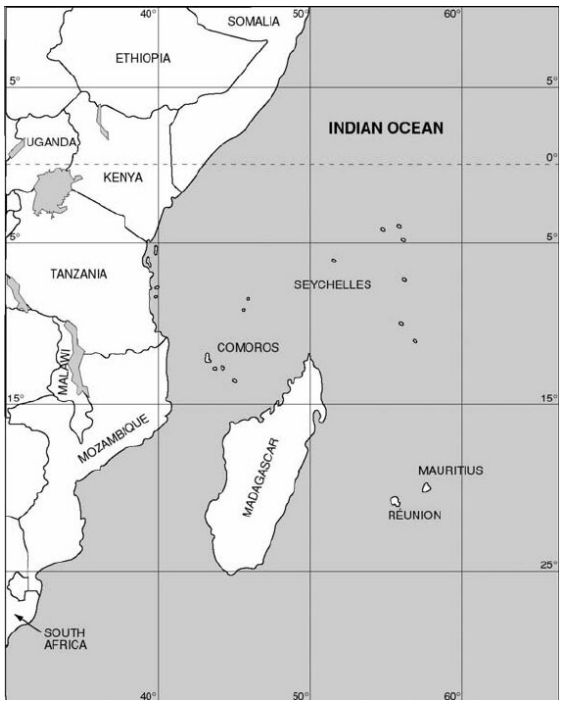


Figure 3.2: *Map of the South-Western Indian Ocean region, including the location of Réunion Island in relation to Madagascar, Seychelles, Mauritius and the eastern coast of Africa. Figure taken from (Kiszkajeremy et al., 2008).*

In the Indian Ocean, the waters surrounding Réunion Island have an average annual temperature of around 26°C (Tessier et al., 2008). Since coral reefs grow in warm waters (18-35°C) with an average thermal optimum between 26°C and 28°C (Birkeland, 1997), the average temperature of the surface waters in Réunion Island is included in the thermal optimum of a large number of coral species (thus encouraging their development).

The coral units are not evenly distributed around the island. The absence of coral reefs on the northern and eastern sides of the island can be explained by the narrowness of the continental shelf and the terrigenous inputs of rivers carrying large volumes of

detrital and sedimentary material (Faure and Montaggioni, 1976).

On the west and south-west sides 4 fringing-type coral units are observed, running from north to south in the following order: Saint-Gilles-La Saline, Saint-Leu, Étang-Salé and Saint-Pierre. Fringing reefs are defined as bio constructed formations against a reef or emerged land. These reefs are in direct contact with the coastline but may in some cases be separated from the coastline, by a channel (Cuvillier, 2016).

In 2015, the French Initiative for Coral Reefs (IFRECOR) estimated that all reef formations in Réunion Island covered a total surface area of 12 km^2 for a linear length of 25 km (Wickel, 2015). These different coral units do not form a continuous belt. This is again linked to the presence of numerous gullies carrying large quantities of organic matter and fresh water via rainwater and thus limiting the growth of corals in these areas. The situation has worsened in recent years (Broudic et al., 2025), accelerating reef degradation as extreme events like coral bleaching become increasingly frequent (Frattini et al., 2025).

Corals grow in oligotrophic waters and are generally concentrated within the upper 70 meters of the water column (D'ELIA, 1990). The coral reefs of Réunion Island have a small total surface area compared to those of Mauritius (300 km^2) and Rodrigues (200 km^2). On Réunion Island, knowledge of the biodiversity of coral reefs documented more than 2,800 species, including 984 species of fish and 150 species of scleractinians.

Aldabra

Aldabra Atoll (9°24′S, 46°22′E) is one of the most isolated regions on Earth and is among the largest raised coral atolls in the world with a total area of 365 km^2 (Hamylton et al., 2012). Aldabra's semi-enclosed lagoon covers 203 km^2 and is encircled by four main islands (Grand Terre, Malabar, Picard, Polymnie). These islands are separated by channels that are 10 to 30 meters deep, creating pathways that connect the lagoon to the ocean (Hamylton et al., 2018). Driven by a high tidal range (mean spring-tide range of 2.74 m), tidal currents are extreme: 1.5 $m\,s^{-1}$ at Passe du Bois (one of the western channels), probably reaching 3 $m\,s^{-1}$ at Passe Houareau and over 3 $m\,s^{-1}$ at Passe Gionnet, peaking at 3.7 $m\,s^{-1}$ (Farrow and Brander, 1971).

It is a unique biodiversity hotspot for its remarkable marine and terrestrial ecosystems (Stoddart, 1968; Obura et al., 2012) and has remained largely untouched by humans for the majority of its existence due to its remoteness and inaccessibility.

Despite its protected status, Aldabra now faces a growing number of environmental threats. Of particular concern is the recent land clearance and construction activity on Assomption Island, located just 27 km from Aldabra. Plans for a luxury resort have raised alarm among conservationists and researchers. While the developers claim to adhere to sustainability practices, concerns remain about the lack of independent environmental scrutiny. Moreover, the potential introduction of invasive species, habitat disturbance and increased marine traffic could deal a final blow to Aldabra's fragile ecosystems. In this context, the high-resolution baseline dataset collected during the 2022 survey provides an essential reference for detecting and quantifying any future ecological changes resulting from these emerging pressures.

In October 2022, as part of the Indian Ocean expedition led by Monaco Explorations, we conducted a seven-day data collection survey on the atoll. Two locations, ARM01 and ARM06, near the research station on Picard Island were chosen due to their inclusion in Seychelles Islands Foundation's ongoing marine monitoring program. The addition of La Gigi as site, easily accessible by foot from the research station and located within a lagoonal channel, added diversity to the dataset. The fourth survey site, Passe Dubois, is the deepest channel in a series along the west coast of the atoll, between Picard and Grand Terre islands. Collecting data at this site required careful planning because of strong tidal currents which made it difficult for the ASV, which had to move at $1\,\mathrm{m\,s^{-1}}$ for optimal data acquisition.

Two ASVs completed a total of 24 missions, contributing to an overall acquisition time of approximately 20 hours and surveying a surface area of $151{,}060\,\mathrm{m^2}$. In parallel, a UAV conducted 33 flights, representing 8 hours and 15 minutes of acquisition. Track positions are shown on Figure 3.3.

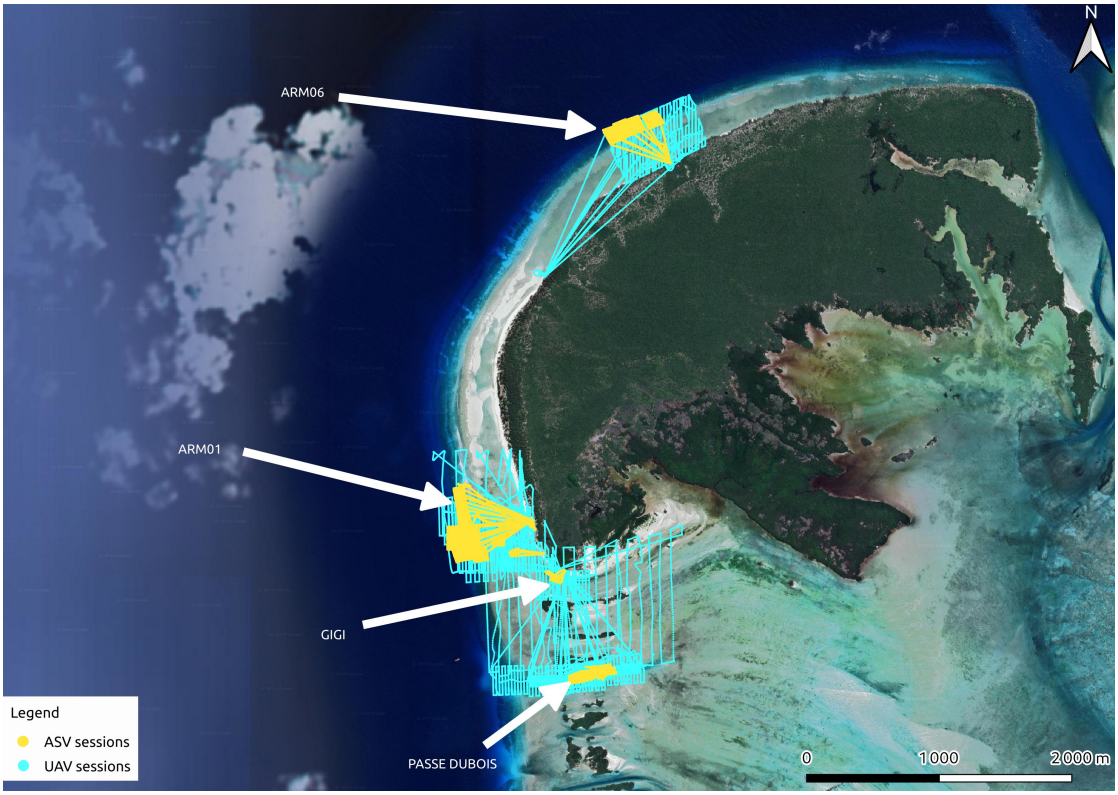


Figure 3.3: *GPS tracks of UAV (aerial drone in blue) and ASV transects (autonomous surface vehicle in yellow) during acquisitions in Aldabra.*

3.2.2 Underwater image acquisition

The *Plancha* ASV used in this work builds upon a previous version described in (Gogendeau et al., 2025), which was originally designed for bathymetric surveys using a single-beam echosounder and a differential GPS *Emlid* Reach M2. The cost of this ASV, approximately $2,434, is significantly lower compared to other ASV platforms, making it accessible for use in institutes in developing regions.

During the course of this thesis, we integrated an action camera into the system and developed a workflow to automatically enrich each captured image with metadata, such as geolocation, depth and attitude (roll, pitch and yaw). To accomplish this, the following steps were taken:

1. Time synchronization between the camera and the GPS clocks.
2. GPS position correction.
3. Bathymetry data correction using local geoid parameters and attitude data of the ASV.
4. Image geolocation using the corrected GPS position and attitude data.

Further details regarding the specifications of the ASV and the camera setting can be found in Appendix 2.

Time synchronization

Videos were cut into frames with a rate of 2.997 fps, so that the cutting frame rate $f_c = 2.997$ fps is a divisor of the video frame rate $f_v = 23.976$, ensuring that the ratio $\frac{f_v}{f_c}$ is an integer.

Since we use time in order to synchronize metadata and images, we need a method to assign a precise timestamp to each frame. Before each data acquisition, as time discrepancies of several seconds or minutes can occur between the clocks of the different devices (the GPS receiver clock is not the same as the camera), the user films the time given by a GPS application on their mobile phone with the camera in order to associate the exact satellite time (UTC+0) to a specific frame or image. In the case where the time filmed with the camera follows UTC standards, leap seconds caused by the difference between UTC time and GPS time must be taken into account when synchronizing the GPS position with the images. This specific frame can then be used as a starting point to correct the timestamp of all images by using the frame rate f_c and the number of frames between the starting frame and the frame of interest.

Cutting frames with a rate that is a divisor of the video frame rate is particularly important when working with precise position accuracy.

Indeed if the cutting frame rate is not a divisor of the video frame rate, it may happen that during the cutting process of the video a frame is skipped or duplicated, causing a misalignment between the timestamp and the corresponding frame. This is represented in Figure 3.4, where blue arrows represent frames extracted from the video

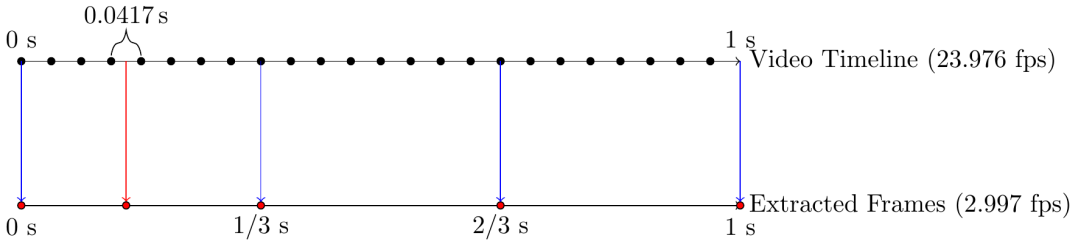


Figure 3.4: *Frame rate extraction from GoPro video. If the cutting frame rate is not a divisor of the video frame rate, a misalignment between the timestamp and the corresponding frame is introduced. This later causes an error in assigning the GPS position to the frame.*

timeline with a rate of f_c = 2.997 fps and red arrow represent a frame extracted with a random frame rate. In the first case, there is a perfect alignment between video frames and frames extracted from it. In the second case, we can see that for the required frame rate there is no corresponding frame in the original video timeline. So that, depending on the chosen option, either the frame is skipped or the closest frame is duplicated. In both cases, an error is introduced in the extracted frame timestamp. A difference in the timestamp will result in a misalignment between the real GPS position and the calculated one, which is proportional to the speed at which the ASV acquired the data. For more information about camera video setting, please refer to Appendix 2.2.

Metadata correction

During data collection with the ASV, centimetre-level GPS accuracy is achieved through the use of a differential GPS receiver and the application of Post-Processing Kinematic (PPK) corrections. This process requires both a differential GPS on the ASV and one on the base station. Two different configurations can be adopted:

- Standard configuration: in field missions conducted on Réunion Island, a mobile GPS base station was systematically deployed near each data collection site. This base station was connected to the *CentipedeRTK* network, which provides real-time correction data to improve the accuracy of the base station's own GPS position (Ancelin et al., 2023). The base station can either be directly connected to the rover in order to provide Real Time Kinematic (RTK) corrections using a 4G connection or PPK techniques can be applied to combine the rover and base station data, yielding centimetre-level accuracy for all geolocated measurements.

- Remote configuration: in missions conducted on remote islands where no local GPS base infrastructure or internet connection is available, a different approach was adopted. In these cases, communication between the ASV and the base station was established using Long Range (LoRa) radio transmission. However, since real-time correction via *CentipedeRTK* was not possible, an offline correction workflow was

implemented. One base station was set up to record Global Navigation Satellite System (GNSS) data for the longest duration possible. After the mission, the base station's position was corrected using Precise Point Positioning (PPP), a high-accuracy method that relies on long-term satellite observations. Once the corrected position of the base station was obtained, standard PPK techniques were used to refine the rover's trajectory. This method allowed centimetre-level accuracy to be maintained even in isolated locations. For more information about the geolocation process adopted at Aldabra Island during the Indian Ocean mission, please refer to Appendix 2.4.

Using the timestamp as a reference, the GPS position can then be define for each frame.

Underwater images positioning was checked in two different ways:

1. Firstly we computed the standard deviation on the east and north axes of the GPS position of the rover. If the standard deviation was less than one centimetre on both axes then the image was considered as a good one.

2. Secondly a visual check was done by visually comparing images from the same session that had a very close GPS position but a different timestamp. If the two images represent the same zone then the session was considered as a good one. An example is given in Figure 3.5. The two images taken in the *Trou d'eau* lagoon in Réunion Island are at a distance of 6.062 cm from each other and are taken 24 minutes and 49 seconds apart. It is clear how, except from the sea cucumber which has moved a little bit between the two images, the two frames represent the same zone, validating in this way the data collection event.

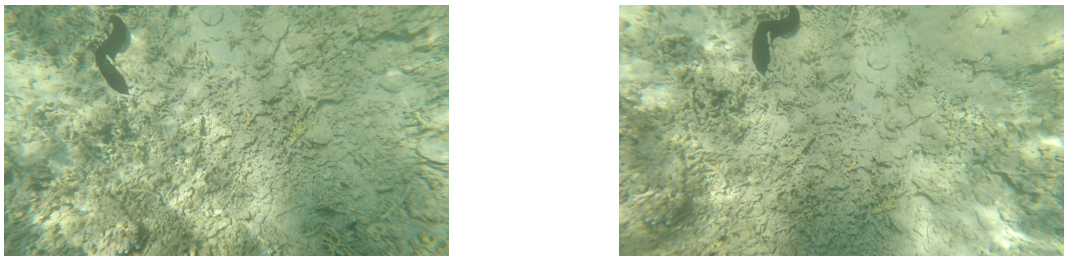


Figure 3.5: *Example of two images taken in the Trou d'eau lagoon in Réunion Island at a distance of* 6.062 cm *from each other and 24 minutes and 49 seconds apart. The similarity between the two images is an important visual criterion to validate a data collection event.*

Images extracted from video frames do not inherently contain standardized metadata. To streamline post-processing and ensure interoperability, we enriched these images by embedding standardized metadata in the headers using the `Exiftool` Perl library, following the Exchangeable Image File Format (EXIF) standard. Each image was associated with a timestamp, which served as a reference to synchronize and attach various sensor

readings recorded during the mission. These include spatial coordinates (from GPS), depth measurements (from the echosounder) and the ASV's attitude data (roll, pitch and yaw) obtained from the onboard Inertial Measurement Unit (IMU) via the Autopilot system. More information about bathymetric data collection and post-processing can be found in Appendix 2.3.

3.2.3 Image classification

Dataset

We used the open source dataset *Seatizen Atlas image dataset* composed of 51 distinct classes of corals, associated fauna and habitats (Contini et al., 2024). The image classification dataset is composed of 14,492 images annotated with the multilabel convention (i.e., the annotator indicates all the classes that he/she observes on the image without highlighting the contours of each instance for each class). These images were collected from various locations within the western Indian Ocean region, specifically from Réunion (REU), Mauritius (MUS), Mayotte (MYT), the French Southern and Antarctic Lands (ATF) and Seychelles (SYC). The distribution of images across these regions is detailed in Table 3.1.

Table 3.1: *Number of images in the Seatizen Atlas image dataset per territory*

Country	Number of Images
Réunion Island (REU)	8,229
Mauritius (MUS)	5,579
Mayotte (MYT)	323
The French Southern and Antarctic Lands (ATF)	306
The Seychelles (SYC)	55

We filter the image dataset to retain only classes with more than 200 annotations, i.e. a sufficient amount of data for model training. This results in a dataset with 31 labels composed of the scenario 2 coral morphotypes classes of the Global Coral Reef Monitoring Network (GCRMN) classification, completed with classes added by the project team. Classes can be divided into 6 groups (Table 3.2).

To support transparency and reproducibility, the DL annotated dataset has also been made available as a `Dataset` object at Huggingface, improving its discoverability and usability for AI practitioners.

Data splitting

As explained in Morais et al. (2019), the process of data splitting is a key step in the construction of image classification models. This is essential to assess their effectiveness objectively. To achieve this, we employ preprocessing techniques and leverage prior knowledge of the dataset. The approach involves dividing the samples into three distinct parts: the training set D_{tr}, the validation set D_{val} and the test set D_{test}. It's important

Table 3.2: *Labels used to annotate images composed of the scenario 2 coral morphotypes classes of the GCRMN classification and classes added by the project team*

Category	**Items**
Algae	*Algae Coralline*, *Algae Halimeda*, *Algae Turf*, *Algal Assemblage*
Coral	*Acropora Branching*, *Acropora Digitate*, *Acropora Submassive*, *Acropora Tabular*, *Bleached coral*, *Dead coral*, *Living coral*, *Non-acropora Millepora*, *Non-acropora Encrusting*, *Non-acropora Foliose*, *Non-acropora Massive*, *Non-acropora Coral free*, *Non-acropora Submassive*
Seagrass	*Syringodium isoetifolium*, *Thalassodendron ciliatum*
Habitat	*Rock*, *Rubble*, *Sand*
Other Organisms	*Fish*, *Sea cucumber*, *Atra/Leucospilota*, *Sea urchin*, *Sponges*
Custom classes	*Blurred*, *Homo Sapiens*, *Human object*, *Useless*

to note that our task involves multilabel classification, where an input can have multiple labels. This complicates the stratification process compared to monolabel classification. Traditional stratification methods designed for single-label tasks are insufficient for ensuring balance in multilabel scenarios because it's not feasible to create *tuples* (image, label), given that each image corresponds to a variable number of labels. Consequently, once an image is assigned to a dataset, all corresponding labels are assigned to it.

To build a model with high generalization capabilities, we opt for a temporal criterion to independently divide the dataset into three subdatasets. The goal is to maximize the diversity of the training dataset by including images from different periods and islands in the Indian Ocean. In our case, this temporal splitting criterion also corresponds to a spatial one, since surveys are carried out in specific locations. This approach ensures that the training dataset is not only temporally diverse but also representative of various spatial contexts. We then apply the `scikit multilabel` data stratification technique to split each subdataset into three subsets (Sechidis et al., 2011). The key goal of this stratification method is to ensure that each label is represented in each fold (train, validation and test) in proportion to its frequency in the overall dataset. To achieve this, the authors of the study first distribute samples associated with rare labels, which are harder to balance and then adjust the placement of samples with more common labels accordingly. For the purpose of achieving optimal predictive performance for our neural network, we implement the algorithm represented by the following pseudo-code:

Algorithm 1 Dataset splitting algorithm

1: **for** $i = 1, \ldots, Nb_years$ **do**
2: 1.Split the D_i into $train_i$ and $val - test_i$ sets using the Multi-label data stratification technique
3: 2.Split the $val - test_i$ set into val_i and $test_i$ sets using the Multi-label data stratification technique.
4: 3.Concatenate the current $train_i$, val_i and $test_i$ datasets to the overall D_{train}, D_{val} and D_{test} datasets.
5: **end for**

In Figure 3.6, we present the total number of images for each class, along with their corresponding distribution in the training, validation and test sets.

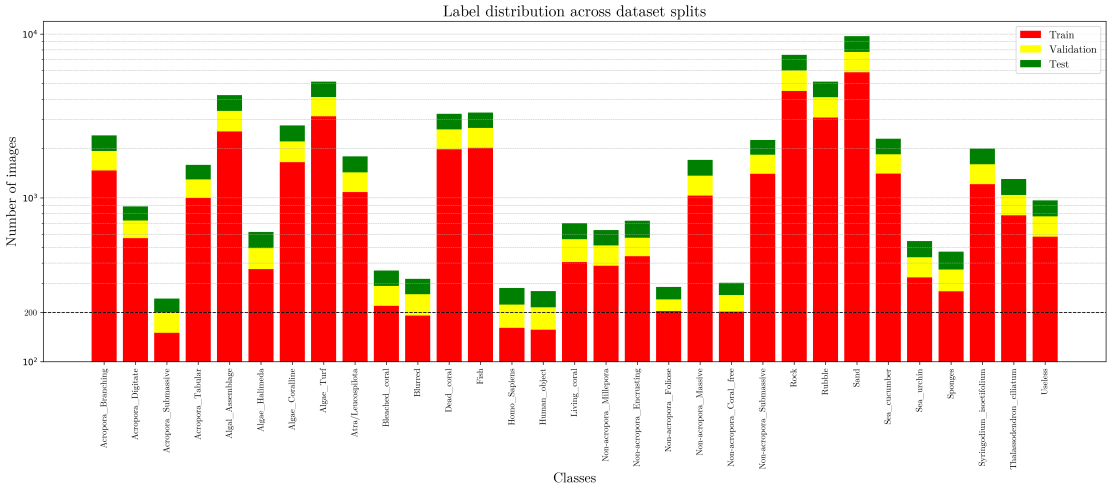


Figure 3.6: *Class-wise distribution of image annotations across the training, validation and test subsets for the 31 annotated categories in the Seatizen Atlas image dataset*

The results indicate a well-balanced class distribution.

Deep-learning model

DinoV2, a family of SOTA transformer models in CV, produces general-purpose visual features (i.e., features that work across image distributions and tasks without fine-tuning). This means that the model can be readily applied to various tasks like image classification or segmentation without necessitating the fine-tuning of the encoder (Oquab et al., 2023). Indeed, instead of training the entire model on a specific dataset, we can simply fine-tune a classification head atop the frozen encoder. [1]

To address the need for a systematic evaluation of architectural choices, we compared DinoV2 with four widely used Convolutional Neural Network (CNN) and transformer-based encoders commonly employed in CV. All models were trained under identical conditions (same training set, same classification head, same optimisation procedure and hyperparameters) ensuring a fair backbone-level comparison. Table 3.3 summarises the results. DinoV2 clearly outperforms all alternative encoders, confirming its strong transfer capabilities in the context of fine-scale marine imagery.

DinoV2 comprises four distinct models based on different sizes: small, base, large and giant. Within the DinoV2 family, we evaluated the four model sizes and, following a performance evaluation against training time experiment, we opted for the large model. Transitioning from the small to base version and from base to large version, shows

[1] To give an idea, the large version of the model (`Vit-L`) has 0.3B of parameters and the giant version (`Vit-g`) 1.1B parameters.

Table 3.3: *Fine-scale backbone comparison*

Model	Loss	F1 Micro	F1 Macro	Accuracy
resnet-50	0.1906	0.6954	0.4462	0.1827
clip-vit-base	0.1981	0.6811	0.4890	0.1745
vit-base	0.1528	0.7688	0.6487	0.2470
beit-base	0.1647	0.7446	0.6015	0.2171
dinov2-base	**0.1260**	**0.8131**	**0.6976**	**0.3014**

performance gains across all metrics (F1 Micro, F1 Macro, Roc Auc and Accuracy) with a reasonable uptick in training time, see Table 3.4. The giant version of the model, despite having slightly lower loss than the large model, has lower performances for all the other metrics and the training steps per second are almost multiplied by three, inducing a considerable increase in total training time.

Table 3.4: *Model size comparison*

Model	Loss	F1 Micro	F1 Macro	Roc Auc	Accuracy	Steps per second
DinoVdeau-small	0.1320	0.8009	0.6614	0.8649	0.2903	**0.164**
DinoVdeau-base	0.1260	0.8131	0.6976	0.8760	0.3014	0.207
DinoVdeau-large	0.1209	**0.8228**	**0.7175**	**0.8813**	**0.3111**	0.21
DinoVdeau-giant	**0.1208**	0.8209	0.7101	0.8812	0.3080	0.66

For the classification head, we used a more complex architecture than a simple linear classifier in order to capture more complex interactions between the variables extracted by the backbone model. Specifically, we added a bottleneck layer block consisting of a linear layer, a ReLU, a batch normalization and a dropout layer. The resulting model is called *DinoVdeau*[2] and the architecture is shown in Figure 3.7.

Introducing the bottleneck layer block instead of a single linear classifier results in a performance gain of approximately 0.006% and 0.015% for the F1 Micro and the Accuracy metrics respectively.

Using DinoV2 as the backbone compared to a *Resnet50* baseline results in a performance gain of approximately 0.098% and 0.140% for the F1 Micro and the Accuracy metrics respectively.

The model was trained for 93 epochs, incorporating an early stopping mechanism with a patience of 10 epochs to mitigate overfitting. Maintaining a fixed batch size of 32, the initial learning rate was established at 10^{-3} and was reduced by a factor of 10 whenever the model's performance plateaued for more than 5 epochs. Network weight updates were executed using the Adam optimization algorithm with a weight decay of 10^{-4}. More information about the model training can be found on DinoVdeau Github.

The training lasted 75 hours, on the *Datarmor* supercomputer equipped with an `NVIDIA Tesla V100 PCIe 32 GB` Graphics Processing Unit (GPU) and `32 Intel`

[2]Pronounced as /diːnoʊ viː doʊ/

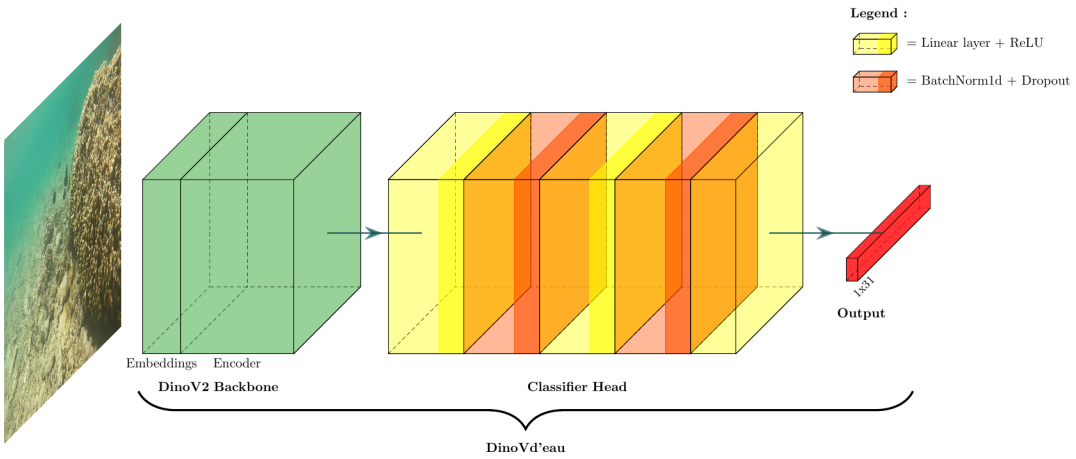


Figure 3.7: *Architecture of DinoVdeau model*

`Xeon-Gold 4216 (2.1GHz/16-core/100W)` CPU. All training was done using the Huggingface `Transformers` library described in Wolf et al. (2020). The model can be found here.

Model metrics

During the training, several metrics are calculated over the validation set: the evaluation loss, the F1 micro score, the F1 macro score and the accuracy. For a more detailed definition of these metrics, please refer to (Schlosser et al., 2024). The function used in order to monitor the training phase was the Binary Cross-Entropy (BCE) loss.

The conventional accuracy metric, calculated as a percentage based on the agreement between the model's predictions and the ground truth label, may not be the most suitable measure for multi-label classification. This is due to its limitation in capturing the nuanced aspects of errors within an image, such as scenarios where one model correctly predicts 4 out of 5 classes, while another predicts 1 out of 5 on the same image, resulting in identical accuracy scores for both, despite significant differences in prediction quality.

To address this limitation, our evaluation will employ two metrics based on the harmonic mean of precision and recall: the F1 micro score and the F1 macro score. The F1 micro score averages the F1 scores across instances, making it insensitive to imbalances in class distribution and capturing the global performance of the model. In contrast, the F1 macro score calculates the F1 score for each class and then computes the average across all classes. This metric assigns equal importance to under-represented classes, ensuring a balanced evaluation that accounts for classes with varying annotation frequencies.

Finally, to get a more detailed view of the distribution of incorrectly classified samples, we will use an extension of the confusion matrix for multi-label problems. As explained by Heydarian et al. (2022), obtaining a distribution of incorrectly classified

instances in a multilabel classification problem is non-trivial. While aggregate metrics (such as those previously analysed) offer a broad overview of the model's performance, they fall short in elucidating the specifics of incorrect prediction distribution across other classes. To gain a deeper understanding of the model's errors, we compute the normalized Multi Label Confusion Matrix (MLCM), where the diagonal indicates the recall for the corresponding class, while the off-diagonal elements represent the percentage of FN for each row class.

3.2.4 Photogrammetry survey

Overlapping images collected over a survey area can be used to generate photogrammetric models. As the ASV is equipped with a differential GPS and an autopilot system that records positioning and attitude information, each image contains metadata that significantly improve the reconstruction accuracy of the photogrammetric model (see Section 1.2.4).

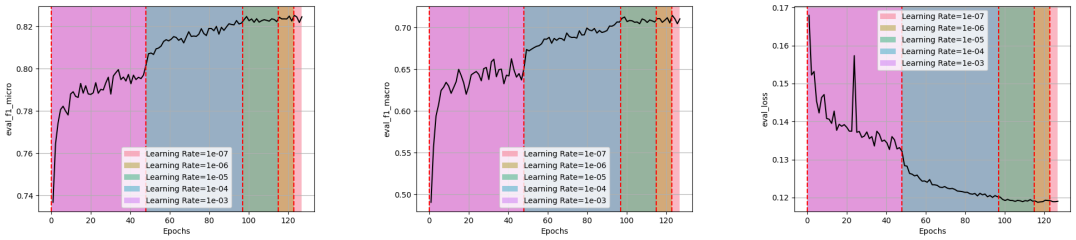
To achieve a three-dimensional reconstruction of the survey area, each image must overlap adjacent images by at least 70%, photos must be clear, free of surface reflections on the seabed and without shadows from the ASV. Using survey-specific data and camera characteristics (e.g., camera Field of View (FOV), water depth), it is possible to calculate the optimal distance between transects to satisfy the required overlap.

For the examples provided in this chapter, the ASV speed was set at $1\,\mathrm{m\,s^{-1}}$, with a transect spacing of 1 m, in survey areas ranging from 0.6 m to 7 m depth. Further information on image footprint calculations on the seabed are provided in Section 4.2.5.

3.3 Results

3.3.1 Image classification

Figure 3.8 shows the evolution of three distinct metrics employed throughout *DinoVdeau* training process on the validation dataset.



(a) Evaluation F1 micro *(b) Evaluation F1 macro* *(c) Evaluation loss*

Figure 3.8: *Evolution of evaluation metrics throughout training. Coloured background bands indicate different learning rate phases used during the training schedule.*

For all three graphs, a colorimetric division is performed in order to better appreciate the variations of the metrics as the learning rate changes. The last Figure 3.8c represents the evolution of the loss on the validation dataset, the metric chosen to monitor the evolution of the model during the training phase and monitored to change the learning rate or stop the model if there is no improvement for more than 10 epochs in order to avoid overfitting on training data. As can be seen, it causes the training of the neural network to stop after 127 epochs. The F1 macro and F1 micro metrics show a trend that aligns with the validation loss trend, finally plateauing at 0.710 and 0.825, respectively. Evaluations on the test dataset give similar results, see Table 3.5. Additional results (per-class metrics) are provided in Appendix 2.5.

Table 3.5: *Performance metrics on the test dataset*

Metric	Value
Test loss	0.119
F1 Macro	0.750 ± 0.169
F1 Micro	0.826

In Figure 3.9 we show the MLCM, the last row NTL (No True Label) is added for cases where an instance does not belong to any of the classes under consideration. Since this scenario does not occur in our dataset, the row is filled with zeros.

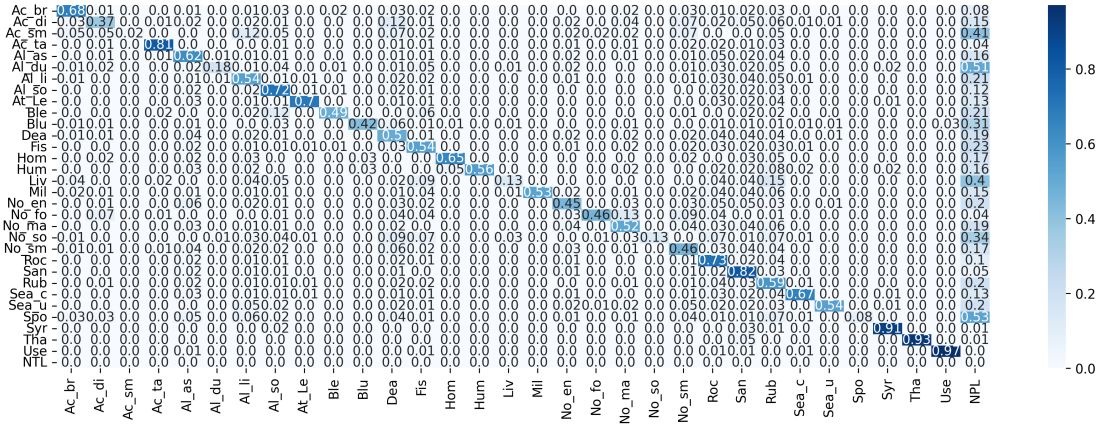


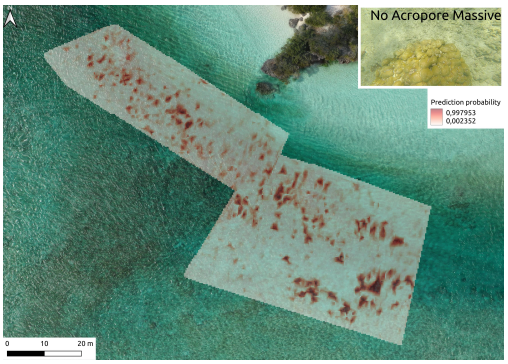
Figure 3.9: *Normalized multilabel confusion matrix showing class-wise prediction performance. Each row corresponds to the true class and each column to the predicted class. Higher values along the diagonal indicate better classification accuracy for each class, while off-diagonal values represent misclassification errors between classes.*

Similarly, the NPL (No Predicted Label) column accounts for instances where the classifier does not assign any label to an instance. This could occur when the model's confidence in all possible labels is below a certain threshold, leading to an instance being left unpredicted.

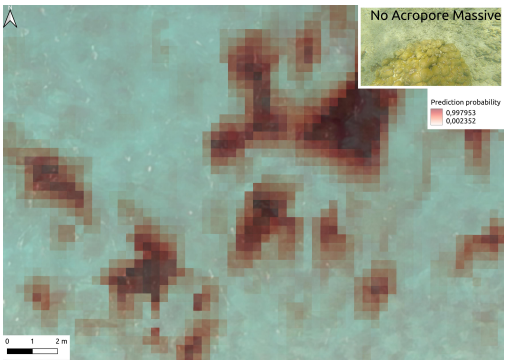
3.3.2 Prediction maps

Using *DinoVdeau*, we predicted the presence or absence of each class outlined in Section 3.2.3 for each collected image. The precise geolocation associated with each frame enables the generation of very high resolution distribution maps. This is achieved by constructing a uniform grid over the study area and then interpolating the model predictions over the entire area.

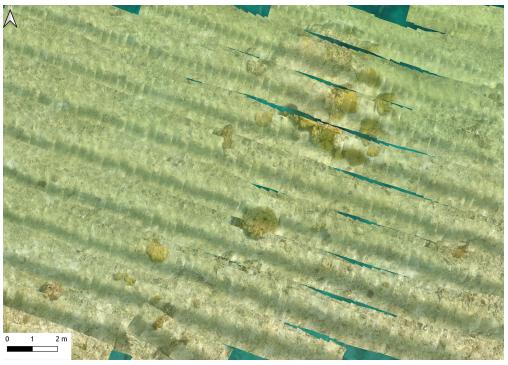
Figure 3.10 provides an illustration of the distribution maps for massive corals that are not part of the Acropora genus (*Non-acropora Massive*) at La Gigi site in Aldabra.



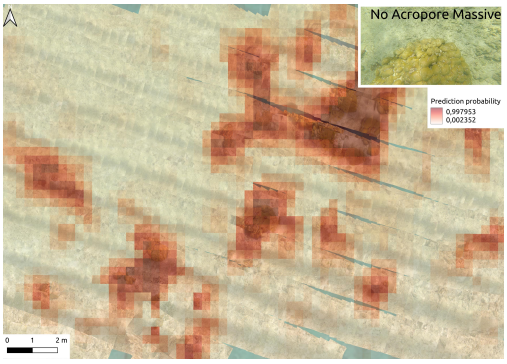
(a) *Prediction map for the Non-acropora Massive coral class at La Gigi. Red patches indicate the presence of corals.*



(b) *Zoomed-in prediction map for the Non-acropora Massive coral class at La Gigi. Red patches indicate the presence of corals.*



(c) *Zoomed-in underwater images overlaid on aerial orthophotos.*



(d) *Zoomed-in prediction map for the Non-acropora Massive coral class at La Gigi, superposed on underwater images. Red patches indicate the presence of corals.*

Figure 3.10: *Prediction maps at La Gigi for the Non-acropora Massive coral class.*

Figure 3.10a shows the prediction map from two ASV sessions covering 3,391 m^2. The background image is the corresponding orthophoto derived from aerial drone data.[3]

[3]The georeferencing was refined through a technique involving the identification of distinctive objects

The highlighted red regions represent the areas where the neural network predicted the presence of individuals belonging to the class *Non-acropora Massive*. Figure 3.10b shows a zoomed-in prediction map for the same zone, emphasizing the resolution of the prediction map. In order to visually compare the prediction map with ground truth data, underwater images were superposed on the aerial orthophoto (Figure 3.10c). The presence of *Porites* coral is then clearly visible and highlighted in Figure 3.10d.

3.3.3 ASV bathymetry

To illustrate the capabilities of the ASV-based bathymetric workflow detailed in Appendix 2.3, we present four example reconstructions from sites surveyed at Réunion Island and in Aldabra Atoll in Figure 3.11. These case studies showcase the ability of the *Plancha* ASV to generate detailed bathymetric data in both shallow and deeper reef environments.

3.3.4 Photogrammetry survey

Figure 3.12 illustrates photogrammetric reconstruction results derived from 9,143 images captured at Réunion Island.

We evaluated four photogrammetry datasets collected in Saint-Leu (21.1638°S, 55.2863°E), to assess photogrammetric reconstruction accuracy. Each dataset covered an average area of 6,182 m^2 using approximately 9,231 images.

Positioning accuracy of the reconstructed models was evaluated using two metrics: horizontal accuracy CE90 (radius containing 90% of points horizontally) and vertical accuracy LE90 (vertical difference encompassing 90% of points). These metrics were calculated using two approaches:

- Absolute accuracy: includes intrinsic GPS errors
- Relative accuracy: comparing GPS data and reconstructed values without considering GPS systematic errors

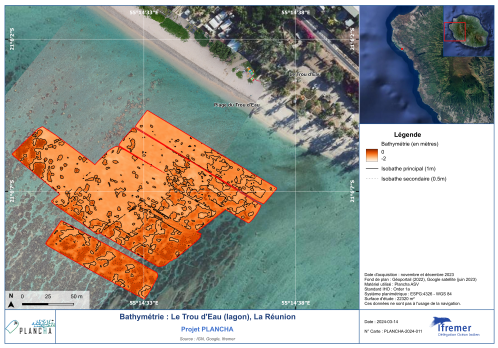
Table 3.6 summarizes the accuracy results.

Table 3.6: *Average absolute and relative accuracy metrics*

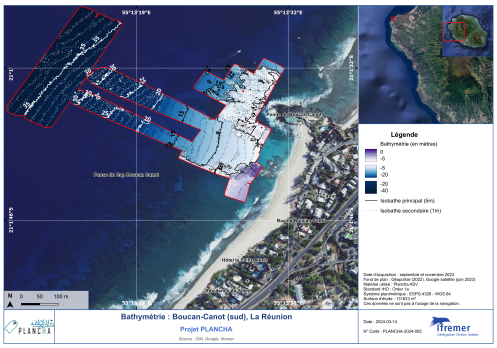
Accuracy Type	Absolute Mean (m)	Relative Mean (m)
Horizontal Accuracy CE90	0.513	0.006
Vertical Accuracy LE90	0.663	0.013

Detailed GPS and 3D reconstruction errors (mean, standard deviation, RMS) for each session are provided in session-specific photogrammetry reports, available in open-access via Zenodo (Contini et al., 2025a,b,c,d).

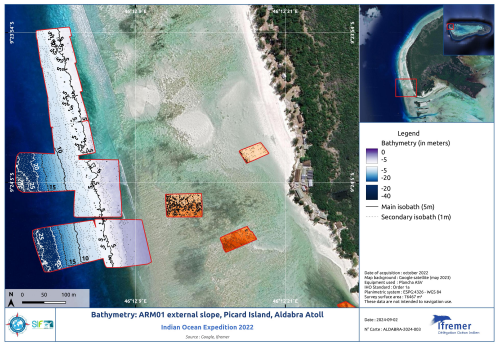
visible in both the orthophotos and underwater images, which were then overlapped using Geographic Information System (GIS) software.



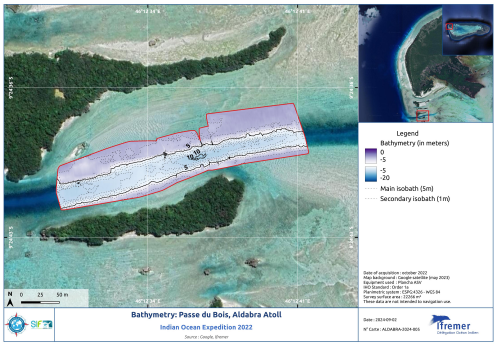
(a) Bathymetric reconstruction of the Trou d'eau lagoon at Réunion Island



(b) Bathymetric reconstruction of the Boucan Canot external slope at Réunion Island



(c) Bathymetric reconstruction of the ARM01 internal and external slope at Aldabra Island



(d) Bathymetric reconstruction of the Passe Dubois channel at Aldabra Island

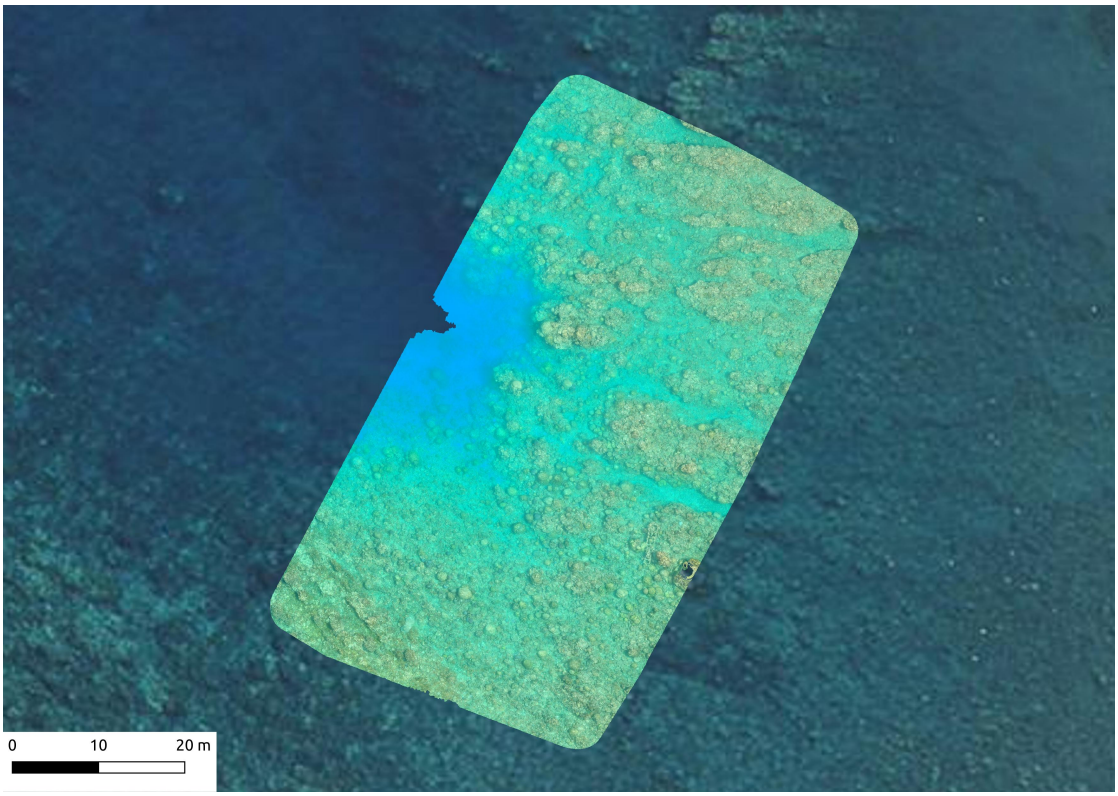
Figure 3.11: *Bathymetry maps of shallow and deep areas around Réunion and Aldabra Islands, obtained using the ASV*

3.4 Discussion

This study introduces a rapid, cost-effective and replicable method for marine biodiversity monitoring in remote regions like Aldabra Atoll. By automating data collection, this approach removes the need for divers and large survey vessels, thereby reducing safety risks and logistical constraints typically associated with bathymetric and photogrammetric surveys.

3.4.1 Image classification

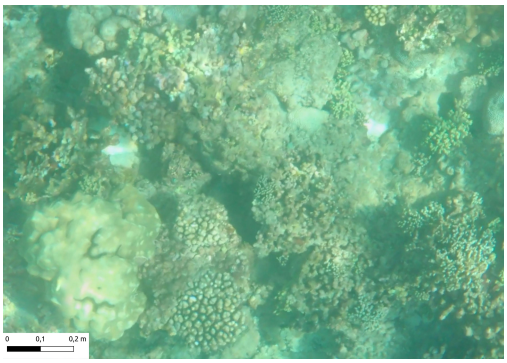
In light of the results of the MLCM displayed in Figure 3.9, it is pertinent to analyse the relationship between the model's classification performances and the dataset's characteristics, as detailed in Figure 3.6. It is interesting to note that the model's performance is often linked to the number of annotated images in the dataset. Take, for



(a) *Top-view orthophoto reconstruction covering 5,733 m².*



(b) *Zoomed-in view of a coral-colonized wreck.*



(c) *Zoomed view of diverse coral assemblages.*

Figure 3.12: *Photogrammetric reconstructions from 9,143 images captured by the ASV during a 2024 survey at Réunion Island (21.1638°S, 55.2863°E).*

example, the four classes with the lowest scores in the MLCM: *Acropora Submassive*, *Sponges*, *Living coral* and *Non-acropora Coral free* with recall values of 0.02, 0.08, 0.13 and 0.13 respectively. They have 243, 470, 250 and 303 annotated images, largely below the mean of 2048 annotated images per class. However, there are a few exceptions such as *Sea urchin* and *Non-acropora Foliose*, which have few annotations, but significantly higher recall values: 0.54 and 0.46. This highlights the model's ability to extract important features with high zero-shot capacity, achieved thanks to the training of DinoV2 backbone on 1.2 billion images. The model's performance is also related to the difficulty of recognizing certain classes due to their morphology. For instance, *Sea urchin* has a shape and colour that is easily recognizable compared to the surrounding substrate whereas *Sponges* tends to have shapes and colours more challenging to discern from the substrate. These observations confirm that the visual distinctiveness of classes plays a central role in the model's ability to classify them correctly, emphasizing the need for considering ecological and biological attributes of the subjects in model training and evaluation.

The MLCM also provides insights on potential misclassification issues, such as between visually similar coral morphotypes that can have a very different shape depending on the camera's point of view (e.g. *Acropora Tabular* and *Acropora Digitate*) or between habitat classes that may coexist on the same image (*Algae*, *Rock*, *Sand* and *Rubble*). The confusion of the *Non-acropora Submassive* class with other categories likely stems from its generic definition, which encompasses any submassive hard coral that is not an *Acropora*, see Figure 3.13. This catch-all category includes corals with highly varied colours and shapes, such as *Pocilloporidae* (pink coral on Figure 3.13a, brown coral on Figure 3.13c), *Merulinidae* (corals on Figure 3.13b) and *Dendrophylliidae* (grey coral on Figure 3.13d) genera.

The diversity within this group challenges the model's classification accuracy, as the class lacks the visual uniformity seen in more narrowly defined categories. It is clear that subdividing the *Non-acropora Submassive* class into more specific and visually coherent classes would likely improve classification precision and reduce confusion with other coral types. This adjustment would better align the model with ecological realities by recognizing the distinct characteristics and importance of each coral type within the broader marine ecosystem. The same reasoning cannot be applied to the *Non-acropora Millepora* class, which, despite encompassing both encrusting (coral on Figure 3.14b) and arborescent forms (corals on Figure 3.14a), maintains a high classification score, likely attributable to its distinctive colouration even between different species of the same genus.

This unique hue serves as a strong identifying feature, enabling the model to distinguish *Non-acropora Millepora* from other coral types with a good degree of accuracy, even though the class spans diverse growth forms. In contrast, lower performance on classes like *Sponges* are principally related to the difficulty of the model to detect them (usually they are not confused with other classes but simply no predictions are made). This highlights the need for substantial efforts to collect and annotate more images of this class.



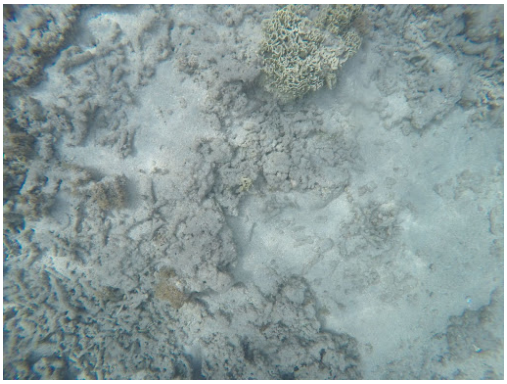
(a) *First Pocilloporidae submassive coral*



(b) *Merulinidae submassive coral*



(c) *Second Pocilloporidae submassive coral*



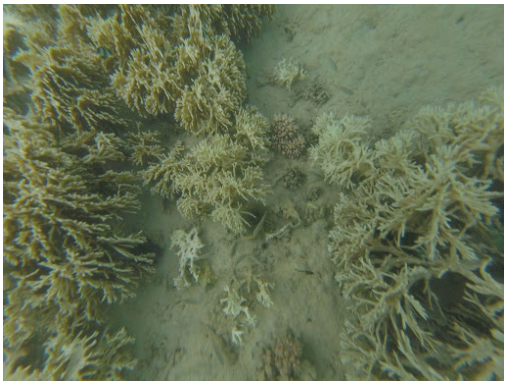
(d) *Dendrophylliidae submassive coral*

Figure 3.13: *Four different types of Non-acropora Submassive corals*

Global performance metrics reported in Table 3.5 provide an additional perspective on the model's overall behaviour. On the test dataset, the model achieves micro and macro F1 scores of 0.826 and 0.750 respectively, indicating strong overall classification performance. However, the standard deviation across class-wise F1 scores (0.169) reveals a huge variability between classes, as confirmed by the per-class F1 scores provided in Appendix 2.5. This is consistent with the patterns observed in the multilabel confusion matrix: classes with low representation or high intra-class visual diversity contribute disproportionately to this variability. While *DinoVdeau* model performs well on average, these results show that its performance is uneven across ecological categories. This reinforces the need for better balancing of training data and more refined class definitions for visually heterogeneous groups.

3.4.2 Prediction maps

Our method falls somewhere in between the Underwater Visual Census (UVC) and the UAV methods described in Section 1.2. Indeed, although the accuracy of the predictions



(a) *Arborescent form of Non-acropora Millepora coral*



(b) *Encrusting form of Non-acropora Millepora coral*

Figure 3.14: *Two different types of Non-acropora Millepora coral*

is not equivalent to that of an experienced diver-ecologist, the ability to recognise corals down to the morphotype and to locate the predictions accurately allows us to cover larger areas than UVC while still retaining key ecological information. As an example, in Figure 3.10, we show how the coverage of *Non-acropora Massive* class can be computed in a specific area.

Finally, storing the raw data on an open-source platform offers two main advantages:

- It enables temporal and cross-project comparisons. Returning to the same locations for exact result comparisons over time is of particular importance, especially given that comparisons can be challenging when methodologies differ. Furthermore, the dissemination of raw data in an open access format establishes a baseline for evaluating other research initiatives, thus enabling external teams to conduct complementary studies or integrate the dataset into broader monitoring programs.
- It ensures long-term reusability of the dataset. By preserving the raw imagery and associated metadata, the same data can later be reanalysed with more advanced or specialized AI models, such as neural networks optimized for detecting stony corals (Rusli and Mohtar, 2023) or identifying early signs of coral stress (Thamarai and Aruna, 2023), extending the scientific value of each survey.

3.4.3 ASV bathymetry

Our methodology allows the production of high-resolution bathymetric maps that comply with the requirements of the *order 1a classification* of the *IHO Standards for Hydrographic Surveys, Ed. 6.0.0 Sept 2020.* Nevertheless, in order to achieve such precision, a local geoid model specific to the surveyed region must be applied. This guarantees that depth measurements are accurately referenced to mean sea level, taking into account local variations in the Earth's surface. In surveys conducted at Réunion Island, geodesic corrections have been applied (see Figure 3.11a and Figure 3.11b) resulting in

consistency of the bathymetric models across different missions. By contrast, Aldabra Atoll currently lacks a specific local geoid model, which limits the ability to produce high-precision depth maps referenced to mean sea level. Establishing a geodetic control network similar to the campaign conducted in 2019 that led to the creation of the *SeyGeoid 2019* model for other parts of the Seychelles, would improve the accuracy of spatial data collected around Aldabra. This would allow for the accurate geolocation of past and future data collected on the island and its surrounding waters, making possible the comparison over time of bathymetric data (Stoykov, 2022).

Moreover, although the installation of a GNSS base for six days and the application of PPK and PPP techniques allowed us to obtain centimetre-level accuracy in positional measurements, the establishment of a permanent GNSS base station on Aldabra Atoll would facilitate the geolocation of both future survey campaigns and the continuous monitoring activities of the permanent Seychelles Islands Foundation station.

3.4.4 Photogrammetry survey

The photogrammetric reconstructions presented in Figure 3.12 demonstrate the feasibility of collecting geolocated images with an ASV to generate 3D models of shallow reef environments. Despite the constraints imposed by surface wave motion and colour distortion due to water depth, the resulting orthophotos show a good level of detail capturing fine-scale features such as geological structures (Figure 3.12a), anthropogenic elements like ship wrecks (Figure 3.12b) and coral assemblages (Figure 3.12c).

The positioning accuracy assessment confirms the robustness of the workflow, with sub-decimetric relative positioning errors (see Table 3.6). Although absolute accuracy is lower, this can be improved by integrating Ground Control Points (GCPs) during the survey process. These points would serve as fixed spatial references, allowing for more accurate georeferencing of the final orthomosaics (see Section 4.2.2).

These findings validate the feasibility of ASV-based photogrammetric mapping in marine environments. In particular, they highlight the ability to map large areas (6,000 m^2 in the provided example) without requiring underwater deployment or diver-operated systems, thus reducing logistical complexity and enabling surveys in shallow water environments with restricted access. However, a key limitation of surface-based imaging is that 3D reconstructions are constrained to features visible from the water surface. Moreover, while the native image resolution of the camera remains constant, increasing the camera-to-target distance enlarges the Ground Sampling Distance (GSD), meaning that each pixel represents a larger area on the seafloor. As a result, fine-scale details become less discernible, so while these 3D datasets can still be used to compute ecologically relevant metrics such as rugosity, fractal dimension and structural complexity, these indicators are typically more accurate when derived from underwater imagery.

Finally, the resulting orthophotos present valuable opportunities for ecological applications. When combined with semantic segmentation models, they can be used to monitor spatial and temporal changes in benthic communities (Sauder et al., 2024), as well as to derive structural complexity metrics relevant to reef health assessments (Young et al., 2018). This integration opens the door to scalable, automated ecologi-

cal monitoring pipelines built on affordable and replicable ASV-based photogrammetry workflows.

3.4.5 Potential improvements

While the approach presented in this chapter provides a foundation for cost-effective marine habitat mapping, there are several ways it could be improved.

One key area for progression is the enrichment of the training datasets. As observed in the classification results, under-represented classes such as *Sponges* or *Acropora Submassive* corals suffer from low recall values due to a limited number of annotated samples. A targeted annotation campaign could help reduce this imbalance. Closely tied to this is the need to refine the class taxonomy. Some categories, like *Non-acropora Submassive*, currently serve as catch-all groups for morphologically varied organisms, which introduces ambiguity during training and prediction. Dividing such classes into more ecologically and visually coherent subclasses would not only improve classification accuracy but also allow more meaningful ecological interpretations of model outputs. In parallel, incorporating metadata into the training process such as water depth, geographic coordinates or spatial relationships between objects (e.g., distance to the coastline, distance to the reef barrier) could further improve the model's ability to distinguish visually similar classes that are ecologically distinct. For example, some species may inhabit only shallow reef crests, while others are found primarily in deeper or more sheltered zones like seagrass close to the coast in shallow waters. Providing the model with such spatial or environmental information can guide its predictions and reduce confusion between morphologically similar but habitat-specific classes, especially in multi-site applications where regional endemism plays a role.

From a geospatial standpoint, integrating GCPs into photogrammetry surveys would substantially increase absolute georeferencing accuracy, essential for long-term monitoring efforts.

A promising extension of the current methodology would involve coupling ASV-based data collection with UAV surveys, an approach that will be explored in the next chapters.

Data availability

The data that support the findings of this study are openly available in Zenodo at Seatizen Atlas. All code for data processing associated with the current submission is available on plancha-workflow Github.

The code for downloading data associated with the current submission is available on zenodo-tools Github.

The code used to train the neural network model used in the current submission is available on DinoVdeau Github.

Conclusion and continuity

This study shows a monitoring method that is suitable for long-term, cost-effective and accurate conservation strategies. By integrating low-cost technologies with cutting-edge AI methodologies, we have demonstrated the feasibility of developing a workflow that acquires precise data and then performs automated analysis.

The DL model we developed, *DinoVdeau*, was able to accurately predict the presence of different classes of corals and habitats in the collected images. "Beyond the model itself, the methodology and workflow developed in this chapter are designed to be both reusable and easily adaptable. The same raw data can be reprocessed in the future using updated versions of *DinoVdeau* or alternative AI models, thereby continuously improving the quality and scope of the generated maps.

In addition, the integration of ASV-based bathymetric and photogrammetric analyses has proven to be a reliable method for generating fine-resolution 3D maps of the seafloor. These outputs provide spatial context for habitat interpretation and open the door to multi-source reef monitoring that combines depth structure with biological classification and can also be used as input for AI segmentation models.

The integration of underwater imaging with open-source ASV platforms presented in this chapter serves as the foundation for the multi-scale remote sensing strategy developed in the following chapters. Building on the fine-resolution, geolocated predictions derived from *DinoVdeau*, the next chapter introduces a method for extending ecological information across spatial scales using aerial imagery acquired by drones. In this framework, the deep learning model trained on underwater imagery acts as a teacher, providing pointwise predictions over small areas. These predictions are then spatially transferred to co-located aerial tiles, where they serve as annotations for training a second model (the student), which is tasked with classifying larger reef areas based solely on UAV imagery. This transfer of information across modalities and spatial resolutions enables the student model to learn from the fine-scale ecological detail of the underwater domain, while leveraging the broader coverage and operational flexibility of aerial platforms.

Bibliography

Adam, L., Čermák, V., Papafitsoros, K., and Picek, L. (2024). Seaturtleid2022: A long-span dataset for reliable sea turtle re-identification. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 7146–7156.

Ancelin, J., Ladet, S., and Heintz, W. (2023). Le Real Time Kinematic collaboratif, lowcost et open source. Positionnement GNSS temps réel, cinématique, collaboratif et en accès libre et à faible coût. In Badard, T., Pouliot, J., Noucher, M., and (Eds), M. V.-O., editors, *Spatial Analysis and GEOmatics 2023*, Actes de la conférence Spa1al Analysis and GEOma1cs (SAGEO) 2023, pages 184–197, Québec, Canada. GDR MAGIS Méthodes et Applications pour la Géomatique et l'Information Spatiale and Centre de Recherche en Données et Intelligence Géospatiales de l'Université Laval (Québec).

Ben Tamou, A., Benzinou, A., and Nasreddine, K. (2021). Multi-stream fish detection in unconstrained underwater videos by the fusion of two convolutional neural network detectors. *Applied Intelligence*, 51(8):5809–5821.

Birkeland, C. (1997). *Life and death of coral reefs*. Springer Science & Business Media.

Broudic, L., Pinault, M., Claud, R., Nicet, J., Wickel, J., Bajjouk, T., Rungassamy, T., Bigot, L., Nikolic, N., Crochelet, E., et al. (2025). Toward a spatialized index of the "benthic ecological state" using a new coral reef rapid assessment method (corram): case study at reunion island. *Environmental and Sustainability Indicators*, page 100811.

Chen, Q., Beijbom, O., Chan, S., Bouwmeester, J., and Kriegman, D. (2021). A new deep learning engine for coralnet. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 3693–3702.

Contini, M., Barde, J., Bonhommeau, S., Illien, V., and Joly, A. (2024). Seatizen atlas image dataset.

Contini, M., Illien, V., Bonhommeau, S., Barde, J., and Joly, A. (2025a). Underwater images collected by an autonomous surface vehicle in st-leu, réunion - 2023-11-10.

Contini, M., Illien, V., Bonhommeau, S., Barde, J., and Joly, A. (2025b). Underwater images collected by an autonomous surface vehicle in st-leu, réunion - 2023-11-10.

Contini, M., Illien, V., Bonhommeau, S., Barde, J., and Joly, A. (2025c). Underwater images collected by an autonomous surface vehicle in st-leu, réunion - 2023-11-10.

Contini, M., Illien, V., Bonhommeau, S., Barde, J., and Joly, A. (2025d). Underwater images collected by an autonomous surface vehicle in st-leu, réunion - 2023-11-10.

Cuvillier, A. (2016). *Dynamique et fonctionnement des herbiers marins dans un complexe récifal anthropisé (île de la Réunion, océan Indien).* PhD thesis, Université de la Réunion.

D'ELIA, C. F. (1990). Biogeochemical nutrient cycles in coral-reef ecosystems. *Coral reefs*.

Espinosa, A. R., McIntosh, D., and Albu, A. B. (2023). An efficient approach for underwater image improvement: Deblurring, dehazing, and color correction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 206–215.

Farrow, G. and Brander, K. (1971). Tidal studies on Aldabra. *Philosophical Transactions of the Royal Society of London. B, Biological Sciences*, 260(836):93–121.

Faure, G. and Montaggioni, L. (1976). Les récifs coralliens au-vent de l'île maurice (archipel des mascareignes, océan indien): géomorphologie et bionomie de la pente externe. *Marine Geology*, 21(1):M9–M16.

Frattini, B., Bruggemann, J. H., Goberville, E., Couëdel, M., Bruggemann, F., and Guillaume, M. M. (2025). Seasonal colonisation and ecological succession shape coral reef sessile cryptobenthic communities in autonomous reef monitoring structures. *Scientific Reports*, 15(1):23232.

Gogendeau, P., Bonhommeau, S., Fourati, H., Julien, M., Contini, M., Chevrier, T., Nieblas, A. E., and Bernard, S. (2025). An autonomous surface vehicle for acoustic tracking, bathymetric and photogrammetric surveys. *Ocean Engineering*, 331:121201.

Hamylton, S., Hagan, A., Bunbury, N., Fleischer-Dogley, F., and Spencer, T. (2018). Mapping the lagoon at aldabra atoll, western indian ocean. *Atoll Research Bulletin*, 2018.

Hamylton, S., Spencer, T., and Hagan, A. (2012). Spatial modelling of benthic cover using remote sensing data in the aldabra lagoon, western indian ocean. *Marine Ecology Progress Series*, 460:35–47.

Heydarian, M., Doyle, T. E., and Samavi, R. (2022). Mlcm: Multi-label confusion matrix. *Ieee Access*, 10:19083–19095.

Jian, M., Yang, N., Tao, C., Zhi, H., and Luo, H. (2024). Underwater object detection and datasets: a survey. *Intelligent Marine Technology and Systems*, 2(1):9.

Khan, S., Naseer, M., Hayat, M., Zamir, S. W., Khan, F. S., and Shah, M. (2022). Transformers in vision: A survey. *ACM computing surveys (CSUR)*, 54(10s):1–41.

Kiszkajeremy, J., Muir, C., Poonian, C., Bourjea, J., Razafindrakoto, Y., Wambiji, N., Bristol, N., et al. (2008). Marine mammal bycatch in the southwest indian ocean: review and need for a comprehensive status assessment. *Western Indian Ocean Journal of Marine Science*, 7(2):119–136.

Lin, T., Wang, Y., Liu, X., and Qiu, X. (2022). A survey of transformers. *AI open*, 3:111–132.

Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., and Guo, B. (2021). Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 10012–10022.

Mahmood, A., Bennamoun, M., An, S., Sohel, F., Boussaid, F., Hovey, R., Kendrick, G., and Fisher, R. (2016). Automatic annotation of coral reefs using deep learning. In *Oceans 2016 mts/IEEE monterey*, pages 1–5. IEEE.

Mannocci, L., Derville, S., Seguin, R., and Mouillot, D. (2024). Aerial video surveys and spatial prioritization reveal conservation opportunities for endangered dugongs in new caledonia. *Aquatic Conservation: Marine and Freshwater Ecosystems*, 34(8):e4237.

Marques, T. P. and Albu, A. B. (2020). L2uwe: A framework for the efficient enhancement of low-light underwater images using local contrast and multi-scale fusion. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pages 538–539.

Morais, C. L., Santos, M. C., Lima, K. M., and Martin, F. L. (2019). Improving data splitting for classification applications in spectrochemical analyses employing a random-mutation kennard-stone algorithm approach. *Bioinformatics*, 35(24):5257–5263.

Obura, D., Church, K. B., Gabrié, C., and Macharia, E. (2012). *Assessing Marine World Heritage from an Ecosystem Perspective–N° 32–The Western Indian Ocean*, volume 32. UNESCO.

Oquab, M., Darcet, T., Moutakanni, T., Vo, H., Szafraniec, M., Khalidov, V., Fernandez, P., Haziza, D., Massa, F., El-Nouby, A., et al. (2023). Dinov2: Learning robust visual features without supervision. *Transactions on Machine Learning Research Journal*, pages 1–31.

Rusli, N. N. and Mohtar, I. A. (2023). Stony coral species recognition system using deep learning. In *2023 4th International Conference on Artificial Intelligence and Data Sciences (AiDAS)*, pages 325–330. IEEE.

Sauder, J., Banc-Prandi, G., Meibom, A., and Tuia, D. (2024). Scalable semantic 3d mapping of coral reefs with deep learning. *Methods in Ecology and Evolution*, 15(5):916–934.

Schlosser, T., Friedrich, M., Meyer, T., and Kowerko, D. (2024). A consolidated overview of evaluation and performance metrics for machine learning and computer vision. *Tobias Schlosser, Michael Friedrich, Trixy Meyer, and Danny Kowerko–Junior Professorship of Media Computing, Chemnitz University of Technology*, 9107.

Sechidis, K., Tsoumakas, G., and Vlahavas, I. (2011). On the stratification of multi-label data. In *Machine Learning and Knowledge Discovery in Databases: European Conference, ECML PKDD 2011, Athens, Greece, September 5-9, 2011, Proceedings, Part III 22*, pages 145–158. Springer.

Stoddart, D. (1968). The conservation of aldabra. *The Geographical Journal*, 134(4):471–485.

Stoykov, E. (2022). Geodetic basis when performing marine hydrographic measurements: Geodetic basis when performing marine hydrographic measurements. *Journal scientific and applied research*, 23(1):26–32.

Tessier, E., Bigot, L., Cadet, C., Cauvin, B., Chabanet, P., Conand, C., Nicet, J.-B., and Quod, J.-P. (2008). Les récifs coralliens de la réunion en 2007: état des lieux et réseau de suivi. *Revue d'écologie*, 63(1-2):85–102.

Thamarai, M. and Aruna, S. (2023). Stressed coral reef identification using deep learning cnn techniques. *Journal of Electronic & Information Systems*, 5(2):1–9.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. (2017). Attention is all you need. *Advances in neural information processing systems*, 30.

Villon, S., Mouillot, D., Chaumont, M., Darling, E. S., Subsol, G., Claverie, T., and Villéger, S. (2018). A deep learning method for accurate and fast identification of coral reef fishes in underwater images. *Ecological informatics*, 48:238–244.

Wickel, J. (2015). Etat des récifs coralliens et des écosystèmes associés des outre-mer français en 2016. *Documentation Ifrecor*.

Wolf, T., Debut, L., Sanh, V., Chaumond, J., Delangue, C., Moi, A., Cistac, P., Rault, T., Louf, R., Funtowicz, M., et al. (2020). Transformers: State-of-the-art natural language processing. In *Proceedings of the 2020 conference on empirical methods in natural language processing: system demonstrations*, pages 38–45.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., and Luo, P. (2021). Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34:12077–12090.

Young, G., Dey, S., Rogers, A., and Exton, D. (2018). Correction: Cost and time-effective method for multi-scale measures of rugosity, fractal dimension, and vector dispersion from coral reef 3d models. *PLOS ONE*, 13:e0201847.

Chapter 4

From underwater to drone: a novel multi-scale knowledge distillation approach for coral reef monitoring



Figure 4.1: *Aerial drone view from Massif des Roses (Madagascar), showing clear lagoon waters with scattered coral patches and a traditional pirogue navigating the shallow reef area.*
Image credit: Sylvain Poulain, Mai 2023.

This chapter is an adapted version of the article originally published in *Ecological Informatics* (Contini et al., 2025a). It describes the first upscaling methodology developed during this thesis, transferring underwater multi-label predictions to aerial multi-label classification.

Context and objectives

Precise mapping coral reef ecosystems would ideally require approaches that combine precision at fine-scales with coverage at broader spatial extents. Traditional underwater surveys like Underwater Visual Census (UVC) or modern ones based on Autonomous Surface Vehicles (ASVs) or Unmanned Aerial Vehicles (UAVs) can capture sub-centimetre details of benthic communities, enabling the identification of coral morphotypes, associated fauna and habitats. However, their limited spatial coverage and high operational effort make them challenging for monitoring large reef areas. Conversely, drone-based aerial imagery provides broad spatial coverage with high-resolution orthophotos but lacks the detail necessary to identify individual benthic organisms or fine-scale habitat structures.

This chapter presents a novel multi-scale approach to coral reef monitoring, integrating fine-scale underwater imagery with medium-scale aerial imagery. Underwater images are captured using an ASV, while aerial images are acquired with an UAV. A transformer-based Deep Learning (DL) model is trained on underwater images to detect the presence of 31 classes covering various coral morphotypes, associated fauna and habitats. These predictions serve as annotations for training a second model, which is applied to orthophotos generated from aerial images captured by UAVs. The transfer of information across scales is achieved through a weighted footprint method that accounts for partial overlaps between underwater image footprints and aerial image tiles.

This chapter shows that the integration of underwater and aerial imagery, supported by DL models, can facilitate scalable and accurate reef assessments. By leveraging multi-scale imaging and Artificial Intelligence (AI), our approach aims for the precision of fine-scale analysis while extending coverage to broader reef areas, providing valuable information for coral reef monitoring and conservation.

4.1 Introduction

Monitoring of coral reef ecosystems at a detailed taxonomic level and at large spatial scales remains a significant ecological challenge. While close-range imaging methods such as UVC or ASV-based surveys can provide high-resolution data on coral morphotypes, associated fauna and habitat types (Misiuk and Brown, 2024), they are limited in spatial coverage and scalability. This approach enables fine-grained annotations and predictions, allowing the distinction of coral morphotypes and the identification of specific classes, such as algae, that are often difficult to discern. The main drawback of survey methods based solely on ASVs is that, since underwater images are taken at a very fine-scale, it is difficult to cover large areas of coral reefs. This implies that, when areas on the order of tens of hectares are to be monitored, the method poses challenges in terms of cost, processing time and ease of deployment.

Recent advances in imaging technologies have opened up new possibilities for large-scale coral reef monitoring. Drone-based imaging has emerged as a valuable tool for coastal habitat mapping and monitoring, providing a cost-effective method for high-resolution habitat classification when combined with Machine Learning (ML) techniques (Doukari and Topouzelis, 2022). The main limitation of using only aerial images is that they do not directly provide detailed information on individual benthic organisms. Thus, the annotation of aerial images is usually made on broader classes (e.g., hard bottom, mixed substrate, soft bottom and seagrass). Here we argue that they contain sufficient information to infer the benthic community as a whole if they are combined with fine-grained predictions inferred from underwater images.

To give an illustration of this problem, Figure 4.2 shows an orthophoto obtained from aerial images on the left and an underwater image on the right, collected by an ASV, corresponding to the area highlighted by the red rectangle. The medium-scale image on the left shows a complex assemblage of corals. However, due to the resolution of the image, although one can guess the presence of several morphotypes of corals, identifying them individually remains a difficult task. This challenge becomes significantly easier when the corresponding underwater images are available. The underwater image on the right provides clear distinctions between coral assemblages and other classes, such as *Algae*, which are often difficult to observe in aerial imagery. Propagating those fine-grained annotations to the aerial image yields a finer classification of the benthic habitat.

Integrating aerial imagery with human-collected ground truth data can be a first solution to map coastal habitats with high accuracy, as demonstrated by Øie Kvile et al. (2024). The authors of this study provide a detailed protocol, from drone imagery collection to orthophoto annotation through Geographic Information System (GIS) softwares, allowing the training of segmentation Convolutional Neural Network (CNN) models on aerial images. In Ventura et al. (2023), the authors use both aerial and underwater imagery, processed with Support Vector Machines (SVM) and Object-Based Image Analysis (OBIA) for benthic habitat classification. They used UAVs to capture high-resolution aerial imagery of coastal areas and Unmanned Surface Vehicles (USVs) to collect ultra-high-resolution underwater images. The data from both platforms were

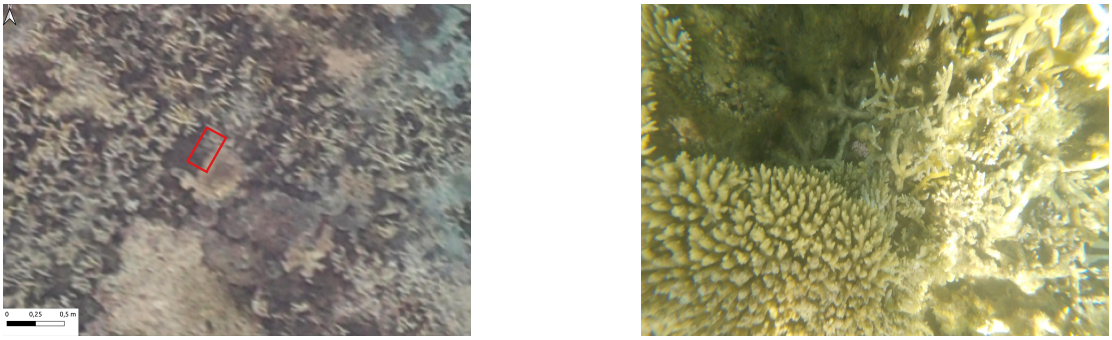


Figure 4.2: *On the left, an aerial orthophoto captures a complex coral assemblage. However, distinguishing between specific morphotypes is challenging because of the limited resolution. The red rectangle highlights the location of the fine-scale underwater image shown on the right. The underwater image provides a higher level of detail, allowing the identification of distinct coral morphotypes and specific classes, such as Algae, that are difficult to identify in aerial imagery.*

processed separately using Structure from Motion (SfM) to create orthophoto mosaics and Digital Surface Models (DSMs). These products were then classified using OBIA and SVM algorithms. The fine-scale data from the USVs improved the classification, particularly in underwater regions where the aerial data lacked detail, allowing a finer level of classification than would have been possible with aerial data alone. However, their approach does not fully leverage the complementarity of the two data sources. Indeed, there is an opportunity to exploit areas where both types of data (fine-scale underwater and medium-scale aerial) are available, in order to improve the model. In particular, they do not employ data from the same zones to train models that can infer fine-scale details from medium-scale data in regions where only aerial imagery is available. This limitation precludes the potential for synergistic use of overlapping datasets to enhance benthic habitat classification on a broader scale.

This chapter introduces a novel multi-scale DL approach that integrates underwater and aerial imagery for fine-grained assessment of coral reefs at broad spatial scales.

To extend this classification to aerial images we employ the concept of knowledge distillation (Gou et al., 2021; Wang and Yoon, 2021), where the underwater model *DinoVdeau* introduced in the Chapter 3 acts as the "teacher" and the aerial model as the "student". The objective is for the student model to learn from teacher's outputs, allowing it to achieve comparable performance by mimicking the teacher's knowledge.

Concretely, as illustrated in Figure 4.3, a first fine-scale model (the teacher) is trained on the underwater images associated with fine-grained manual annotations. A second model is then trained on the aerial images, using the underwater predictions from the teacher model and image metadata to generate annotations. The transfer of information across the two scales is achieved through a weighted footprint method that accounts for partial overlaps between underwater image footprints and aerial image tiles.

By reducing the time-consuming annotation process to a single step on underwater

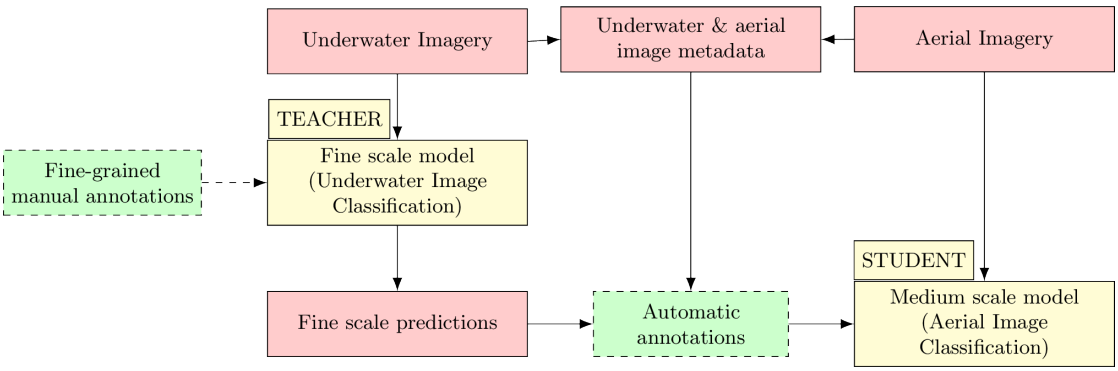


Figure 4.3: *Workflow of the multi-scale approach for coral reef monitoring.*

images, this approach allows the aerial model to classify images at a larger scale while maintaining as much as possible the fine-scale information provided by the underwater model.

Despite the UAV dataset being limited to Réunion Island, it captures key features of Indian Ocean reef ecosystems, making it suitable for demonstrating our methodology. A central advantage of our approach is the ability of the aerial student model to inherit knowledge from the fine-scale teacher model. As a result, when UAV data from additional islands in the Indian Ocean become available, the student model will seamlessly adapt, leveraging the pre-existing knowledge encoded by the teacher model.

This research offers a powerful tool for coral reef monitoring through an accurate classification of coral morphotypes and associated marine organisms. Compared to traditional ASV surveying techniques, this method provides significant advantages in terms of cost efficiency, as it reduces the human time required for ASV deployments to cover the same surface area. UAVs are also easy to deploy from the shore, allowing access to sites located several kilometres away without losing control of the drone, depending on the local aviation regulations (see Figure 4.1[1]), showing an aerial image of the *Massif de Roses* in Madagascar located 2 km offshore). Furthermore, this approach eliminates the risk of sending ASVs into dangerous zones where they could become stuck or damaged, ensuring safer operations and minimal disturbance to the reef ecosystem. By demonstrating the combination of advanced technologies such as ASVs, drones and DL models, the study contributes to the development of more effective and efficient conservation strategies (Ullah et al., 2024), highlighting the potential of multi-scale monitoring as a robust tool for environmental protection.

To our knowledge, this is the first time that such a cascade of DL models has been used to classify aerial images. This opens up new possibilities for upscaling a computer vision model trained on fine-scale images to a larger scale. Although being developed for marine applications, this method could be used for terrestrial ecosystems where images can be collected at different scales and altitudes. Further studies could consider

[1]The image can be found in Poulain et al. (2023) under the relative path:
`20231202_MDG-massif-des-roses_UAV-02_1/DCIM/DJI_0352.JPG`

simultaneous underwater and aerial data collection. This would allow the monitoring of slow-moving species, such as sea cucumbers, which can be observed in aerial imagery but require temporally aligned datasets for precise identification.

4.2 Materials and methods

4.2.1 Underwater image acquisition

Underwater images were collected using the *Plancha* ASV described in Chapter 3. 50 missions were carried out in the lagoon of Réunion Island: 30 in the Saint-Leu lagoon and 20 in the *Trou d'eau* lagoon.

4.2.2 Aerial image acquisition

Aerial drone images were taken with a `DJI Mavic 2 Pro` drone. Images were collected following good practices in use in aerial imagery as outlined by Slocum et al. (2019).

Since this drone is not equipped with a differential GPS, once images were taken and the SfM model was built, the default orthophoto georeferencing has been improved by collecting Ground Control Points (GCPs) using a differential GPS.

To obtain a high-resolution orthophoto with high positioning precision, the following steps were taken:

1. Mission planning: check the equipment, request authorizations from French authorities, check weather conditions and plan the flight mission.
2. Mission execution: fly the drone at an altitude ensuring a Ground Sampling Distance (GSD) below 2 cm, while adapting camera settings to the specific conditions of the day.
3. Image processing: build the SfM model using images taken during the flight mission. This was done using `OpenDroneMap` software.
4. GCPs collection: collect GCPs using a differential GPS with centimetric accuracy.
5. Orthophoto georeferencing: refine the georeferencing of the orthophoto using GCPs.

Two missions were carried out in the lagoon of Réunion Island: one in the *Saint-Leu* lagoon and the other in the *Trou d'eau* lagoon. Flight plans of UAV surveys were set in the `Pix4D Capture` application with an 80% overlap for both frontlap and sidelap, a flight altitude between 40 m and 60 m and an image resolution of 5,472 px × 3,648 px, resulting in a GSD ranging from 0.9 cm px^{-1} to 1.6 cm px^{-1}. Settings used for processing the images using the `OpenDroneMap` software package are shown in Table 4.1.

For further details on mission planning, execution and image processing, please refer to Appendix 3, where these aspects are explained in detail.

Table 4.1: *OpenDroneMap parameter settings*

Setting	**Value**
Auto-boundary	True
DEM Resolution	2.0
DSM	True
Orthophoto Resolution	1.0
Point Cloud Quality	Ultra
Rolling Shutter	True

Orthophoto georeferencing

Since the drone was not equipped with a differential GPS, in order to improve the native georeferencing of the orthophoto, GCPs were chosen on fixed and easily distinguishable objects on land (e.g. manhole covers, corners of basketball courts, etc.) and easy-to-distinguish corals (e.g., large *Porites* or *Acropora Tabular* corals). GCPs positions were collected using a GPS with centimetre-level accuracy and the orthophoto was subsequently reconstructed by aligning the pixel representing GCPs to be at the ground truth collected position. GCPs examples are shown in Figure 4.4 with a pink point on the image.



Figure 4.4: *Examples of GCPs collected in Réunion Island in order to correct the aerial orthophoto position. Both human-made objects (on the left) and easily distinguishable corals (a Porites coral on the right) can be used to set GCPs.*

The *Trou d'eau* lagoon mission measured a surface of 189,682 m^2 and 10 GCP were collected. The *Saint-Leu* lagoon mission covered an area of 204,748 m^2 and 9 GCP were collected.

Final orthophoto positioning was checked by computing the Root Mean Square Error (RMSE) between the GCPs and the orthophoto.

In Table 4.2 we show the error statistics for GCPs collected in the *Trou d'eau* and the *Saint-Leu* lagoons. A RMSE of approximately 1 cm on the xy plane ensures adequate

Table 4.2: *GCP errors statistics for Trou d'eau Lagoon and Saint-Leu Lagoon*

Lagoon	Error Type	Mean	Standard Deviation	RMS Error
Trou d'eau	X Error (meters)	0.001	0.011	0.011
	Y Error (meters)	-0.001	0.010	0.010
	Z Error (meters)	-0.011	0.028	0.030
	Total			**0.022**
Saint-Leu	X Error (meters)	-0.001	0.003	0.003
	Y Error (meters)	-0.000	0.001	0.001
	Z Error (meters)	0.000	0.004	0.004
	Total			**0.003**

positional precision for subsequent analyses.

4.2.3 Multi-scale positioning

Since the objective is to pass information from a fine-scale (underwater images) to a larger scale (drone images), the precision of the relative position between underwater and drone images is crucial. Indeed, the benthic substrate can vary significantly across small distances, meaning that any geolocation error in the underwater images would result in annotations from the underwater teacher model being transferred to an incorrect tile in the aerial student image. Similarly, an error in the georeferencing of the aerial tile position would cause the teacher annotations to be assigned to the wrong tile during the matching process. These potential mismatches emphasize the importance of achieving high positioning accuracy for both underwater and aerial images to ensure reliable multi-scale data integration. To ensure consistency in the georeferencing process, all spatial data were referenced using the `EPSG:32740` coordinate reference system (WGS 84 / UTM zone 40S).

In order to validate the geolocation of underwater images with respect to aerial images, we used data from French National Institute of Geographic and Forest Information (IGN)[2]. Each 3 to 4 years, IGN produces *BD Ortho*: a collection of orthophotos with a default resolution of 20 cm. The last orthophoto produced by IGN on Réunion Island was in 2022, so we decided to use this data as a reference to validate the georeferencing of our data.

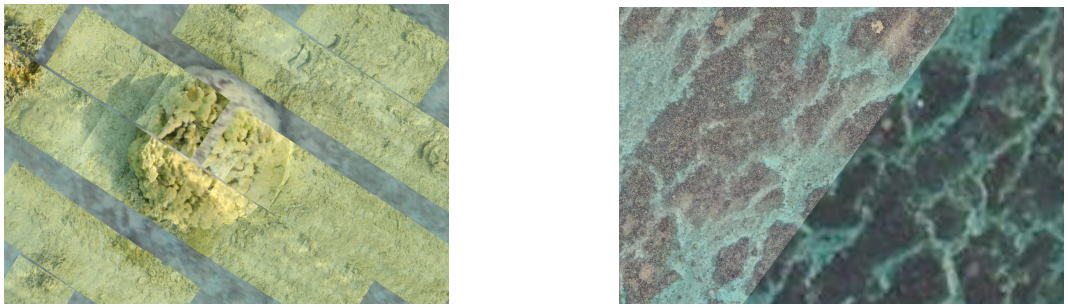
Two visual criteria were then used to confirm the georeferencing of the underwater and aerial images with respect to the *BD Ortho* orthophoto:

- Relative georeferencing: find the presence of easily recognizable objects in both underwater and aerial orthophoto and compare them in a GIS software. Often the presence of *Porites* corals can be used to compare the two scales since their contours are easily recognizable in both types of images. See Figure 4.5a.
- Aerial absolute georeferencing: find the presence of easily recognizable coral

[2]IGN is the French public state administrative establishment whose main objective is to produce and maintain geographical information for France and its overseas territories

colonies in both aerial and *BD Ortho* orthophoto and compare them in a GIS software. See Figure 4.5b.

The combination of the relative georeferencing between underwater and aerial images and the aerial absolute georeferencing with respect to the *BD Ortho* orthophoto allowed us to crosscheck the geolocation of the underwater images with respect to national baseline data.



(a) ***(b)***

Figure 4.5: *Visual georeferencing criteria to validate the geolocation of underwater and aerial images with respect to the BD Ortho orthophoto. On the left (a) underwater images geolocated with respect to the aerial orthophoto. On the right (b) aerial orthophoto georeferenced with respect to the BD Ortho orthophoto produced by the IGN. The lighter part on the left corresponds to the drone-based orthophoto and the darker part on the right corresponds to the BD Ortho.*

4.2.4 Underwater image classification

Using *DinoVdeau* model described in Chapter 3, we ran inference on 56,653 geotagged ASV images from the *Trou d'eau* lagoon and 58,076 images from the *Saint-Leu* lagoon in Réunion Island which correspond to the areas overlapping with the aerial drone survey.

4.2.5 Upscaling predictions

The next key step is to correctly transfer the predictions from the underwater model to the aerial model. The objective is to train an aerial model based on underwater predictions, without requiring manual annotation of aerial images.

This is achieved by following the steps below:

1. Split aerial orthophoto into tiles, ensuring consistency in ground surface representation of each tile across different sessions. See Section 4.2.5.

2. Filter useless aerial tiles (e.g., black tiles issued from SfM processing errors or tiles without corresponding underwater images). See Section 4.2.5.

3. Associate each aerial tile with underwater images whose camera GPS position is within the tile boundaries. See Section 4.2.5.

4. Compute the footprint of each underwater image and filter aerial tiles with not enough underwater coverage. See Section 4.2.5.

5. Transform underwater predictions into aerial annotations. See Section 4.2.5.

Orthophoto tiling

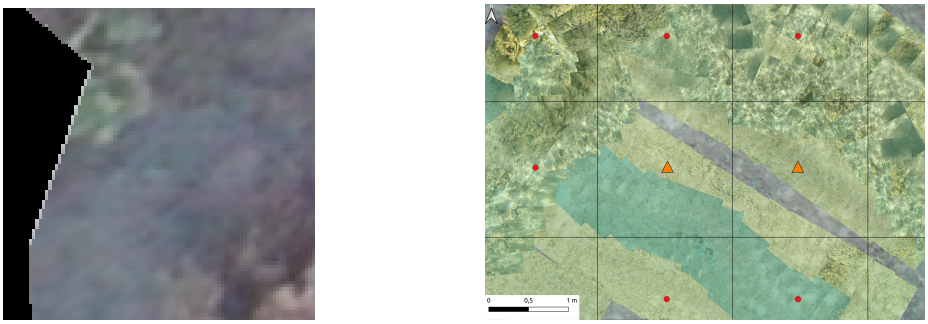
The first step is to split the aerial orthophoto into tiles. This process accounts for the GSD of each orthophoto. This approach guarantees that while tile dimensions in pixels may vary due to different GSDs, each tile consistently represents a fixed area on the ground. This method allows for standardized comparison and analysis of images across different datasets and sessions, maintaining a consistent spatial resolution. Splitting the orthophoto into too small tiles results in images without enough context to be correctly classified and/or with an insufficient resolution. On the contrary, splitting the orthophoto into too large tiles results in good classification performance but does not allow for a fine-grained analysis of the data. Searching for the best compromise, we fixed this area to be 1.5 m × 1.5 m.

Useless tile filtering

The second step is to filter out useless tiles. This is done by removing tiles with a high percentage of black pixels (due to errors in SfM processing) and tiles with no corresponding underwater images. Examples of such tiles are shown in Figure 4.6. In Figure 4.6a we can see an example of a tile extracted from the aerial orthophoto of the *Saint-Leu* lagoon in Réunion Island with a high percentage of black pixels.

Footprint calculation and tile coverage assessment

The third step involves associating underwater predictions with aerial tiles. This assignment is achieved by identifying underwater images whose camera position centre falls within the boundaries of the aerial tiles. After associating underwater images to aerial tiles, the next step is to compute the footprint of each underwater image to filter out aerial tiles with not enough underwater coverage. In Figure 4.7, we outline the process to calculate the footprint of underwater images based on data from ASV sensors. Using bathymetric data from the echosounder, we measure the distance between the camera and the seabed, which determines the scale of the area captured in each image. The camera orientation in the XYZ axis plane is defined by the roll, pitch and yaw angles, which determine how the Field of View (FOV) is directed relative to the seafloor. Finally, the FOV, divided into horizontal (FOV_h) and vertical (FOV_v) angles, defines the area visible to the camera. By projecting these angles down to the seafloor, we calculate the intersection points, forming a polygonal footprint that represents the region covered by the image. This footprint is necessary to associate each underwater image with a



(a) *(b)*

Figure 4.6: *Examples of useless tiles extracted from the aerial orthophoto of the Saint-Leu lagoon in Réunion Island: (a) Example of a tile extracted from the aerial orthophoto of the Saint-Leu lagoon in Réunion Island, with a high percentage of black pixels (b) Example of a group of tiles extracted from the aerial orthophoto of the Saint-Leu lagoon in Réunion Island, with corresponding underwater images. The tiles in the middle do not have enough coverage of underwater images*

specific area of the seafloor. Merging the footprints of all underwater images associated with a tile, we obtain the union of the footprints, which represents the area covered by the underwater images associated with the tile [3].

This allows us to filter out tiles that do not have enough underwater coverage. An example of such a tile is shown in Figure 4.6b, where a group of tiles extracted from the same orthophoto is shown. Tiles whose centre is represented by a red point are classified as useful, since they are completely covered by underwater images. On the contrary, tiles represented by an orange triangle are classified as useless, since they do not have enough coverage of underwater images.

Transforming underwater predictions into aerial annotations

The fifth step is to transform underwater predictions into aerial annotations. The trivial approach would be to associate the presence of a class c in a tile t if at least one underwater image associated with the tile is predicted as belonging to the class c by the teacher model. In other words, if class c is not predicted as being absent on all underwater images associated with tile t, which can be formulated as:

$$\forall c \in \mathcal{C},\ \forall t \in \mathcal{T}, \quad I(y_c = 1 \,|\, t) = 1 - \prod_{x \in \mathcal{X}(t)} \left[1 - h_c^{\text{teacher}}(x)\right] \tag{4.1}$$

where :

- $\mathcal{C}$ is the set of classes described in Section 4.2.5
- $\mathcal{T}$ is the set of tiles

[3] More information about footprint calculation can be found on drone-upscaling GitHub

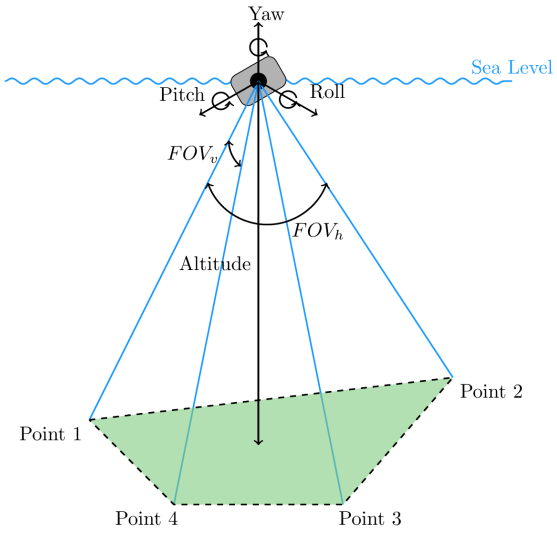


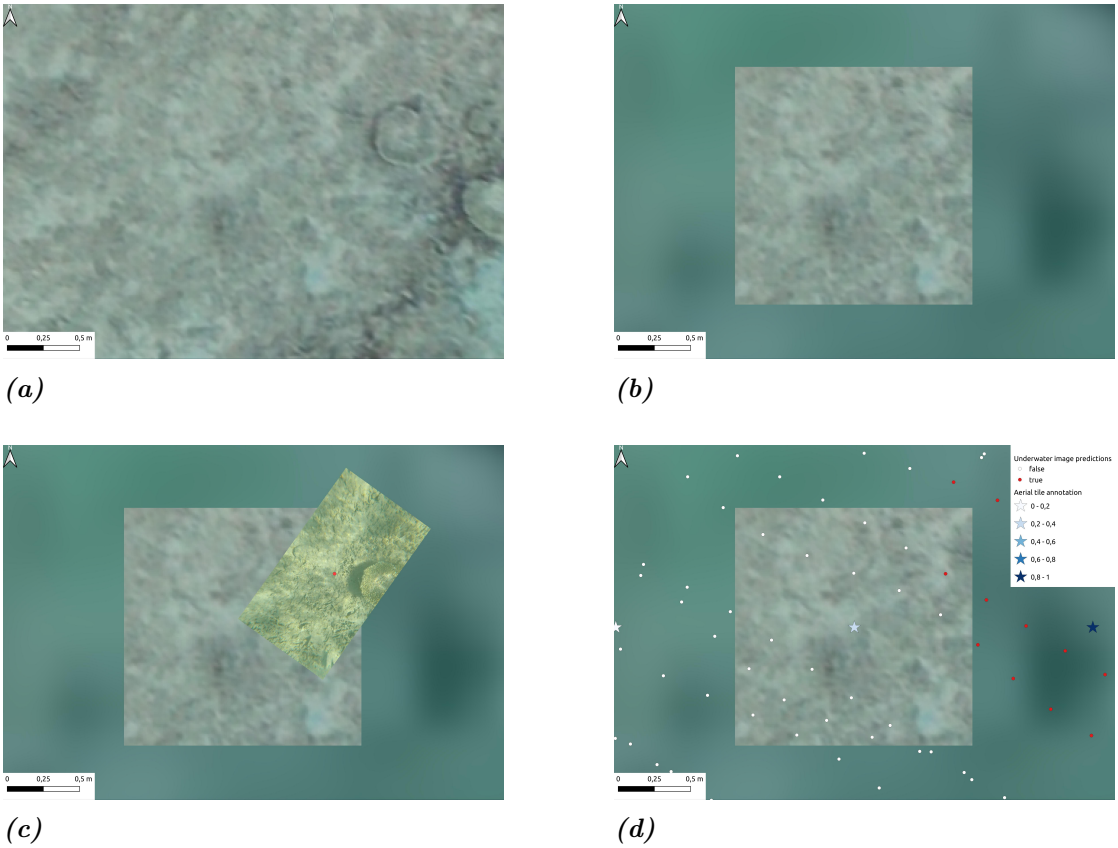
Figure 4.7: *Footprint calculation of underwater images based on echosounder data, camera field of view and ASV angles.*

- $y_c \in \{0; 1\}$ is the binary label associated with the presence/absence of class c
- $I(y_c = 1 \,|\, t) \in \{0; 1\}$ is a binary function indicating the presence or absence of class c in tile t.
- $\mathcal{X}(t)$ is the set of underwater images associated with tile t
- $h_c^{\text{teacher}}(x) \in \{0; 1\}$ is the binary prediction associated with the presence/absence of class c in underwater image x

The drawback of this approach is that it does not consider the footprint of underwater images, tending to overestimate the presence probability of a class in a tile.

A more realistic approach, since not all underwater image footprints fall entirely within the boundaries of a specific tile, requires computing the intersection between the underwater image footprint and the corresponding tile. This allows us to give more weight to underwater images that are completely within a tile and less weight to underwater images that are only partially within a tile.

The orthophoto in Figure 4.8 gives an example with the corresponding predictions on underwater images.



(a) *(b)*

(c) *(d)*

Figure 4.8: *Example of the upscaling process from underwater predictions to aerial annotations, depending on the intersection between underwater images footprint and drone tiles, the probability of presence of a class in a tile can be mitigated: (a) Aerial orthophoto of the Trou d'eau lagoon in Réunion Island (in the right part of the image a colony of Acropora Tabular corals is visible), (b) Drone tile extracted from the aerial orthophoto in Figure 4.8a (the Acropora Tabular corals are not visible in the tile), (c) Superposition of the underwater image in the top right corner of the tile and the drone tile (the Acropora Tabular coral colony is outside the drone tile), (d) Underwater predictions associated with the tile in Figure 4.8b (the Acropora Tabular corals is predicted on the top right underwater image of the tile, but weakly predicted on the tile)*

In the right part of Figure 4.8a, a colony of *Acropora Tabular* corals is visible. Proceeding with tile extraction from the orthophoto, we obtain the tile in Figure 4.8b. Since the *Acropora Tabular* corals do not fall within the tile, we would like that, after computing the tile annotation starting from underwater predictions, the probability for the class *Acropora Tabular* associated with this tile will be low. Unfortunately, it may happen that underwater images that have the centre within the tile (but not all the footprint) include classes that are outside the tile bounds: as shown in Figure 4.8c, where a part of the *Acropora Tabular* coral colony is visible in the right part of the underwater image.

In these cases, weighting underwater predictions based on the intersection between the underwater image footprint and the tile allows us to reduce the impact of these images on the final aerial annotations. This is shown in Figure 4.8d where predictions on underwater images are represented with circles on a red ramp and aerial annotations are represented with stars on a blue ramp. Even if in the underwater image in Figure 4.8c on the right of the tile the presence of the *Acropora Tabular* class is predicted, the overlap between the underwater image and the aerial tile is weak. Consequently, the probability of presence of the *Acropora Tabular* class on the tile is mitigated: ending up with an annotation of 0.4 (while the blue star on the tile just on the right indicates a probability of presence of 0.98).

To take account of the intersection between the underwater image footprint and the tile bounds, we can consider that the probability of presence is proportional to the intersection area relative to the underwater image area. Therefore, we can modify Equation 4.1 as follows:

$$\forall c \in \mathcal{C},\ \forall t \in \mathcal{T}, \quad P(y_c = 1 \,|\, t) = 1 - \prod_{x \in \mathcal{X}(t)} \left[1 - \frac{s(x \cap t)}{s(x)}\, h_c^{\text{teacher}}(x) \right] \tag{4.2}$$

where:

- $P(y_c = 1 \,|\, t) \in [0, 1]$ is the probability of presence of class c in tile t
- $s(x)$ is the area of the underwater image x
- $s(x \cap t)$ is the area of intersection between the underwater image x and the tile t

The product over all underwater images gives the probability that class c is absent in all underwater images associated with the tile t.

To get a better estimation of the presence probability, we can also take into account the confidence of the teacher model. Therefore, we can replace the binary output of the classifier $h_c^{\text{teacher}}(x)$ by the probabilistic output $p_{\text{teacher}}(y_c = 1 \,|\, x)$ leading to:

$$\forall c \in \mathcal{C},\ \forall t \in \mathcal{T}, \quad P(y_c = 1 \,|\, t) = 1 - \prod_{x \in \mathcal{X}(t)} \left[1 - \frac{s(x \cap t)}{s(x)}\, p_{\text{teacher}}(y_c = 1 \,|\, x) \right] \tag{4.3}$$

where:

- $p_{\text{teacher}}(y_c = 1 \,|\, x) \in [0, 1]$ is the probabilistic output of the teacher model for class c in the underwater image x, obtained through a sigmoid function on top of the final layer of the model. If the model is trained with the binary Cross-Entropy (CE) loss function (as in our experiments), the output is asymptotically converging to the true conditional probability that class c is present in image x (Lorieul, 2020).

It is worth noting that when the probability $p_{\text{teacher}}(y_c = 1 \,|\, x)$ is equal to 1, then Equation 4.3 is equivalent to Equation 4.2. But in the general case, it is comprised in the

interval $]0, 1[$. In the literature related to knowledge distillation (Gou et al., 2021), such probabilistic labels passed to the student model are often called *soft labels* in opposition to hard labels such as the one in Equation 4.1. Training the student model on soft labels rather than hard labels enables a better transfer of information from the teacher to the student. The probability of presence captures valuable information on the uncertainty of the teacher model and allows us to place less weight on ambiguous cases in the loss function of the student.

Aerial dataset

Following the upscaling process detailed in Section 4.2.5, starting from two aerial orthophotos of the *Trou d'eau* and *Saint-Leu* lagoons in Réunion Island measuring 189,682 m^2 and 204,748 m^2 respectively, we ended up with 4,911 and 6,832 annotated tiles respectively for a total of 11,743 annotated tiles.

Since the upscaling process implies a loss in the image resolution, we made some changes to the classes to be predicted:

1. The first change was to merge *Algae* classes into a single class called *Algae*, indeed distinguishing between the different types of algae (*Algal Assemblage*, *Algae Halimeda*, *Algae Coralline* and *Algae Turf*) is a task that requires a higher resolution than the one we have[4].

2. The second change was to remove underwater classes that do not have a corresponding aerial class: *Blurred* images (an underwater blurred image does not imply a blurred aerial image) and *Homo Sapiens* (since human body parts in underwater images do not imply human body parts in aerial images).

3. The third change was to remove underwater classes that are not relevant for the aerial images, i.e. *Fish*, *Sea cucumber* and *Sea urchin*. The first two classes, even if visible in some aerial images, were removed because underwater and aerial images are not taken at the same time, so that the presence of a sea cucumber or a fish in an underwater image does not imply the presence of those organisms in the corresponding aerial image (see Section 4.4.3). The last one was removed since those organisms are not visible at all in aerial images.

Finally, we retained only classes for which there was a sufficient number of annotations.

[4]In the case of Equation 4.2 this was done by assigning $h^{\text{teacher}}_{Algae}(x) = 1$ if at least one type of algae (*Algal Assemblage*, *Algae Halimeda*, *Algae Coralline* and *Algae Turf*) was predicted as present on the fine-scale image x, otherwise $h^{\text{teacher}}_{Algae}(x) = 0$. In the case of Equation 4.3 this was done by assigning to $p_{\text{teacher}}(y_{Algae} = 1 \,|\, x)$ the maximum between all the probabilities predicted by the underwater model for algae classes (*Algal Assemblage*, *Algae Halimeda*, *Algae Coralline* and *Algae Turf*).

Thus, after removing classes that have less than 200 annotations in the aerial dataset, we ended up with 12 classes:

- Coral
  1. *Acropora Branching*
  2. *Acropora Digitate*
  3. *Acropora Tabular*
  4. *Dead coral*
  5. *Non-acropora Encrusting*
  6. *Non-acropora Massive*
  7. *Non-acropora Millepora*
  8. *Non-acropora Submassive*
- Habitat
  1. *Rock*
  2. *Rubble*
  3. *Sand*
- Other Organisms
  1. *Algae*

Aerial deep learning model (student model)

To train the student model with soft labels, we used the Binary Cross-Entropy (BCE) loss function. This loss measures the divergence between the predicted probabilities of the student model and the soft labels $P(y_c = 1 \mid t) \in [0, 1]$ generated by the teacher model. Specifically, the loss for class c in tile t is given by:

$$\mathcal{L}_{\text{BCE}}(t, c) = -\Big[P(y_c = 1 \mid t) \cdot \log(p_{\text{student}}(y_c = 1 \mid t)) + (1 - P(y_c = 1 \mid t)) \cdot \log(1 - p_{\text{student}}(y_c = 1 \mid t))\Big] \tag{4.4}$$

where:

- $P(y_c = 1 \mid t) \in [0, 1]$ is the soft label provided by the teacher model for class c in tile t, as described in Equations 4.2 and 4.3
- $p_{\text{student}}(y_c = 1 \mid t) \in [0, 1]$ is the probabilistic output of the student model for class c in tile t, obtained through a sigmoid function on top of the final layer of the model.

To maintain consistency with underwater predictions, we used the same architecture for the student model as the one used for the teacher model (see Section 3.2.3). The only difference is that, since the underwater model was trained with binary values and the aerial model is trained on probabilities, when computing evaluation metrics during the training process we cannot directly use accuracy, precision, recall or F1-score metrics. Instead, we computed the RMSE, the Mean Absolute Error (MAE) and the Kullback-Leibler (KL) divergence metrics. The student model can be found here.

4.2.6 Test zone and model evaluation

To evaluate the performance of the aerial DL model, we selected a test zone within the *Trou d'eau* lagoon, as shown in Figure 4.9. This area was chosen due to its diverse composition of coral morphotypes, habitats and other marine organisms, which represents a challenging environment for model validation. The test zone comprises 194 underwater images, corresponding to 28 aerial tiles, for a total of 63 m^2.

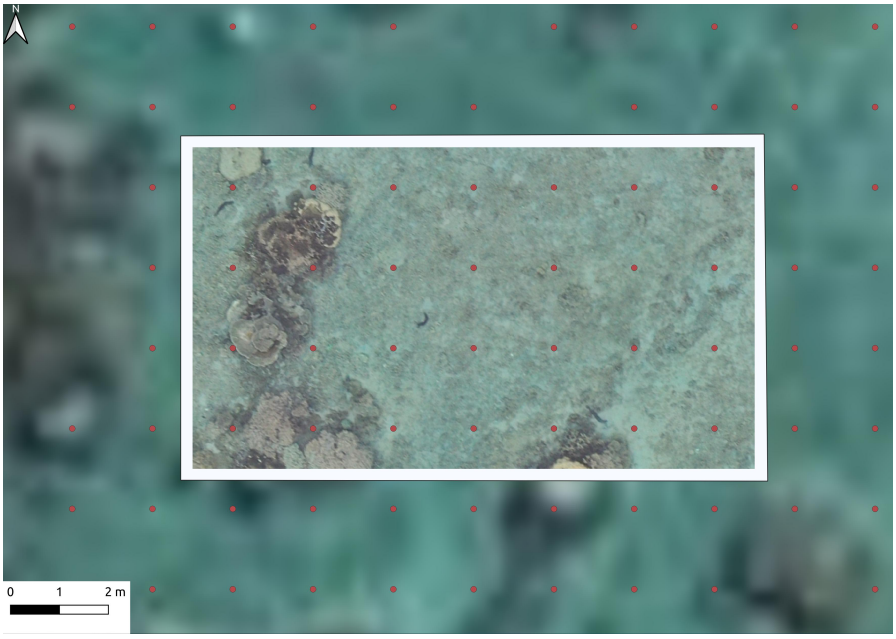


Figure 4.9: *Test zone within the Trou d'eau lagoon, selected for model evaluation. The area measures* 63 m^2 *and comprises 194 underwater images corresponding to 28 aerial tiles.*

The annotation process for the aerial tiles is carried out as follows:

1. For each aerial tile, underwater images with centroids located within the tile boundaries are identified
2. These underwater images are then projected in QGIS to assess the portions of each image that intersect the aerial tile boundaries.
3. Each cropped underwater image is manually annotated with fine-grained precision.
4. As a result, each aerial tile is annotated with a level of detail comparable to that of underwater imagery and is therefore considered as ground truth data.

These aerial tiles were not used during model training, ensuring an unbiased evaluation of the aerial model.

Both equations 4.2 and 4.3 were used to generate aerial annotations starting from underwater predictions in the test zone. The generated annotations were then compared with ground truth data to evaluate the quality of the upscaling process using the Area Under the Curve (AUC) metric, which is a commonly used metric in the evaluation of Species Distribution Models (SDMs) (Elith* et al., 2006).

Finally, in order to evaluate the aerial model, we compared the predictions on the test zone with the ground truth data using the AUC metric.

4.3 Results

4.3.1 Upscaling process evaluation

In order to evaluate the upscaling process, we compared the generated annotations with ground truth annotations in the geospatial test zone in the *Trou d'eau* lagoon. We first evaluated the quality of the soft labels generated by our methods described in Equations 4.2 and 4.3. As they provide a presence probability for each class, we can actually measure their AUC on the ground truth annotations. With a value of 0.9211 for annotations generated through Equation 4.2 and 0.9251 for annotations generated through Equation 4.3, both methods show a high level of accuracy in transferring information across scales.

To further evaluate both methods, we then measured the performance of the student model trained with either method. In the following, we will call `Model_spatial_only` the model trained from annotations generated through Equation 4.2 and `Model_distilled` the model trained from the annotations generated through Equation 4.3. Only the second model integrates information about the teacher's model confidence (= knowledge distillation). The first model integrates the hard labels predicted by the teacher and the spatial coverage. We first looked at the evaluation metrics measured on the soft labels themselves (using the random test set). The results are shown in Table 4.3.

Table 4.3: *Comparison of* `Model_spatial_only` *and* `Model_distilled` *on various performance metrics*

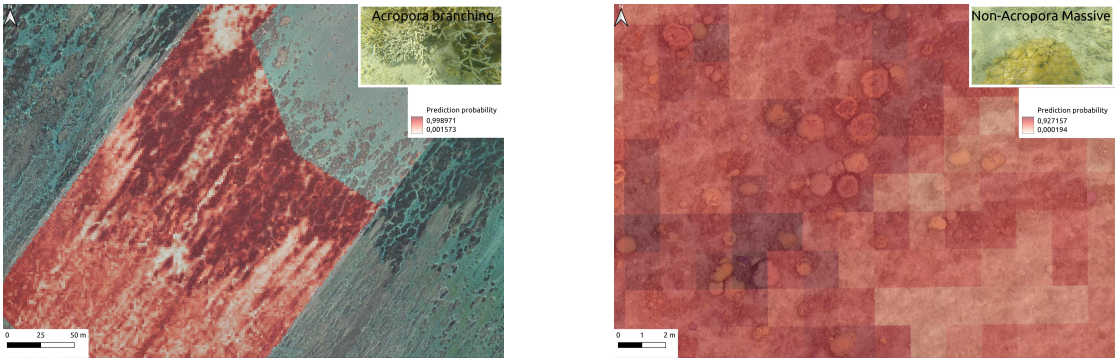
Model	RMSE	MAE	KL Divergence
`Model_spatial_only`	0.2019	0.1446	0.9802
`Model_distilled`	**0.1546**	**0.1143**	**0.3931**

The results show that `Model_distilled` trained using knowledge distillation (i.e. with Equation 4.3) allows a better prediction of the soft labels than `Model_spatial_only` trained without knowledge distillation (i.e. Equation 4.2) on all metrics. This means that the information they contain is more predictable from the aerial image contents.

Finally, comparing the predictions generated with both `Model_spatial_only` and `Model_distilled` on the ground truth data of the geospatial test zone in the *Trou d'eau* lagoon, we obtain an AUC of 0.7753 and 0.7952 respectively. This confirms that the knowledge-distillation approach (Equation 4.3) yields the most accurate upscaling, enabling the aerial model to achieve the highest AUC values.

4.3.2 Prediction maps

Once the aerial (student) model is trained, we use it in inference mode to generate high-resolution maps of large areas. In particular, we ran it on 20,027 tiles in the *Trou d'eau* lagoon and 61,059 tiles in the *Saint-Leu* lagoon. For each tile, we used the output of the student model as the probability of presence of each class and then prediction maps were generated for each class.



(a) *(b)*

Figure 4.10: *Prediction maps generated by the aerial model superposed on the aerial orthophoto: (a) Prediction map for the Acropora Branching class in the Trou d'eau lagoon. Red patches indicate the presence of corals. (b) Zoomed-in prediction map for the Non-acropora Massive class in the Saint-Leu lagoon. Red patches indicate the presence of corals.*

In Figure 4.10 we show two examples of prediction maps generated by the aerial model for the *Acropora Branching* class in the *Trou d'eau* lagoon and the *Non-acropora Massive* class in the *Saint-Leu* lagoon. The granularity of the prediction raster is fixed at 1.5 m × 1.5 m, which is the same as the aerial tiles. Different spatial scales are shown in Figure 4.10a and Figure 4.10b, in order to highlight the model's ability to cover large areas while still being able to capture fine-scale details.

4.4 Discussion

This study demonstrates the potential of combining underwater and aerial imagery to improve the monitoring and management of coral reef ecosystems. Although previous research has highlighted the advantages of using both imaging techniques, this work is the first, to our knowledge, to integrate AI models across multiple spatial scales. By using high-resolution underwater AI predictions to train a larger-scale aerial model, we ensure the precision of underwater analysis while extending it to cover a broader reef area. This multi-scale approach has the potential to advance marine monitoring and it can also be applied to other fields such as agriculture, forestry, urban planning and so forth.

4.4.1 Transferring information across scales: model independence and flexibility

Upscaling process evaluation

While having a well-performing fine-scale model is needed for reliable medium-scale annotations, working on the model architecture in order to gain a few percentage points of accuracy is not the core of our methodology. The primary objective of our workflow is to train a medium-scale student model to mimic the behaviour of a fine-scale teacher model, without the need to reannotate medium-scale images.

Two different techniques were used to generate aerial annotations starting from underwater predictions: the first one was based on presence/absence values obtained by thresholding underwater predictions (Equation 4.2), while the second one was based on knowledge distillation, i.e., on integrating the probability values predicted by the underwater model in the soft labels passed to the student (Equation 4.3). To compare and validate these methodologies, we assessed underwater predictions against ground truth data in the test zone of the *Trou d'eau* lagoon. The distillation-based method appeared to be the best with a high value of the AUC metric (0.9251). This indicates that it is a reliable method for transferring information from fine-scale to medium-scale images, allowing for a more nuanced estimate of class presence compared to using binary predictions (Equation 4.2).

Indeed, the soft labels generated by the teacher model provide a more accurate representation of class presence compared to the hard targets, since they capture the probability distribution across all possible classes, offering continuous values between 0 and 1. This approach allows the student model to better learn from the teacher model's uncertainty, leading to improved generalization capabilities and performance (Hinton et al., 2015). This result is consistent with the performance of the aerial model trained with annotations generated by Equation 4.3, which outperformed the model trained with annotations generated by Equation 4.2 on all metrics.

A key advantage of this methodology is its model independence. If a more accurate or advanced model becomes available in the future (thanks to new annotated images or improved algorithms), our framework allows for easy adaptation. By applying the new model to our existing fine-scale data, we can regenerate high-resolution predictions, which can then be seamlessly transferred to the medium-scale. This adaptability ensures that the model continues to benefit from the most accurate fine-scale insights.

Aerial model evaluation

The model's performance on the neural network test set underscores its robustness. As shown in Table 4.3, the RMSE, MAE and KL divergence metrics are all low, indicating that the model's predictions closely match annotations on the test set. The relatively low MAE compared to the RMSE (0.1143 vs. 0.1546) suggests that predictions on average deviate from the true values by less than 12%, indicating a high level of accuracy. Occasional larger discrepancies make the RMSE slightly higher, but the overall results are still very promising. The high AUC value (0.7952) computed on the test zone further

confirms the model's strong performance, indicating that the model reliably estimates class probabilities consistent with the true label distributions.

With regard to possible improvements concerning the DL model, as previously mentioned, in this study we used DinoV2 as a backbone: one of the State Of The Art (SOTA) computer vision models that currently performs the best on benchmark datasets. Applying transfer learning, we take advantage of the model's strong generalization capabilities while fine-tuning it to address our specific problem. However, we recognize that with the rapid advancements in artificial intelligence, our model may already be on the path to obsolescence (Ma et al., 2024). Continuously updating the model predictions to train the medium-scale model would enable us to incorporate the latest breakthroughs in AI, ensuring increasingly refined annotation quality over time.

4.4.2 Georeferencing challenges in multi-scale monitoring

A framework that enables the transfer of information from fine-scale to broader-scale imagery relies essentially on achieving precise alignment between the two layers of data. As shown in Figure 4.8 the benthic substrate can vary significantly across small distances, so that in order to upscale underwater predictions to aerial annotations, data need to be accurately geolocated.

Using differential GPS technology is the first step to achieve precise positioning, but does not guarantee a centimetric accuracy. To improve the ASV positioning, we used Post-Processing Kinematic (PPK) techniques thanks to the CentipedeRTK network [5], resulting in centimetric ASV positioning accuracy. Unfortunately, the position of the ASV is not the same as the position of the image, since waves can change the attitude of the ASV by tilting the direction of the camera from the vertical axis (see Figure 4.7). To correct this, we used the camera angles on the three axes (Roll, Pitch and Yaw) and the echosounder data to compute the footprint of underwater images. Ending up with the latitude and longitude of the four corners defining the footprint on the seabed of each underwater image. A check on the quality of the geolocation of underwater images is then necessary in order to validate the image positioning accuracy with an unbiased approach. In our case we chose to compare our data with data produced by the IGN, which is a reference in the field of georeferencing. Thanks to a visual comparison between the two datasets, we were able to validate the accuracy of the fine-scale geolocation process.

Although using a differential GPS on the platform acquiring images is a significant advantage, it is neither a necessary nor a sufficient condition to achieve centimetric accuracy.

For instance, our methodology demonstrated that combining PPK techniques, GCPs and validation against an external reference such as IGN data can deliver the required accuracy for coral reef monitoring. More generally, strict georeferencing is not always required for successful multi-scale upscaling. What matters most is that all images are accurately aligned, either in an absolute geographic reference frame or in a consistent relative frame.

[5] https://docs.centipede.fr/

However, since our current setup involves UAVs without embedded differential GPS, precise aerial image alignment still depends on the manual collection of GCPs, making the process time-consuming and labour-intensive. Even if, for long-term monitoring of the same site, previously established GCPs corresponding to fixed and stable objects can be reused, transitioning to UAVs equipped with embedded differential GPS could significantly streamline the data collection process. By applying PPK techniques directly to aerial images, the need for GCP collection could be minimized to a few validation points, reducing mission planning time and increasing efficiency. This advancement would enable faster and more scalable reef monitoring over large areas while maintaining accurate positioning.

4.4.3 Expanding spatial coverage and species identification

Satellite imagery upscaling

This study highlights several areas for future improvement. Although the classification accuracy was high across most classes, certain coral types remain challenging to differentiate at the aerial scale due to image resolution constraints. Addressing this limitation may involve refining aerial image resolution by reducing the flight altitude (while respecting local aviation regulations in the country of data collection), employing more advanced image processing techniques, using higher-quality drones/cameras with better sensors (Giusti et al., 2022) or even using hyperspectral cameras to capture more detailed information about the reef (Rossiter et al., 2020).

Finally, mimicking the idea presented in this study, we could extend the methodology to satellite-scale imagery. Alvarez-Vanhard et al. (2021) discuss the complementary nature of UAV and satellite data, pointing out that integrating these technologies can improve spatial and temporal resolution in remote sensing applications. For instance, initiatives like the Allen Coral Atlas (Partnership, 2020), while offering large-scale reef coverage, often lack region-specific calibration, leading to inaccuracies in predictions (Nomenisoa et al., 2024). Successful integration would allow rapid monitoring of large areas, significantly reducing the need for field data collection, as satellite imagery is often freely available and collected at regular intervals. Furthermore, access to a historical archive of satellite imagery provides a unique opportunity to study ecosystem evolution over past decades, while supporting long-term monitoring efforts in the future. This extended temporal and spatial coverage could greatly improve our understanding of ecological change on a global scale.

Expanding to slow-moving species identification through synchronized imaging

The collection of both fine-scale and aerial images simultaneously has the potential to enable the identification of certain benthic species, which would otherwise be challenging to recognise. To illustrate, slow-moving organisms such as sea cucumbers are frequently visible in aerial images (e.g., Figure 4.9). Given that these species move at a slow

speed, synchronised collection of both image types would provide the temporal and spatial alignment necessary for passing the information from the fine-scale model to the medium-scale model. Although this approach imposes additional constraints compared to the method adopted in this study, where fine-scale and medium-scale data can be collected days or even months apart, it offers the advantage of providing additional ecological insights that would otherwise be inaccessible (Conand et al., 2018).

4.4.4 AI and biodiversity monitoring

The integration of AI into biodiversity and ecological monitoring offers many benefits but also introduces risks that should not be underestimated. On the one hand, AI enables unprecedented efficiency and accuracy in data collection, processing and analysis, allowing researchers to monitor ecosystems at a scale and depth that were previously unattainable (Tuia et al., 2022; Besson et al., 2022; Reynolds et al., 2024). On the other hand, the adoption of AI is not without challenges. The reliance on advanced technologies risks marginalising regions with limited resources or technical capacity, exacerbating global inequalities in biodiversity conservation (Sandbrook, 2024). Transparency of AI algorithms is another critical challenge. Conservation AI models often rely on "black-box" algorithms and restricted-access data, making it difficult for users, including scientists and policymakers, to understand how decisions are made. This opacity can lead to mistrust in AI systems and reduce their acceptance in conservation practices (Ayoola et al., 2024). A balance between exploiting the benefits of AI and addressing these risks is needed to ensure its effective and equitable use in ecology.

To promote transparency, reproducibility and equitable access, all raw and processed data used in this study are publicly available on Zenodo (Contini et al., 2024b,a). These data adhere to community standards outlined in Contini et al. (2025b), which fosters good practices for open science and Findable Accessible Interoperable Reusable (FAIR) data management. Standards implemented include Darwin Core and Global Biodiversity Information Facility (GBIF) for biodiversity data, Global Coral Reef Monitoring Network (GCRMN) protocols for scientific diving and Exchangeable Image File Format (EXIF) & Extensible Metadata Platform (XMP) metadata standards for camera imagery.

Similarly, all code used to process raw data and train AI models is openly accessible on GitHub and the trained models themselves are published on Hugging Face along with their respective configuration files. By making these resources available, we aim to encourage collaboration, facilitate reuse of our work and support the adoption of transparent, FAIR-aligned practices in ecological research.

Data availability

The data that support the findings of this study are openly available in Zenodo at Seatizen Atlas (Contini et al., 2025b, 2024a). To facilitate access to the input and output data of this study, these files have also been grouped in this repository.

All code for data processing associated with the current submission is available on drone-upscaling GitHub.

The code for downloading data associated with the current submission is available on zenodo-tools GitHub.

The code used to train the neural network model used in the current submission is available on DinoVdeau GitHub.

Conclusion and continuity

In this chapter, we presented a novel methodology for transferring information across scales in coral reef monitoring. Our approach leverages the strengths of high-resolution underwater imagery with medium-scale aerial data, ensuring the precision of fine-scale analysis while extending it to cover a broader reef area. This multi-scale framework has the potential to revolutionize marine monitoring, providing a more comprehensive and efficient way to assess coral reef health.

By training a medium-scale model to mimic the behaviour of a fine-scale model, we ensure that the system can easily incorporate new advances in AI without the need for data re-annotation. This flexibility allows us to continuously improve the model's predictions, ensuring that it remains at the forefront of coral reef monitoring. However, it is important to note that the upscaling process leads to the loss of certain classes, particularly those that are rare or below the spatial resolution of the aerial imagery. This limitation highlights the trade-off between spatial coverage and classification detail that is inherent to multi-scale approaches.

Looking ahead, we see great potential for our methodology to be extended to satellite-scale imagery. By integrating UAV and satellite data, we can enhance spatial and temporal resolutions in remote sensing applications, offering a broader, ecosystem-level view on coral reefs. This extended coverage could greatly improve our understanding of ecological change on a global scale, supporting long-term monitoring efforts and contributing to the conservation of these vital marine ecosystems.

Building on the concepts presented in this chapter, Chapter 5 explores a complementary approach to multi-scale coral reef monitoring. While the current methodology focused on transferring fine-scale knowledge from an underwater multi-label classification model to an aerial one, the following chapter will explore how multi-label predictions can be used as weak annotations to train a semantic segmentation model on aerial orthophotos.

Bibliography

Alvarez-Vanhard, E., Corpetti, T., and Houet, T. (2021). Uav & satellite synergies for optical remote sensing applications: A literature review. *Science of remote sensing*, 3:100019.

Ayoola, V. B., Idoko, P., Eromonsei, S. O., Afolabi, O., Apampa, A., and Oyebanji, O. (2024). The role of big data and ai in enhancing biodiversity conservation and resource management in the usa. *World Journal of Advanced Research and Reviews*, 23(02):1851–1873.

Besson, M., Alison, J., Bjerge, K., Gorochowski, T. E., Høye, T. T., Jucker, T., Mann, H. M., and Clements, C. F. (2022). Towards the fully automated monitoring of ecological communities. *Ecology Letters*, 25(12):2753–2775.

Conand, C., Ribes-Beaudemoulin, S., Trentin, F., Mulochau, T., and Boissin, E. (2018). Marine biodiversity of la reunion island: Echinoderms. *Western Indian Ocean Journal of Marine Science*, 17(1):111–124.

Contini, M., Barde, J., Bonhommeau, S., Illien, V., and Joly, A. (2024a). Seatizen atlas.

Contini, M., Barde, J., Bonhommeau, S., Illien, V., and Joly, A. (2024b). Seatizen atlas image dataset.

Contini, M., Illien, V., Barde, J., Poulain, S., Bernard, S., Joly, A., and Bonhommeau, S. (2025a). From underwater to drone: A novel multi-scale knowledge distillation approach for coral reef monitoring. *Ecological Informatics*, page 103149.

Contini, M., Illien, V., Julien, M., Ravitchandirane, M., Russias, V., Lazennec, A., Chevrier, T., Rintz, C. L., Carpentier, L., Gogendeau, P., et al. (2025b). Seatizen atlas: a collaborative dataset of underwater and aerial marine imagery. *Scientific Data*, 12(1):67.

Doukari, M. and Topouzelis, K. (2022). Overcoming the uas limitations in the coastal environment for accurate habitat mapping. *Remote Sensing Applications: Society and Environment*, 26:100726.

Elith*, J., H. Graham*, C., P. Anderson, R., Dudík, M., Ferrier, S., Guisan, A., J. Hijmans, R., Huettmann, F., R. Leathwick, J., Lehmann, A., et al. (2006). Novel methods improve prediction of species' distributions from occurrence data. *Ecography*, 29(2):129–151.

Giusti, E., Capria, A., Saverino, A. L., Gelli, S., Muñoz-Castañer, J., Dosil, R., Naya, J., and Menéndez, J. (2022). A drone-based multisensory payload for maritime pollutants detections. *IEEE Aerospace and Electronic Systems Magazine*, 38(3):4–18.

Gou, J., Yu, B., Maybank, S. J., and Tao, D. (2021). Knowledge distillation: A survey. *International Journal of Computer Vision*, 129(6):1789–1819.

Hinton, G., Vinyals, O., and Dean, J. (2015). Distilling the knowledge in a neural network. *stat*, 1050:9.

Lorieul, T. (2020). *Uncertainty in predictions of deep learning models for fine-grained classification*. PhD thesis, Université Montpellier.

Ma, Y., Song, Z., Zhuang, Y., Hao, J., and King, I. (2024). A survey on vision-language-action models for embodied ai. *arXiv preprint arXiv:2405.14093*.

Misiuk, B. and Brown, C. J. (2024). Benthic habitat mapping: A review of three decades of mapping biological patterns on the seafloor. *Estuarine, Coastal and Shelf Science*, 296:108599.

Nomenisoa, A. L. D., Todinanahary, G., Edwin, H. Z., Razakarisoa, T., Israel, J. B., Raseta, S., Jaonalison, H., Mahafina, J., and Eeckhaut, I. (2024). Remote sensing of coral reef habitats in madagascar using sentinel-2 satellite images. *Western Indian Ocean Journal of Marine Science*, 23(2):41–56.

Partnership, A. C. A. (2020). Allen coral atlas maps of the benthic cover and geomorphic zones of coral reefs in western micronesia.

Poulain, S., Nomenisoa, A. L. D., Andriamasinoroa, H., MD-NODC, and Barde, J. (2023). Orthophoto & dem (mne) issues d'images drone, uav, massif des roses, madagascar - 20230507 - 02_1.

Reynolds, S. A., Beery, S., Burgess, N., Burgman, M., Butchart, S. H., Cooke, S. J., Coomes, D., Danielsen, F., Di Minin, E., Durán, A. P., et al. (2024). The potential for ai to revolutionize conservation: a horizon scan. *Trends in Ecology & Evolution*.

Rossiter, T., Furey, T., McCarthy, T., and Stengel, D. B. (2020). Uav-mounted hyperspectral mapping of intertidal macroalgae. *Estuarine, Coastal and Shelf Science*, 242:106789.

Sandbrook, C. (2024). Beyond the hype: Navigating the conservation implications of artificial intelligence. *Conservation Letters*, page e13076.

Slocum, R. K., Wright, W., Parrish, C., Costa, B., Sharr, M., and Battista, T. (2019). Guidelines for bathymetric mapping and orthoimage generation using suas and sfm, an approach for conducting nearshore coastal mapping.

Tuia, D., Kellenberger, B., Beery, S., Costelloe, B. R., Zuffi, S., Risse, B., Mathis, A., Mathis, M. W., Van Langevelde, F., Burghardt, T., et al. (2022). Perspectives in machine learning for wildlife conservation. *Nature communications*, 13(1):1–15.

Ullah, F., Saqib, S., and Xiong, Y.-C. (2024). Integrating artificial intelligence in biodiversity conservation: bridging classical and modern approaches. *Biodiversity and Conservation*, pages 1–21.

Ventura, D., Grosso, L., Pensa, D., Casoli, E., Mancini, G., Valente, T., Scardi, M., and Rakaj, A. (2023). Coastal benthic habitat mapping and monitoring by integrating aerial and water surface low-cost drones. *Frontiers in Marine Science*, 9:1096594.

Wang, L. and Yoon, K.-J. (2021). Knowledge distillation and student-teacher learning for visual intelligence: A review and new outlooks. *IEEE transactions on pattern analysis and machine intelligence*, 44(6):3048–3068.

Øie Kvile, K., Gundersen, H., Poulsen, R. N., Sample, J. E., Salberg, A.-B., Ghareeb, M. E., Buls, T., Bekkby, T., and Hancke, K. (2024). Drone and ground-truth data collection, image annotation and machine learning: A protocol for coastal habitat mapping and classification. *MethodsX*, 13:102935.

Chapter 5

The point is the mask: scaling coral reef segmentation with weak supervision

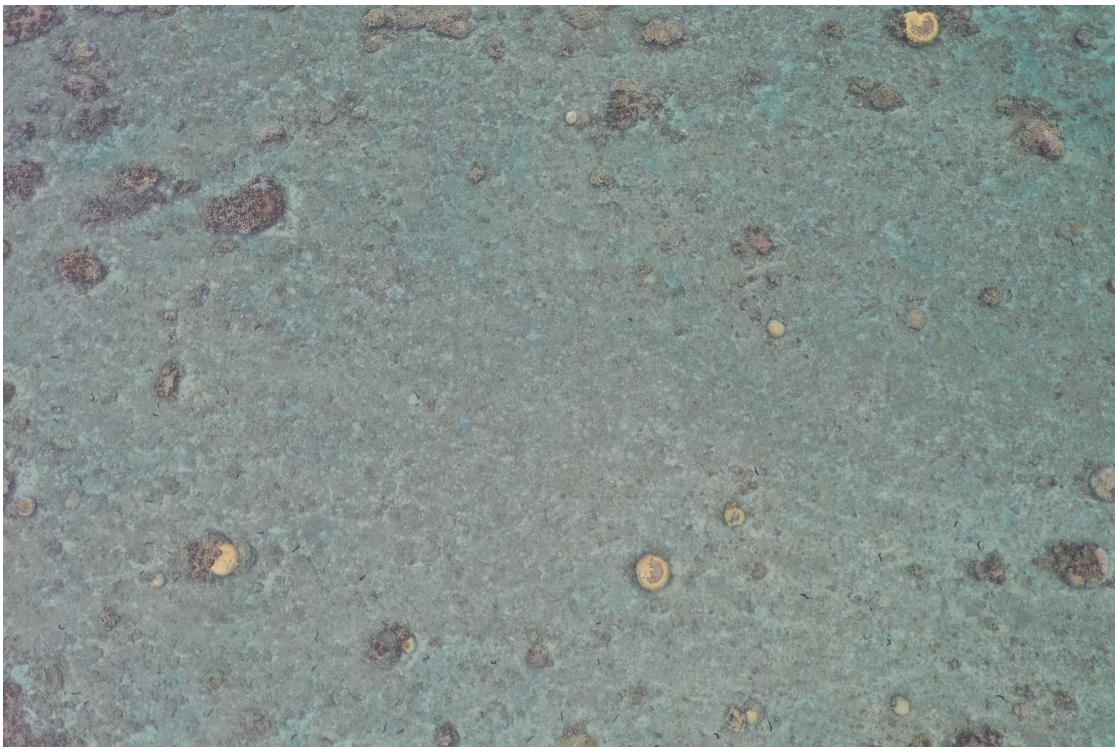


Figure 5.1: *Aerial drone view of benthic features in the shallow waters of the Trou d'Eau lagoon, Réunion Island, showing different coral morphotypes and sea cucumbers. Image credit: Serge Bernard, December 2023.*

This chapter is an adapted version of the article accepted in the *ICCV 2025 - Joint Workshop on Marine Vision* (Contini et al., 2025f). It describes the second upscaling methodology developed during this thesis, transferring underwater multi-label predictions to aerial semantic segmentation. Several sections have been expanded to provide more detailed explanations that could not be included in the conference paper due to its strict eight-page limit.

Context and objectives

Semantic segmentation plays a key role in ecological monitoring by enabling spatial mapping of ecological classes at the pixel level, thus allowing not only their identification but also the quantification of their extent and distribution within an image. In the context of coral reef monitoring, segmentation can, for example, be used to estimate surface cover for benthic classes such as corals, algae or seagrass. Moreover, when applied to georeferenced imagery, it can be used to derive morphological metrics such as the size or length of colonies or individual organisms (e.g., corals, sea cucumbers or turtles).

Despite its potential, semantic segmentation in marine environments remains a challenging task. The complexity of benthic scenes, combined with the presence of optical distortions such as caustics (patterns of light and shadow caused by the refraction of sunlight through the wavy water surface) often leads to visually noisy and ambiguous imagery where object contours are difficult to discern. Additionally, the annotation process for segmentation tasks is particularly challenging. Based on our experience, annotating a single image for semantic segmentation takes nearly four times longer than multilabel classification.

To address these challenges, this chapter introduces a multi-scale Weakly Supervised Semantic Segmentation (WSSS) framework that enables the training of high-resolution segmentation models on aerial imagery using classification-based outputs. Chapter 4 explored how to transfer classification knowledge from an underwater model to an aerial model to perform the same multilabel classification task on aerial imagery. In this chapter, we take a different approach by using those same multilabel outputs (interpolated into dense rasters) as weak annotations to train a student model for semantic segmentation on aerial images, without requiring manually labelled aerial data.

This work represents an advancement in efficient annotation pipelines for marine ecology. By reducing reliance on manual labelling through knowledge distillation from models trained on heterogeneous data sources, it enables the development of scalable segmentation models for coral reef monitoring using Unmanned Aerial Vehicle (UAV) imagery.

5.1 Introduction

Segmentation-based image analysis has transformed marine monitoring applications, allowing automated classification of underwater complex coral environments (Alonso et al., 2019; Raine et al., 2022; Zhong et al., 2023; Zheng et al., 2024).

Fully supervised semantic segmentation models require pixel-wise annotations, but producing such detailed labels for large-scale reef monitoring is labour-intensive and time-consuming (Sauder et al., 2024, 2025). Moreover, while broad habitat classes such as sand, coral, or seagrass can be reliably identified on UAV images (Doukari and Topouzelis, 2022; Qu et al., 2024; Øie Kvile et al., 2024), distinguishing between different coral morphotypes is challenging due to the limited resolution of UAV imagery that often does not provide enough discriminative detail to differentiate fine-scale morphological variations (see Figure 5.1[1]).

The problem is further exacerbated by the visual complexity of benthic scenes and the presence of optical distortions such as caustics (patterns of light and shadow caused by the refraction of sunlight through the wavy water surface) that produce noisy and ambiguous images where object boundaries are difficult to discern. This challenge demands alternative approaches that reduce annotation efforts while preserving annotation accuracy.

WSSS applied to remote sensing images offers a promising solution by leveraging weak labels, such as image-level annotations, bounding boxes, or rough segmentation masks, instead of manually labelled pixel-wise ground truth masks (Huang et al., 2023; Wang et al., 2023). This technique reduces the need for extensive manual annotation by training models on approximate labels (human or machine-generated) and learning to predict masks from these weak annotations. However, to our knowledge, WSSS approaches have not yet been applied in multi-scale contexts, where cross-modal knowledge transfer (e.g., from Autonomous Surface Vehicle (ASV)-based underwater classification to UAV-based aerial segmentation) can be used to improve the quality and scalability of segmentation models.

To address these challenges, this chapter introduces a multi-scale WSSS approach that bridges the gap between fine-scale coral reef monitoring based on underwater images and large-scale mapping based on aerial images. The workflow of the proposed method is illustrated in Figure 5.2.

Instead of manually annotating UAV images, we propose using classification-based probability maps derived from underwater images to generate segmentation annotations for aerial imagery. Our method spatially interpolates ASV-derived class predictions, producing rough segmentation masks that serve as training labels for SegFormer, a transformer-based semantic segmentation model (Xie et al., 2021). The segmentation mask outputs of the initial SegFormer model are then refined using the SAMRefiner algorithm described by Lin et al. (2025), a universal and efficient approach that adapts Segment Anything Model (SAM) to the mask refinement task (Kirillov et al., 2023). The

[1] The image can be found in Contini et al. (2025b) under the relative path:
`20231202_REU-TROU-DEAU_UAV-01_01/DCIM/DJI_0007.JPG`

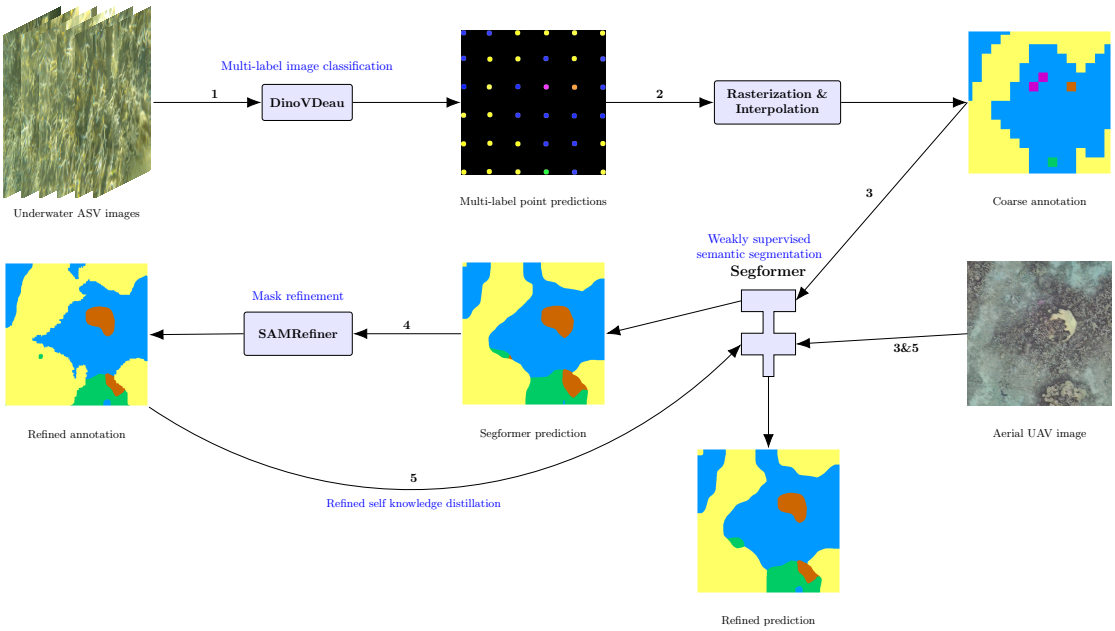


Figure 5.2: *Workflow of the proposed weakly supervised semantic segmentation (WSSS) multi-scale coral reef segmentation approach. Underwater images are classified using DinoVDeau model (step 1), generating spatialized predictions that are interpolated (step 2) to create continuous rasters. These rasters are used as coarse annotations for training an UAV-based segmentation model (step 3) using the SegFormer model. Model predictions are then refined using SAMRefiner (step 4) and used to retrain Segformer (step 5), improving segmentation accuracy.*

refined masks are then used to retrain the UAV model, improving final segmentation accuracy.

This approach:

- Transfers fine-scale information from ASV images to UAV segmentation models, reducing the need for manual labelling to the simpler task of classifying underwater images.
- Improves the spatial representation of coral morphotypes and habitats in aerial imagery, by passing from image classification to semantic segmentation.
- Enables large-scale coral mapping with high-resolution detail, supporting scalable conservation monitoring.

5.2 Materials and methods

This chapter presents a multi-scale method for generating fine-scale segmentation rasters of coral morphotypes and habitats on UAV imagery. The approach relies on collecting

and automatically classifying dense underwater images over a portion of the study area to provide fine-scale details for aerial model training.

First, a Deep Learning (DL) model (the teacher model) is used to classify coral morphotypes in ASV images collected in the Indian Ocean. These predictions are then linearly interpolated to generate continuous probability maps across the surveyed area. To obtain high-resolution rasters, a high density of images is required in the surveyed zone.

For this reason, an ASV is ideal as it allows transects to be spaced less than one metre apart, ensuring complete spatial coverage of the area. This enables the generation of reliable, fine-grained rasters representing coral morphotype distributions. Since ASV images are geolocated, these rasters can be aligned with aerial images (see Section 4.2.3 for more information on georeferencing), serving as weak segmentation annotations for training a second model (student) on UAV imagery. Finally, the student model predictions are corrected through a mask refinement model, improving the quality of the segmentation masks and enabling the retraining of the model for better accuracy. To sum up, this new methodology follows the following key steps:

1. Underwater data acquisition and image processing: collect high-density geolocated ASV images and classify them using *DinoVdeau* model (step 1 in Figure 5.2).

2. Interpolation & rasterization: create continuous spatial rasters from image classification outputs of the teacher model (Contini et al., 2024b) (step 2 in Figure 5.2).

3. Aerial data acquisition and image processing: build a high-resolution aerial orthophoto from UAV images, ensuring map overlay between ASV spatialized annotation rasters and UAV orthophoto (Contini et al., 2025c,a).

4. Aerial segmentation dataset construction: split fine-scale annotation rasters and aerial orthophoto into tiles, giving more weight to rare classes.

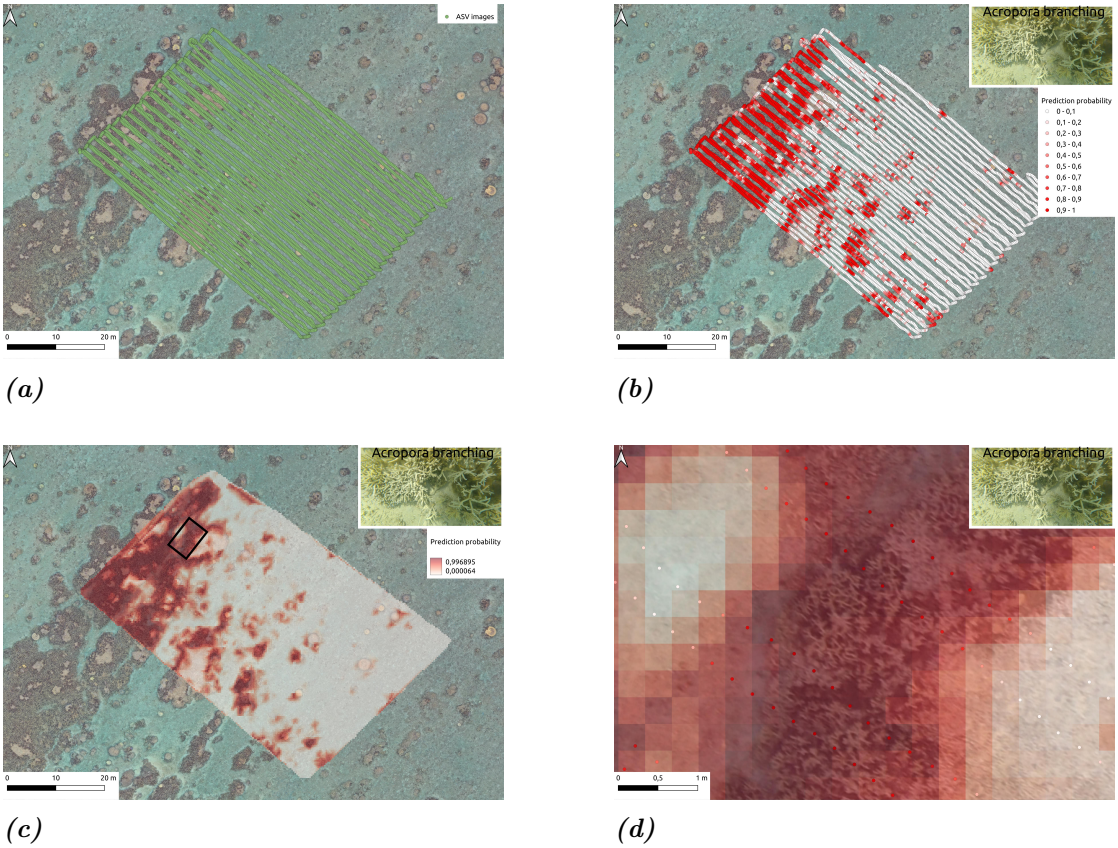
5. Aerial model training: use the generated annotations to train a student model for UAV-based coral and habitat semantic segmentation (step 3 in Figure 5.2).

6. Mask refinement and model retraining: refine student model predictions and retrain a new model to improve segmentation accuracy (steps 4 and 5 in Figure 5.2).

We provide the methodological details of each step in the following subsections.

5.2.1 Underwater data acquisition and image processing

Underwater images were collected using the *Plancha* ASV described in Chapter 3. Six ASV missions were conducted along predefined transects, spaced between 0.5 and 1 meter apart (Figure 5.3) in the *Trou d'eau* lagoon on Réunion Island.



(a) *(b)*

(c) *(d)*

Figure 5.3: *Example of the rasterization process on an ASV data collection event in the Trou d'eau lagoon in Réunion Island: (a) GPS coordinates of ASV images along the session. The high density of images allows for fine-scale spatial interpolation. (b) Point-wise predictions of Acropora Branching presence in each ASV image. (c) Interpolated raster representing the probability of Acropora Branching presence at each grid point, with the black rectangle indicating the area shown in Figure 5.3d. (d) Zoomed-in view of the raster and corresponding underwater predictions within the black rectangle from Figure 5.3c. The raster provides a continuous spatial representation of coral morphotype distributions, enabling the generation of rough segmentation masks that can be used for UAV model training.*

Each ASV image (frame extracted from raw videos) was enriched with metadata, including latitude, longitude and attitude metadata (roll, pitch and yaw angles), so that the images could be geolocated and orientated correctly in the survey area (Figure 5.3a).

To obtain fine-scale predictions of coral morphotypes and habitat types, the publicly available model *DinoVdeau* presented in Section 3.2.3 is used as a teacher model. The model outputs probability scores for 31 classes described in Section 3.2.3, indicating the likelihood of each class being present in a given image (in Figure 5.3b, predictions for the *Acropora Branching* class are shown on an ASV session). Since image classification

identifies whether a class is present in an image but does not generate segmentation masks of the class's exact location, no information about their spatial coverage or location within the frame is provided.

The teacher model was applied to six ASV sessions conducted in the *Trou d'eau* lagoon, Réunion Island, generating predictions over a total surveyed area of 16,904 m^2 (step 1 in Figure 5.2).

5.2.2 Interpolation & rasterization

To generate dense training labels for UAV image segmentation, ASV pointwise predictions are transformed into continuous spatial rasters (step 2 in Figure 5.2).

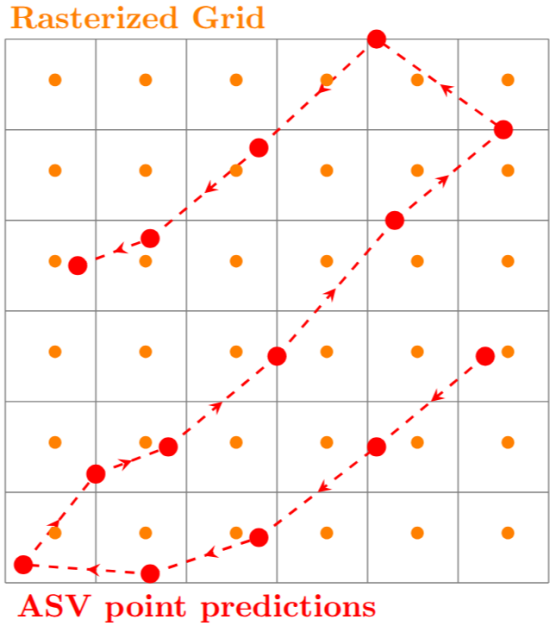


Figure 5.4: *Illustration of the rasterization process used to convert ASV-derived point predictions into a continuous grid. Red dots represent ASV image locations along the survey path, with arrows indicating the acquisition sequence. Orange dots mark the centres of the raster grid cells. The grid spacing is set to the median distance between two consecutive ASV points, typically ranging from* 0.3 m *to* 0.35 m*, to ensure sufficient resolution for spatial interpolation.*

To ensure a fine spatial resolution, the grid spacing for rasterization was set to the median distance between two consecutive ASV frames along the survey (Figure 5.4). Depending on the survey, this distance varied between 0.3 m and 0.35 m. A uniform grid was then created over the survey area, with each grid point representing a pixel in the final raster. Raster generation was performed using linear interpolation.

These rasters are aligned with UAV orthophotos to provide approximate segmentation masks for UAV model training. Figure 5.3c shows the raster generated from ASV

predictions for the *Acropora Branching* class and in Figure 5.3d a zoomed-in view of the raster and the corresponding underwater prediction points.

For each ASV data collection event and for each class predicted by the teacher model, a probability raster was generated, representing the likelihood of its presence at every grid point.

5.2.3 Aerial data acquisition and image processing

Aerial drone images are taken with a `DJI Mavic 2 Pro` drone. A high-resolution UAV orthophoto is constructed by mosaicking individual UAV images. Since the drone is not equipped with a differential GPS, once images were taken and the Structure from Motion (SfM) model was built, the orthophoto positioning is refined by collecting Ground Control Points (GCPs) on the surveyed area using a differential GPS.

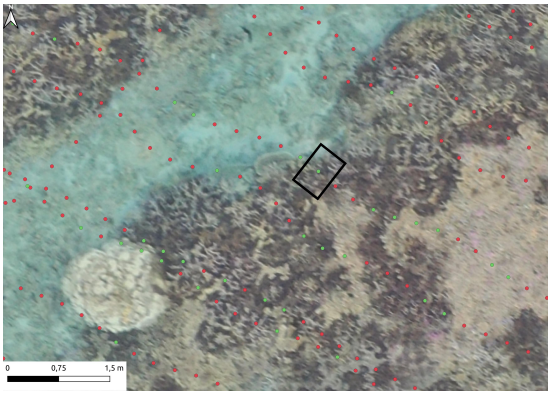
Two missions were carried out in the lagoon of Réunion Island: one in the *Saint-Leu* lagoon (Contini et al., 2025a) and the other in the *Trou d'eau* lagoon (Contini et al., 2025c), measuring 204,748 m^2 and 189,682 m^2 respectively. The Ground Sampling Distance (GSD) of the resulting orthophotos was 0.9 cm for *Trou d'eau* and 1.6 cm for *Saint-Leu*, reflecting differences in flight altitude and mission setup. The model is trained on the *Trou d'eau* lagoon and tested on the *Saint-Leu* lagoon, in order to evaluate its ability to generalize across geographically distinct sites without site-specific tuning.

Further details on mission planning, execution and orthophoto georeferencing are provided in Section 4.2.2.

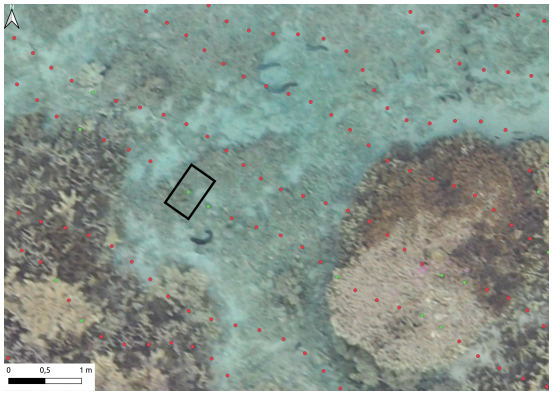
5.2.4 Aerial segmentation dataset

Classes selection

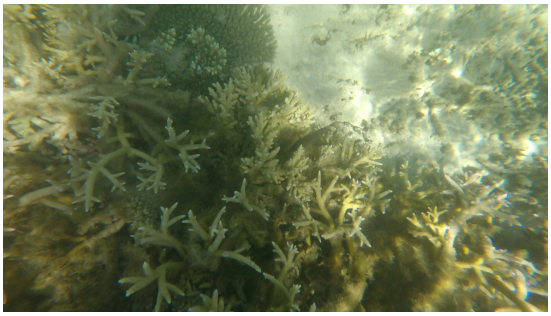
Some classes used in the underwater teacher model cannot be confidently identified in UAV orthophotos due to their size (*Algae* classes, small coral colonies like *Acropora Digitate*) colour similarity (*Dead coral* and *Rubble*) or simply lack of resolution in the aerial images. Figure 5.5 shows two illustrative examples comparing aerial and underwater views of the same locations, highlighting this limitation.



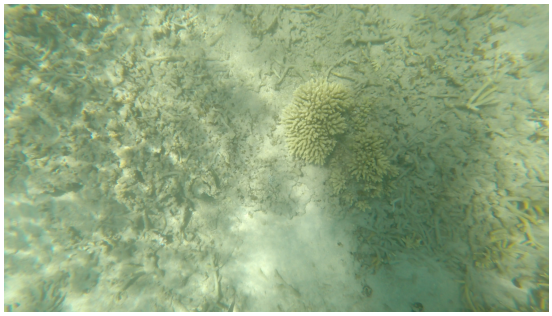
(a) *Aerial view of a reef zone with predicted Algal Assemblage presence. Green dots indicate ASV image locations where the underwater DinoVdeau model predicted the presence of Algal Assemblage; red dots indicate absence. The black rectangle highlights the location of the underwater image shown in Figure 5.5c, where a dense algal cover is confirmed.*



(b) *Aerial view of a reef zone with predicted Acropora Digitate presence. Green dots indicate ASV image locations where the underwater DinoVdeau model predicted the presence of Acropora Digitate; red dots indicate absence. The black rectangle highlights the location of the underwater image shown in Figure 5.5d, where a small colony of Acropora Digitate is confirmed.*



(c) *Underwater image taken by an ASV camera showing a dense algal assemblage, confirming DinoVdeau prediction at this location.*



(d) *Underwater image showing a small colony of Acropora Digitate, which is ecologically relevant but not distinguishable in the corresponding aerial orthophoto due to its small size and low contrast.*

Figure 5.5: *Comparison of aerial orthophotos (top) and corresponding underwater images (bottom) for two example locations. These examples illustrate how some benthic classes, such as Algal Assemblage (left) and small coral morphotypes like Acropora Digitate (right), can be clearly identified in underwater imagery but remain indistinct in aerial views due to resolution and colour similarity limitations.*

Figure 5.5a and Figure 5.5b show two aerial orthophotos with green and red dots indicating the presence or absence of *Algal Assemblage* and *Acropora Digitate* predicted by the underwater *DinoVdeau* model on the corresponding ASV images. The black

rectangles highlight the locations of the underwater images shown in Figure 5.5c and Figure 5.5d, where the presence of *Algal Assemblage* and *Acropora Digitate* is confirmed. These examples illustrate how some benthic classes, such as *Algal Assemblage* (left) and small coral morphotypes like *Acropora Digitate* (right), can be clearly identified in underwater imagery but remain indistinct in aerial views due to resolution and colour similarity limitations. To simplify the segmentation task, we retained only classes that are clearly distinguishable in aerial images.

Other classes, although ecologically important and distinguishable on UAV orthophotos like seagrass species *Syringodium isoetifolium* and *Thalassodendron ciliatum* were excluded. This decision was based on the limited number of presence predictions available from the underwater model, which would have increased label noise during training. Figure 5.6 presents the relative abundance of each of the 31 benthic classes predicted in the underwater imagery by *DinoVdeau* model, on the six ASV sessions retained for this study.

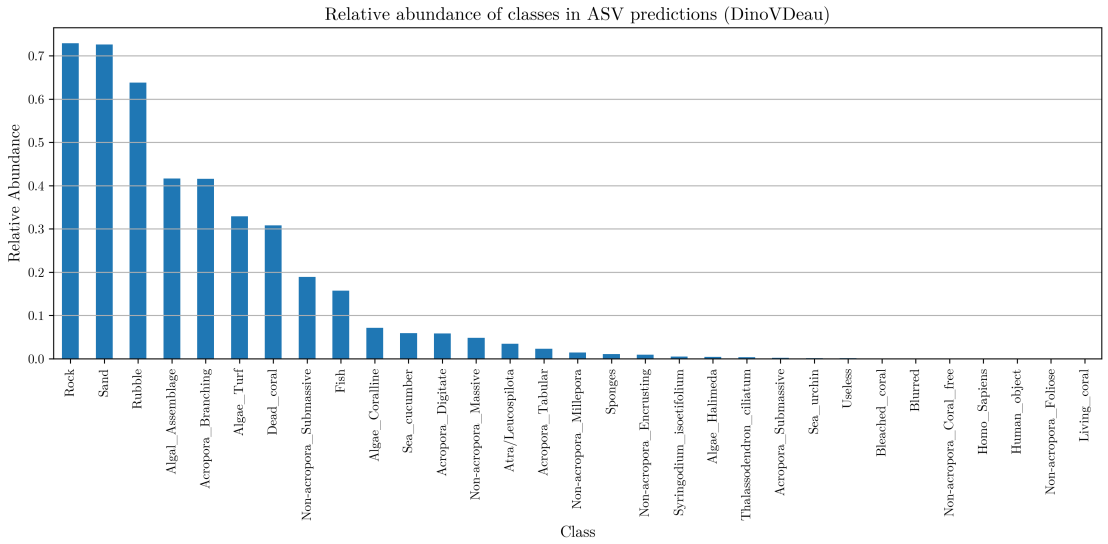


Figure 5.6: *Relative abundance of benthic classes predicted in the underwater imagery by DinoVdeau model across the six training ASV sessions.*

Hard-substrate classes like *Rock*, *Sand* and *Rubble* dominate the distribution but only *Sand* was retained in this study. *Rock* and *Rubble* were excluded because they appear too similar to *Sand* in UAV imagery at the given resolution. Additionally, their frequent co-occurrence in underwater predictions would have introduced significant label noise during the generation of coarse annotations. It is also important to note that the relative abundance values shown in Figure 5.6 do not sum to 1. This is because the underwater model *DinoVdeau* is a multi-label image classification model, allowing multiple classes to be predicted for a single image, reflecting the natural co-occurrence of benthic features in reef environments.

Thus, for the purpose of a proof of concept, we selected the following subset of classes,

which exhibit clear visual patterns in aerial images:

1. *Acropora Branching*
2. *Acropora Tabular*
3. *Non-acropora Massive*
4. *Other Corals*
5. *Sand*

The first two classes are common coral morphotypes of the genus *Acropora* that can be found in Réunion Island's lagoons. *Acropora Branching* is a branching coral easily distinguishable from aerial images due to its particular spectral signature (dark brown) and is often found in large colonies. *Acropora Tabular* is a tabular coral a little bit more difficult to identify on aerial images but still distinguishable due to the particular layer shape of its colonies. The third class *Non-acropora Massive* is a stony coral with a light spectral signature, often found alone surrounded by sand. The fourth class *Other Corals* encompasses many different species of corals that can range from small colourful colonies of the genus *Pocillopore* to large stony corals that are darker in colour than *Non-acropora Massive* corals.

Annotation generation

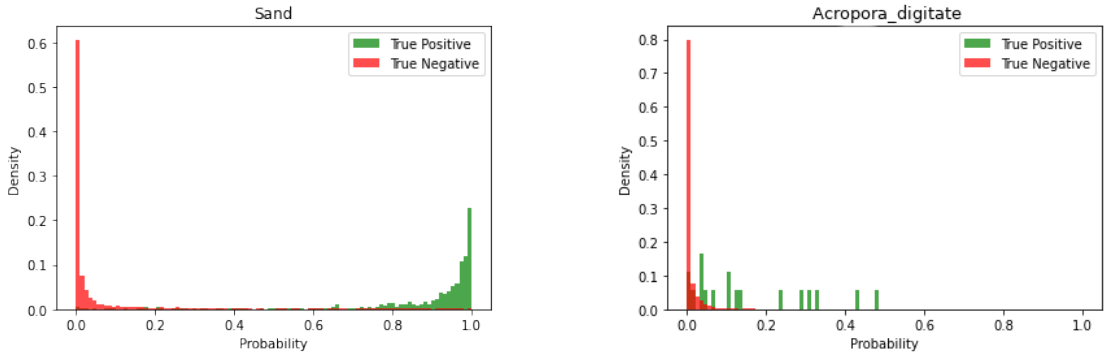
In the semantic segmentation task, a model is trained to predict a class label for each pixel in an image. To achieve this, a set of images must be labelled so that each pixel corresponds to one and only one class. Let C be the set of classes, I be the set of images and N the cardinality of C.

In our case, for an image $i \in I$, we have N continuous rasters representing the probability of presence for each class $c \in C$. Thus, we need a function that, for each pixel, maps these probabilities to a class label. A straightforward approach is to assign to each pixel the label of the class with the highest probability of presence.

However, this approach has limitations. Since the underwater teacher model is a multilabel classifier (trained with the Binary Cross-Entropy (BCE) loss function), the predicted presence/absence probabilities of each class are independent and not necessarily calibrated in the same way. As an illustration, Figure 5.7 shows the distribution of the predictions of true positives and true negatives for two classes.

For common classes such as *Sand*, the teacher model tends to assign very high probabilities to true positives and low values to true negatives, resulting in well-separated bimodal distributions (Figure 5.7a). In contrast, for rare classes such as *Acropora Digitate*, the model shows a lower confidence, with true positive predictions spread more uniformly throughout the probability range (Figure 5.7b). This leads to a systematic underestimation of rare classes when using naïve maximum-probability labelling.

To address this, we implement a class-specific quantile normalisation strategy. Following the principles outlined in Zhao et al. (2020), we normalise the predicted probability maps for each class independently using their empirical percentiles. Specifically, for each class c, we aggregate all predicted probabilities generated by the teacher model across all ASV sessions and compute the 1st and 99th percentiles. We then linearly



(a) *(b)*

Figure 5.7: *Probability distribution of true positives and true negatives for two classes in the underwater image classification teacher model. (a) Sand class is a common class with high probabilities of presence when correctly detected and low probabilities when absent. (b) Acropora Digitate class is a rare class with a low probability of presence, even when correctly detected.*

scaled the per-pixel probability values within this range. This ensures that for each class, the normalised values reflect relative confidence within that class's distribution, rather than across the entire set.

Formally, let $p_c(x, y)$ denote the predicted probability of presence for class c at pixel location (x, y) in the ASV-derived raster. Let $q_1(c)$ and $q_{99}(c)$ be the empirical 1st and 99th percentile values of the predicted probability distribution for class c across all sessions. The normalised probability $\hat{p}_c(x, y)$ is computed as:

$$\hat{p}_c(x, y) = \text{clip}\left(\frac{p_c(x, y) - q_1(c)}{q_{99}(c) - q_1(c) + \epsilon},\ 0,\ 1\right) \tag{5.1}$$

where ϵ is a small constant added for numerical stability and the clip function ensures that values remain within the valid probability range. This normalisation method improves the naïve maximum selection by taking into account the class-specific prediction dynamics and mitigating the bias caused by imbalanced raw probability distributions. The final label assigned to each pixel is then:

$$L(x, y) = \arg\max_{c \epsilon C} \hat{p}_c(x, y) \tag{5.2}$$

This quantile-based normalisation is conceptually similar to class-specific normalisation strategies explored in gene expression data preprocessing, where non-uniform expression distributions across classes can lead to biased analyses if left uncorrected (Zhao et al., 2020). Applying this principle to semantic segmentation in ecology helps rectify inter-class imbalance, leading to a more equitable representation of rare classes in the resulting annotations.

Finally, since the aerial orthophoto and the ASV annotation rasters do not share the same resolution and since semantic segmentation models require a one-to-one pixel correspondence between the image and annotation, we upsample the ASV annotation rasters to the resolution of the orthophoto by replicating nearest-neighbour pixel values.

5.2.5 Aerial deep learning model (student model)

WSSS provides a framework for training models using approximate labels instead of manually annotated ground truth masks (Chan et al., 2021). In our case, the annotations are derived from ASV-based probability maps, which are inherently coarse and may contain noise (step 3 in Figure 5.2). These weak labels, combined with the challenges of UAV-based segmentation (such as variable lighting conditions, water clarity differences and limited spatial resolution) introduce significant uncertainty into the training process. As a result, the segmentation model must be robust to label noise and able to generalise to different spatial conditions and environmental contexts.

To address these requirements, we used SegFormer, a transformer-based segmentation architecture that offers several advantages over conventional Convolutional Neural Networks (CNNs) (Xie et al., 2021; Ru et al., 2022):

- Hierarchical feature representation: SegFormer processes images at multiple scales, allowing it to capture both global contextual information and fine-grained details. This is particularly useful for UAV imagery, where coral morphotypes may have highly variable spatial patterns.
- Robust to weak annotations: unlike models that rely on precise pixel-wise labels, SegFormer effectively learns from weak labels by employing an efficient self-attention mechanism, improving segmentation accuracy even when trained on approximate masks.
- Lightweight and efficient: unlike traditional vision transformers, SegFormer does not require heavy computational resources, making it feasible for large-scale aerial image processing.

Choosing an appropriate loss function is critical in WSSS, where training annotations are coarse and often affected by noise. Region-based losses, such as Dice loss, are especially suitable in these scenarios because they evaluate the overlap between predicted and annotated regions as a whole, rather than focussing on individual pixels or precise edges (Azad et al., 2023). Unlike pixel-wise losses, which are sensitive to misaligned labels, or boundary-based losses, which require accurate contour annotations, region-based losses are more tolerant of uncertainty in object boundaries, making them a better fit for learning from approximate masks. Given the nature of this problem, we opted to minimise the region-based Dice loss with an Adam optimiser (Kingma and Ba, 2014).

The input and output resolutions are chosen to be 512 px × 512 px, with a batch size of 16 images. The initial learning rate was initially set to 10^{-5} and decayed by a factor of 10 whenever the validation loss plateaued for five consecutive epochs.

5.2.6 Mask refinement and model retraining

The final step in our methodology involves refining the initial segmentation masks generated by the student model (step 4 in Figure 5.2). A common technique in Computer Vision (CV) for improving model performance without additional supervision is Self Knowledge Distillation (SKD), where the model is retrained using its own predictions as pseudo-labels (Xu et al., 2024). In this configuration, the same model assumes the roles of both teacher and student, iteratively improving itself by distilling and refining its own previously generated output.

In our case, we explored both this technique and a variant in which the masks are first refined before retraining using a mask refinement model. The chosen model is SAMRefiner, a universal and efficient approach that adapts SAM to the mask refinement task (Lin et al., 2025).

Once annotations refined, we finally retrained Segformer (step 5 in Figure 5.2).

5.2.7 Test zone and model evaluation

Due to the lack of human annotations, the DL model cannot be evaluated on weakly generated annotations using standard CV metrics on the test set. For this reason, we selected two test zones: one within the *Trou d'eau* lagoon measuring 84 m^2 and the other in the *Saint-Leu* lagoon measuring 194 m^2.

We then manually annotated all coral morphotypes and habitats present in these areas.

These zones were chosen due to their diverse composition of coral morphotypes, habitats and other marine organisms, representing a challenging environment for model validation (Contini et al., 2025c,a).

5.3 Results

5.3.1 Evaluation of model variants during pipeline development

To justify the contribution of each component in the proposed pipeline (illustrated in Figure 5.2), we evaluated the segmentation performance of multiple model variants on manually annotated ground-truth data from the test zones in the *Trou d'eau* and *Saint-Leu* lagoons (see Section 4.2.6). The following model configurations were tested to isolate the effect of each step in the pipeline:

- *Coarse model*: Segformer trained solely on coarse annotations derived from ASV predictions (step 3 in Figure 5.2). No mask refinement or self-distillation is applied.
- *Coarse refined*: coarse annotations are first refined using the SAMRefiner model. Segformer is then trained on these refined masks. No self-distillation is applied.
- *Coarse self-distilled*: Segformer is initially trained on coarse annotations and then retrained using its own predictions as supervision (self-distillation), without involving SAMRefiner.

- *Coarse refined + distilled*: full pipeline. Segformer is first trained on coarse annotations, whose predictions are then refined using SAMRefiner. A final Segformer is trained via self-distillation on these refined predictions. This is the final proposed model (step 5 in Figure 5.2).

Table 5.1 presents total pixel accuracy, mean Intersection over Union (IoU) and per-class IoU for each model. For a more detailed definition of these metrics, please refer to (Schlosser et al., 2024).

Table 5.1: *Comparison of total accuracy and mean IoU across different model variants, along with IoU per class.*

Model	Accuracy	Mean IoU	Acropora B.	Acropora T.	Non-acro M.	Other Corals	Sand
Coarse	0.8250	0.4625	0.5109	0.2087	0.2877	**0.4201**	0.8854
Coarse Refined	0.8467	0.4658	0.4676	0.1532	0.3721	0.4078	**0.9285**
Coarse Self-Distilled	0.8300	0.4745	0.5175	0.2114	0.3395	0.4190	0.8852
Coarse Refined + Distilled	**0.8572**	**0.5010**	**0.5282**	**0.2313**	**0.4095**	0.4111	0.9249

The results show that although the *Coarse* and the *Coarse refined* model variants achieve a slightly better IoU for *Other Corals* and *Sand* classes respectively, only the full pipeline offers a more balanced trade-off between pixel-level accuracy and spatial precision, as indicated by its superior accuracy and mean IoU scores. This progressive evaluation validates the effectiveness of the proposed weakly supervised multi-scale training framework.

In the following sections, we focus on the results obtained with the full pipeline model to evaluate the quality and ecological relevance of the segmentation outputs across annotated and unannotated reef zones.

5.3.2 Model performance on annotated reef zones

We now focus on the performance of the final model (i.e., the *Coarse refined + distilled* variant) on the same annotated test zones used in the previous section, namely the *Trou d'eau* and *Saint-Leu* lagoons (see Section 4.2.6). This assessment aims to characterise the spatial accuracy and ecological relevance of the predicted segmentation masks at the UAV scale. Per-zone and per-class metrics are presented in Table 5.2.

Table 5.2: *IoU per class, total accuracy and mean IoU by zone.*

Zone	Accuracy	Mean IoU	Acropora B.	Acropora T.	Non-acro M.	Other Corals	Sand
Trou d'eau	0.9053	0.4493	0.2354	0.3623	0.3295	0.3707	0.9484
Saint-Leu	0.8160	0.5793	0.5641	/	0.4318	0.4242	0.8972
Total	**0.8572**	**0.5010**	**0.5282**	**0.2313**	**0.4095**	**0.4111**	**0.9249**

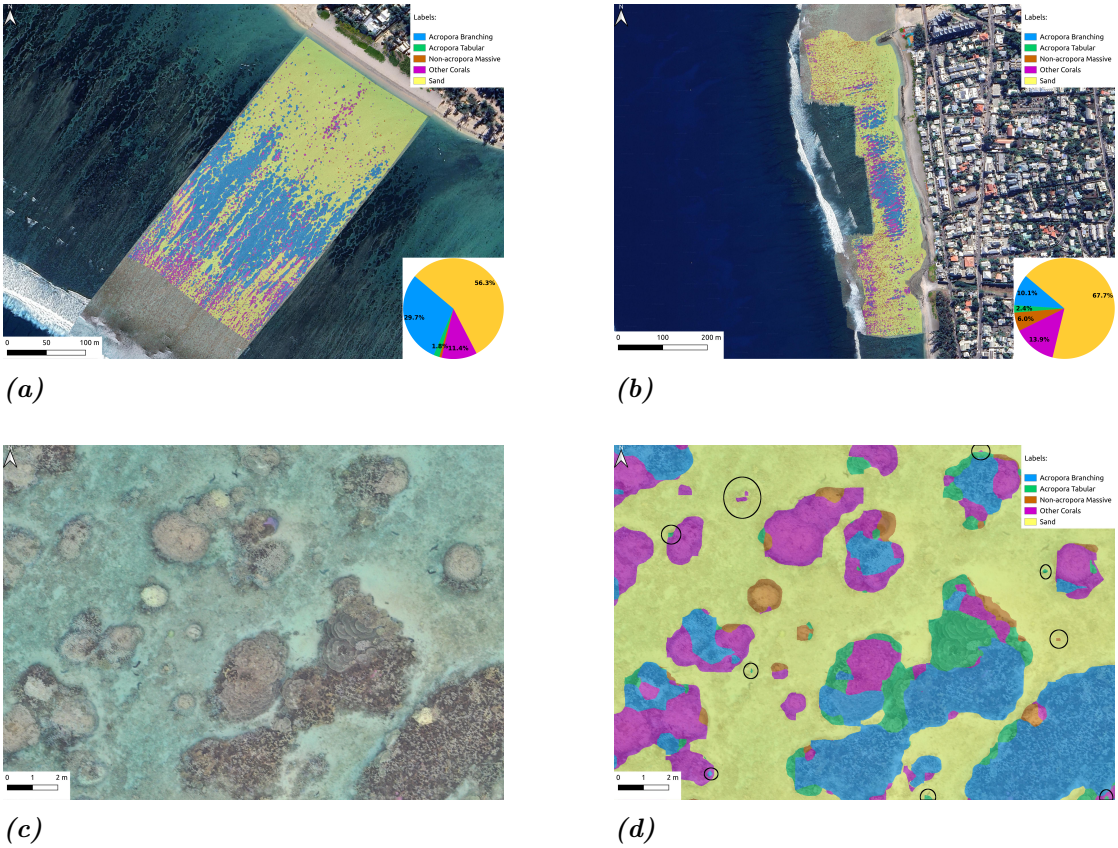
The total pixel accuracy reached 85.72%, while the mean IoU was 50.10%. The confusion matrix, showing the percentage of true class pixels predicted for each class, along with misclassification errors, is presented in Figure 5.8.



Figure 5.8: *Normalized confusion matrix computed across both test zones (Trou d'eau and Saint-Leu). Values represent the percentage of true class pixels predicted for each class.*

5.3.3 Predicted habitat composition across reef zones

To assess the spatial consistency and ecological relevance of the segmentation predictions across large reef areas, we applied the trained UAV model to the entire UAV orthophotos of the *Trou d'eau* and *Saint-Leu* lagoons, measuring 189,682 m^2 and 204,748 m^2 respectively, see Figure 5.9. As a reminder, the six ASV sessions used to train the model covered a total area of total of 16,904 m^2. Therefore, on average a single UAV session covers approximately 100 times more area than an ASV survey of equivalent time duration, assuming both platforms maintain horizontal overlap between transects. However, this gain in spatial extent comes at the cost of resolution: the ASV imagery typically achieves a GSD of about 1 mm px^{-1} in the lagoon, whereas UAV imagery provides a coarser resolution of roughly 1 cm px^{-1}.



(a) *(b)*

(c) *(d)*

Figure 5.9: *Results of the UAV model applied to the entire orthophoto of the Trou d'eau lagoon (a) and Saint-Leu lagoon (b). The model predicts the presence of coral morphotypes and habitats across the entire reef area. Piecharts can then be computed to quantify the predicted habitat composition. (c) Zoomed-in view of the UAV orthophoto of the Trou d'eau lagoon. (d) Corresponding predicted segmentation mask for the zoomed-in area of the Trou d'eau lagoon orthophoto, showing the spatial distribution of coral morphotypes and habitats. Black circles highlight spurious patches of different corals predicted by the model, which are not present in the orthophoto.*

In Figures 5.9a and 5.9b, the model predictions over the entire orthophotos of the *Trou d'eau* and *Saint-Leu* lagoons are displayed, accompanied by pie charts summarising the percentage cover of each predicted class. Finally, Figure 5.9d presents a zoomed-in view of the UAV orthophoto of the *Trou d'eau* lagoon, while Figure 5.9c shows the corresponding predicted segmentation mask.

5.4 Discussion

The results of the multi-scale upscaling process demonstrate the feasibility of training a UAV-based coral segmentation model using weak annotations derived from ASV-based

image classification. As shown in Figure 5.9, the benthic coverage for each class can be quantified while highlighting spatial patterns throughout the reef scape. In Figures 5.9a and 5.9b, the model predictions over the entire orthophotos of the *Trou d'eau* and *Saint-Leu* lagoons are displayed, accompanied by pie charts summarising the percentage cover of each predicted class. Although surface area is a useful metric for many benthic classes, it may be less informative for others (e.g., *Non-acropora Massive* or mobile species) where the number of detected individuals serves as a more ecologically relevant indicator. As demonstrated in Section 5.4.3, the proposed method also enables individual-level detection and counting for such classes. Figures 5.9c and 5.9d present a zoomed-in view of a specific area in the *Trou d'eau* lagoon and the corresponding segmentation mask. Black circles highlight spurious patches of different corals predicted by the model, which are not present in the orthophoto. While minor misclassification errors can be observed, such as isolated spurious patches of *Non-acropora Massive*, the overall quality of the segmentation is encouraging. The predicted masks accurately reflect the structure and extent of the main coral morphotypes, with clear and coherent shapes that align well with the visible habitat patterns in the aerial imagery. This qualitative assessment highlights the potential of the model to support spatial analyses of reef composition and guide targeted conservation actions.

5.4.1 WSSS loss function

When dealing with WSSS the choice of the loss function can have a significant impact on the model performance. Indeed during training, the loss function is used at each iteration to compute the error between the predicted segmentation mask and the ground truth mask. Since the ground-truth masks are generated from weak annotations, they contain noise and inaccuracies, which must be taken into account when choosing the loss function. For the first training of Segformer model, we tried four different loss functions: boundary loss, dice loss, Cross-Entropy (CE) loss and focal loss.

The comparison between the coarse annotation mask and the prediction masks obtained by training the model shows substantial differences between the four different loss functions (Figure 5.10). The CE loss and focal loss are pixel-level losses that focus on individual pixels (Azad et al., 2023), making them sensitive to label noise and class imbalance. This is reflected in the predictions, that often do not recognise under-represented classes such as *Acropora Tabular* and *Non-acropora Massive*. Boundary level losses, like the boundary loss, are designed to focus on the edges of objects, which can be useful for fine-grained segmentation tasks. For WSSS however, boundary losses are too sensitive to noise in the annotations and this is clearly reflected in the predictions in Figure 5.10, which exhibits high pixelation effects.

Finally dice loss (Sudre et al., 2017), although sometimes produces noisier masks than the ones obtained with other losses, is the only one that correctly handled the class imbalance. In fact, the dice loss is a region-based loss that evaluates the overlap between predicted and annotated regions as a whole, rather than focusing on individual pixels or precise edges. This makes the loss robust to class imbalance even if less sensitive to fine edge details or precise boundary alignment.

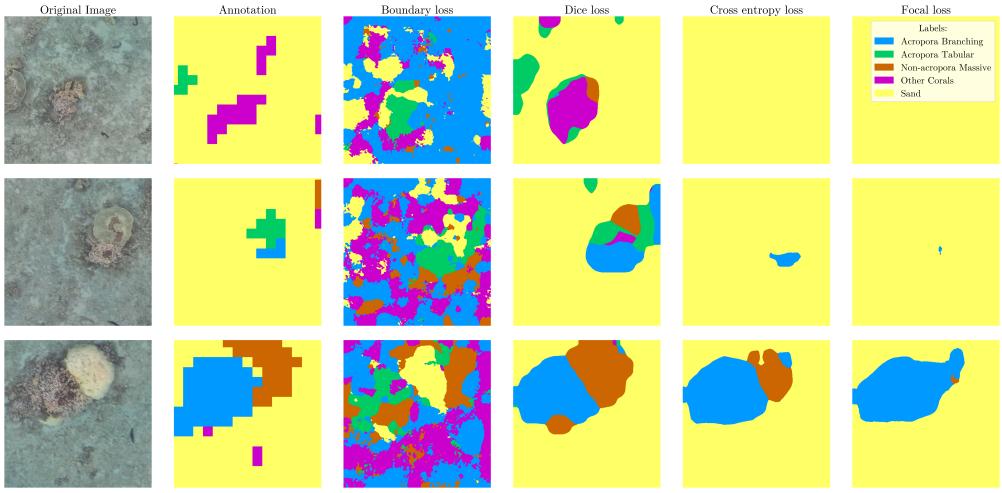


Figure 5.10: *Comparison of four different loss functions for WSSS training with coarse annotations. The first column shows the original UAV tile, followed by the coarse annotation used for training. The remaining columns display the predicted segmentation masks obtained using boundary loss, dice loss, CE loss and focal loss, respectively.*

5.4.2 Aerial model evaluation

With a total accuracy of 85.72% and a mean IoU of 50.10%. (see Table 5.2), the model effectively captures the spatial distribution of coral morphotypes and habitats in aerial imagery. The difference between the two metrics indicates that the model is biased towards the most common class, *Sand*, which is easily recognisable and segmentable in aerial images. Its distinct spectral signature and broad spatial coverage in aerial imagery likely contributed to the high recall of 93.6% (see Figure 5.8). The model also performed well on the three coral classes *Acropora Branching*, *Acropora Tabular* and *Non-acropora Massive*, correctly identifying 79.6%, 85.5% and 83.6% of the annotated pixels for each class respectively.

The lowest performance is encountered for the *Other Corals* class. This class, grouping various coral morphotypes, is characterised by high intra-class variability and spectral ambiguity, leading to a lower performance of 47.3%. For example the *Acropora Digitate* morphotype, which is a small coral colony being part of this class, can be easily confused on aerial images with *Acropora Branching* or *Acropora Tabular* morphotypes, which explains the high misclassification rates of 29.2% and 15.6% respectively, as shown in Figure 5.8. Furthermore, the *Other Corals* class groups a wide variety of coral morphotypes that differ greatly in appearance and structure. For instance, it includes both large, dome-shaped colonies that resemble massive corals and also flat, spreading forms like encrusting corals, which are morphologically and spectrally distinct. This high intra-class variability makes it difficult for the model to learn coherent and consistent features, reducing its classification performance. While the current model was not trained to recognize the full spectrum of coral morphotypes, the methodology itself is generic and has

been validated for its ability to incorporate new classes as needed (see Section 5.4.3). Although subdividing the *Other Corals* category into more specific morphotypes would likely improve accuracy, this was not pursued in the present study due to the need for more detailed annotations, an expanded underwater dataset and higher-resolution UAV imagery to ensure reliable training.

5.4.3 Expanding to new ecological classes

The presented methodology can be easily extended to include additional benthic classes, such as specific types of seagrass or other coral morphotypes not found in Réunion Island. To incorporate new classes, it is sufficient to collect both underwater and aerial images in a region where the new class is present and to train the student model using the ASV-derived annotations following the same upscaling methodology. It is important to note that the method remains applicable regardless of the specific architecture of the underwater model, as long as it provides reliable class-level predictions (whether from monolabel, multilabel or segmentation classifiers). This flexibility ensures the continued relevance and adaptability of the approach as Artificial Intelligence (AI) models evolve, while also enabling its transferability to different reef systems for broader regional monitoring without requiring changes to the core pipeline.

Another way to incorporate new species, including mobile ones such as sea cucumbers or turtles, is to manually annotate a small number of instances directly in the aerial imagery. For slow-moving species such as sea cucumbers, the collection of fine-scale ASV images and UAV images simultaneously can provide the spatial and temporal alignment required to propagate class labels from underwater predictions to aerial images (Contini et al., 2025d). However, weather conditions or logistical constraints may prevent synchronised data collection. In addition, for fast-moving species such as turtles, simultaneous acquisition is not feasible. In these situations, a few manual annotations can be added on top of ASV-generated masks to introduce new classes to the training dataset.

The introduction of SAM in 2023 by Kirillov et al. (2023) has greatly simplified the annotation process. Tools such as the *Geo-SAM* plugin for QGIS (Zhao et al., 2023) allow users to create object masks by clicking on points or drawing bounding boxes, reducing the time required for detailed labelling. For instance, using Geo-SAM, we manually annotated 338 sea cucumbers in a 13,840 m^2 area of the *Trou d'eau* lagoon in approximately 55 minutes.

These annotations were integrated with the existing model predictions in the training zone described in Section 5.2.1 by overlaying them onto the predicted masks. The student model was then retrained from scratch to include the new *Sea cucumber* class. Given that sea cucumbers are small organisms occupying a limited number of pixels in aerial images, their representation in the training data was proportionally low. To address this imbalance and ensure the model could effectively learn to detect them, we adjusted the loss function by applying a weighted Dice loss, assigning a higher penalty (seven times the weight of other classes) to errors involving sea cucumbers. The training process did not involve fine-tuning from previous weights to avoid bias toward earlier class distributions and to assess whether the inclusion of new categories interferes with performance on

existing ones.

The model was tested on the same test zones as before, specifically the *Trou d'eau* and *Saint-Leu* lagoons, by adding the *Sea cucumber* class to the ground truth annotation masks. Model evaluation showed that the new class was detected with high accuracy:

- Pixel Accuracy: 0.7545
- IoU: 0.4817

demonstrating the model's ability to learn from a small number of manual annotations.

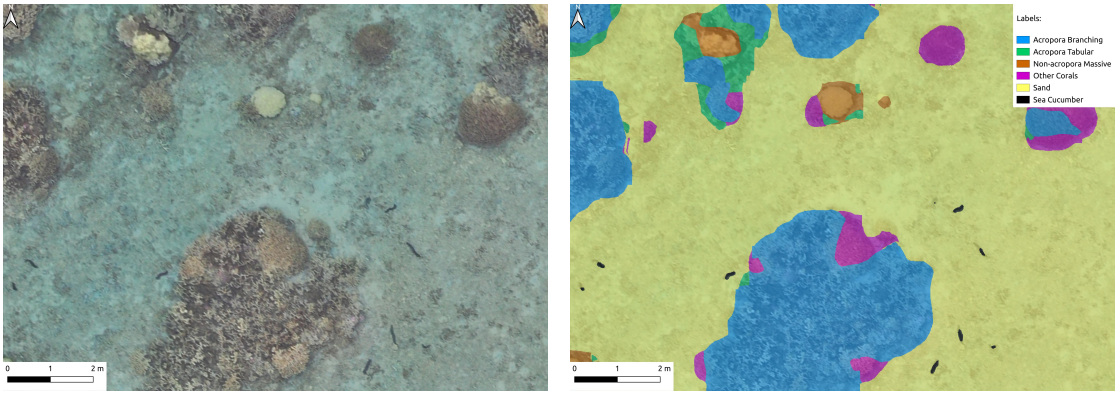
To assess whether introducing *Sea cucumber* impacted the model's performance on other classes, we compared standard metrics with those from the original 5-class version. As shown in Table 5.3, a decrease in overall accuracy (from 85.72% to 83.35%) and mean IoU (from 50.10% to 44.48%) was observed, indicating a modest reduction in segmentation performance after the addition of the new class.

Table 5.3: *IoU per class, total accuracy and mean IoU with the extended 6-class model including Sea cucumber.*

Zone	Accuracy	Mean IoU	Acropora B.	Acropora T.	Non-acro M.	Other Corals	Sand	Sea Cucumber
Total	**0.8335**	**0.4448**	**0.4638**	**0.1560**	**0.3905**	**0.2511**	**0.9254**	**0.4817**

The IoU scores for all coral-related classes declined: for *Other Corals* (from 0.4111 to 0.2511) and *Acropora Tabular* (from 0.2313 to 0.1560). Similarly, *Acropora Branching* and *Non-acropora Massive* saw moderate reductions (from 0.5282 to 0.4638 and 0.4095 to 0.3905, respectively). These drops suggest that the integration of a new, visually distinct class, such as *Sea cucumber*, may have introduced some degree of class interference, likely due to the high penalty weighting assigned to it in the loss function. Importantly, the performance on the *Sand* class remained stable (0.9249 vs 0.9254), suggesting that well-represented, visually dominant classes are more robust to such extensions.

Figure 5.11a illustrates a zoomed-in view of the UAV orthophoto of the *Trou d'eau* lagoon showing the presence of sea cucumbers and in Figure 5.11b the corresponding predicted segmentation mask. Some key ecological information can be extracted from such maps, including the spatial density of individuals per square metre and estimates of the size of the organism. For instance, the average length of each detected sea cucumber can be approximated by measuring the longest axis within each predicted segmentation mask. In the back reef depression of the *Trou d'eau* lagoon, this method yielded an average length of 19.6 ± 0.24 cm, calculated over 742 predicted individuals. These estimates are consistent with *in situ* measurements reported in Pierrat et al. (2024), where average lengths of 24.6 ± 0.5 cm and 12.5 ± 0.7 cm were recorded for the sea cucumber species *Holothuria leucospilota* and *Stichopus chloronotus*, respectively. Based on 72 individuals per species (144 in total), this yields a combined average length of 18.5 cm, in close agreement with our estimate.



(a) *(b)*

Figure 5.11: *Example of the segmentation output for the Sea cucumber class. (a) Zoomed-in view of the UAV orthophoto of the Trou d'eau lagoon showing the presence of sea cucumbers. (b) Corresponding predicted segmentation mask, showing the spatial distribution of sea cucumbers.*

5.4.4 Comparing multi-label classification and semantic segmentation approaches

In Chapter 4, a multi-label classification framework was adopted to transfer knowledge from ASV-based underwater predictions to UAV-based aerial imagery. This method associates each aerial tile with a list of predicted classes, indicating which coral morphotypes or habitats are likely present in the image, without specifying their exact location. In contrast, the current chapter introduces a semantic segmentation approach, where the model must assign a class label to every pixel in the image.

One of the main advantages of the multi-label classification framework is its flexibility in data collection. Since this method requires only the presence or absence of classes at the tile level, the image dataset does not need to be densely distributed over a specific area. As long as the underwater images are collected across representative zones and sufficiently cover each aerial tile, the model can learn to infer class presence on UAV orthophotos. This makes the upscaling process easier to deploy, especially when fine scale data is limited or hard to collect. In comparison, the semantic segmentation approach demands a much denser distribution of underwater images over the area of interest. This is because generating weak segmentation masks from ASV predictions involves interpolating spatially continuous rasters. To reliably locate where each class appears within a tile, a complete and high-resolution coverage of the zone is recommended.

Another important difference lies in how class context is transferred from underwater to aerial imagery. In multi-label classification, the aerial model can learn to associate contextual information with the presence of a class, even if the class is not clearly visible on the image. For example, certain coral types like *Non-acropora Encrusting* may not be directly visible in the aerial image, but their presence can be inferred from surrounding

textures, structures, or typical spatial patterns at the tile level. Semantic segmentation, on the other hand, requires the model to assign a class to each pixel. This makes it harder to transfer class presence if the class is not clearly visible, since there is no straightforward way to represent contextual cues in pixel-wise annotations.

Despite these limitations, semantic segmentation provides a clear advantage in terms of spatial details since it allows for precise mapping of coral morphotypes and enables the computation of spatial coverages or counting of marine organisms. This level of detail is essential for certain ecological applications, such as tracking coral recovery (Zhang et al., 2022). Furthermore, as demonstrated in Section 5.4.3, segmentation masks enable the extraction of additional metrics beyond simple presence/absence, such as organism density and size distributions.

In summary, the multi-label classification approach offers greater flexibility and simplicity in deployment, especially in scenarios with sparse underwater data. Semantic segmentation, while more demanding in terms of data and annotation requirements, enables richer and more spatially explicit predictions that are better suited for detailed ecological analyses. These two approaches are therefore complementary and the choice between them depends on the monitoring objectives and data availability.

Data availability

All code for data processing associated with the current submission is available on the-point-is-the-mask Github. The data that support the findings of this study are openly available in Zenodo at Seatizen Atlas (Contini et al., 2024a) and described in Contini et al. (2025e). ASV and UAV sessions used for training and testing in this study are available for download either via the zenodo-tools Github or will be automatically retrieved when running the main repository the-point-is-the-mask Github.

Conclusion and continuity

This study introduces a multi-scale WSSS framework for coral reef monitoring that transfers fine-scale information from underwater imagery to large-scale aerial assessments. Using dense ASV image collections and a DL classification model, we generated spatially interpolated probability maps that serve as weak annotations for training UAV-based segmentation models, eliminating the need for pixel-level labels.

Despite coarse supervision, the method enables accurate segmentation of key benthic classes and captures ecologically relevant spatial patterns. Region-based loss functions and model refinement via self-distillation further improve segmentation performance.

The approach is flexible and extensible. New classes can be incorporated either by collecting a small number of additional ASV sessions in areas where these classes are present or by manually annotating a limited number of instances directly in aerial imagery, as demonstrated for sea cucumbers.

Beyond UAV-based coral mapping, this multi-scale knowledge transfer strategy offers broader potential for hierarchical remote sensing. The same approach could be extended to transfer information from UAV orthophotos to airborne or satellite imagery, enabling the creation of large-scale coral reef maps that combine local ecological detail with regional coverage.

This framework offers a practical and cost-effective solution for ecological monitoring in complex and changing marine environments. By reducing annotation requirements and leveraging multi-scale data sources, it contributes to the development of robust, scalable tools for automated coral reef assessment.

Bibliography

Alonso, I., Yuval, M., Eyal, G., Treibitz, T., and Murillo, A. C. (2019). Coralseg: Learning coral segmentation from sparse annotations. *Journal of Field Robotics*, 36(8):1456–1477.

Azad, R., Heidary, M., Yilmaz, K., Hüttemann, M., Karimijafarbigloo, S., Wu, Y., Schmeink, A., and Merhof, D. (2023). Loss functions in the era of semantic segmentation: A survey and outlook. *arXiv preprint arXiv:2312.05391*.

Chan, L., Hosseini, M. S., and Plataniotis, K. N. (2021). A comprehensive analysis of weakly-supervised semantic segmentation in different image domains. *International Journal of Computer Vision*, 129(2):361–384.

Contini, M., Barde, J., Bonhommeau, S., Illien, V., and Joly, A. (2024a). Seatizen atlas.

Contini, M., Bonhommeau, S., Bernard, S., and Barde, J. (2025a). Aerial images collected by an unmanned aerial vehicle in st-leu, réunion - 2023-12-08 (processed_data).

Contini, M., Bonhommeau, S., Bernard, S., and Barde, J. (2025b). Aerial images collected by an unmanned aerial vehicle in trou-deau, réunion - 2023-12-02.

Contini, M., Bonhommeau, S., Bernard, S., and Barde, J. (2025c). Aerial images collected by an unmanned aerial vehicle in trou-deau, réunion - 2023-12-02 (processed_data).

Contini, M., Illien, V., Barde, J., Poulain, S., Bernard, S., Joly, A., and Bonhommeau, S. (2025d). From underwater to drone: A novel multi-scale knowledge distillation approach for coral reef monitoring. *Ecological Informatics*, page 103149.

Contini, M., Illien, V., Julien, M., Ravitchandirane, M., Russias, V., Lazennec, A., Chevrier, T., Rintz, C. L., Carpentier, L., Gogendeau, P., et al. (2025e). Seatizen atlas: a collaborative dataset of underwater and aerial marine imagery. *Scientific Data*, 12(1):67.

Contini, M., Illien, V., Poulain, S., Bernard, S., Barde, J., Bonhommeau, S., and Joly, A. (2025f). The point is the mask: scaling coral reef segmentation with weak supervision.

Contini, M., Leblanc, C., and Illien, V. (2024b). Dinovdeau-large-2024_04_03-with_data_aug_batch-size32_epochs150_freeze – Hugging Face. https://doi.org/10.57967/hf/2947.

Doukari, M. and Topouzelis, K. (2022). Overcoming the uas limitations in the coastal environment for accurate habitat mapping. *Remote Sensing Applications: Society and Environment*, 26:100726.

Huang, L., Jiang, B., Lv, S., Liu, Y., and Fu, Y. (2023). Deep-learning-based semantic segmentation of remote sensing images: A survey. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:8370–8396.

Kingma, D. P. and Ba, J. (2014). Adam: A method for stochastic optimization. *International Conference on Learning Representations*.

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., et al. (2023). Segment anything. In *Proceedings of the IEEE/CVF international conference on computer vision*, pages 4015–4026.

Lin, Y., Li, H., Shao, W., Yang, Z., Zhao, J., He, X., Luo, P., and Zhang, K. (2025). Samrefiner: Taming segment anything model for universal mask refinement. In *The Thirteenth International Conference on Learning Representations*.

Pierrat, J., Urbistondoy, L., Modi, A., Viramoutou, B., and Frouin, P. (2024). Searching for drivers of the patchy distribution of sympatric deposit-feeding sea cucumbers: A multi-scale monitoring study. *Limnology and Oceanography*, 69(9):2057–2070.

Qu, L., Song, X., Zhang, M., Wang, J., Wen, R., and Wang, S. (2024). Gan-based defogging and multiscale fusion approach for uav-based seagrass bed imagery semantic segmentation in challenging marine environments. In *International Conference of Pioneering Computer Scientists, Engineers and Educators*, pages 55–72. Springer.

Raine, S., Marchant, R., Kusy, B., Maire, F., and Fischer, T. (2022). Point label aware superpixels for multi-species segmentation of underwater imagery. *IEEE Robotics and Automation Letters*, 7(3):8291–8298.

Ru, L., Zhan, Y., Yu, B., and Du, B. (2022). Learning affinity from attention: End-to-end weakly-supervised semantic segmentation with transformers. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 16846–16855.

Sauder, J., Banc-Prandi, G., Meibom, A., and Tuia, D. (2024). Scalable semantic 3d mapping of coral reefs with deep learning. *Methods in Ecology and Evolution*, 15(5):916–934.

Sauder, J., Domazetoski, V., Banc-Prandi, G., Perna, G., Meibom, A., and Tuia, D. (2025). The coralscapes dataset: Semantic scene understanding in coral reefs. *arXiv preprint arXiv:2503.20000*.

Schlosser, T., Friedrich, M., Meyer, T., and Kowerko, D. (2024). A consolidated overview of evaluation and performance metrics for machine learning and computer vision. *Tobias Schlosser, Michael Friedrich, Trixy Meyer, and Danny Kowerko–Junior Professorship of Media Computing, Chemnitz University of Technology*, 9107.

Sudre, C. H., Li, W., Vercauteren, T., Ourselin, S., and Jorge Cardoso, M. (2017). 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.

Wang, J., Wang, Y., Liu, L., Yin, H., Ye, N., and Xu, C. (2023). Weakly supervised forest fire segmentation in uav imagery based on foreground-aware pooling and context-aware loss. *Remote Sensing*, 15(14):3606.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., and Luo, P. (2021). Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34:12077–12090.

Xu, X., Li, M., Tao, C., Shen, T., Cheng, R., Li, J., Xu, C., Tao, D., and Zhou, T. (2024). A survey on knowledge distillation of large language models. *arXiv preprint arXiv:2402.13116*.

Zhang, H., Grün, A., and Li, M. (2022). Deep learning for semantic segmentation of coral images in underwater photogrammetry. *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 2:343–350.

Zhao, Y., Wong, L., and Goh, W. W. B. (2020). How to do quantile normalization correctly for gene expression data analyses. *Scientific reports*, 10(1):15534.

Zhao, Z., Fan, C., and Liu, L. (2023). Geo sam: a qgis plugin using segment anything model (sam) to accelerate geospatial image segmentation. *Zenodo. Available from: https://doi. org/10.5281/zenodo*, 8191039(10969837):0.

Zheng, Z., Liang, H., Hua, B.-S., Wong, Y. H., Ang, P., Chui, A. P. Y., and Yeung, S.-K. (2024). Coralscop: segment any coral image on this planet. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 28170–28180.

Zhong, J., Li, M., Zhang, H., and Qin, J. (2023). Combining photogrammetric computer vision and semantic segmentation for fine-grained understanding of coral reef growth under climate change. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 186–195.

Øie Kvile, K., Gundersen, H., Poulsen, R. N., Sample, J. E., Salberg, A.-B., Ghareeb, M. E., Buls, T., Bekkby, T., and Hancke, K. (2024). Drone and ground-truth data collection, image annotation and machine learning: A protocol for coastal habitat mapping and classification. *MethodsX*, 13:102935.

Chapter 6

Assessing the potential of upscaling from low- to high-altitude aerial imagery

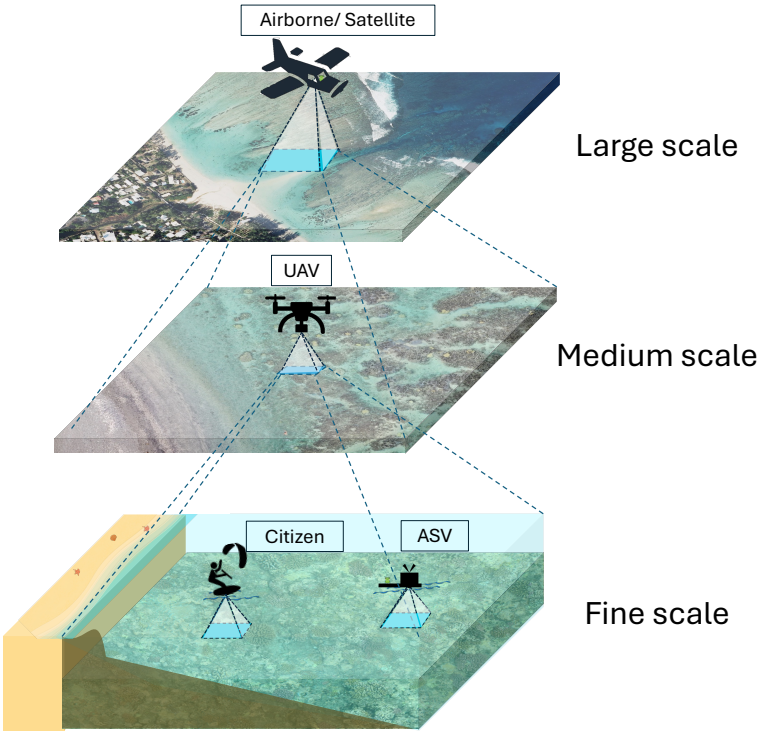


Figure 6.1: *Workflow of the thesis, showing how underwater images and low-altitude aerial imagery are progressively upscaled to high-altitude aerial imagery.*
Image credit: Letizia Lamperti.

This chapter presents an exploratory study conducted during the concluding phase of this doctoral research. Unlike previous chapters based on published articles, this work has not yet been submitted for peer review. It aims to evaluate the feasibility of scaling ecological annotation from low-altitude aerial imagery to high-altitude aerial imagery as part of the multi-scale coral reef monitoring framework.

Context and objectives

While underwater and Unmanned Aerial Vehicle (UAV)-based observations offer high detail, their limited spatial extent restricts their utility for large-scale conservation planning. In Chapter 5, we presented a Weakly Supervised Semantic Segmentation (WSSS) framework that transfers benthic predictions from underwater imagery to UAV orthophotos. This approach enabled the generation of fine-scale benthic habitat maps. However, UAV imagery is inherently constrained by flight duration and weather conditions, making it difficult to scale beyond localized surveys. To overcome these constraints, the next logical step is to integrate high-altitude aerial imagery, such as the orthophotos provided by the national French National Institute of Geographic and Forest Information (IGN) *BD Ortho* program at $20\,\mathrm{cm\,px^{-1}}$ resolution.

This chapter builds on the multi-scale WSSS framework introduced in Chapter 5, extending it from UAV to high-altitude aerial imagery. Specifically, We use the UAV-based predictions as high-resolution annotations for training a segmentation model on the coarser-resolution orthophotos. By aligning UAV and IGN imagery through photogrammetric and geospatial techniques, we establish a training dataset that allows ecological information to be propagated over larger spatial extents.

This chapter aims to demonstrate the feasibility of a complete upscaling pipeline that transfers ecological information from underwater imagery to satellite-resolution platforms, using UAV-collected data as an intermediate layer. Similar to the upscaling from Autonomous Surface Vehicle (ASV) images to aerial drone imagery, this step also reduces the number of classes that can be reliably detected.

Our case study focuses on the Hermitage / La Saline reef system in Réunion Island and explore how this framework can serve as a foundation for broader applications involving satellite-level observations.

6.1 Introduction

Modern Earth observation systems deliver high-resolution imagery with frequent revisit times, producing satellite image time series that are widely used in ecology, agriculture and habitat mapping (Interdonato et al., 2019; Ienco et al., 2019). In coral reef research, multispectral and hyperspectral imagery from satellites and aerial platforms has enabled habitat classification, coral cover estimation and temporal change detection (Lazuardi et al., 2021; Burns et al., 2022).

At the global scale, the Allen Coral Atlas initiative mapped coral reef habitats using PlanetScope satellite imagery, with $3.1\,\mathrm{m\,px^{-1}}$ resolution. This effort has enabled the production of benthic and geomorphic habitat maps worldwide, offering a baseline for global reef monitoring. However, the relatively coarse spatial resolution of PlanetScope imagery limits reliable discrimination to broad categories such as rock, sand, rubble, seagrass, coral and algae. As a result, classification performance is degraded in many reef areas, limiting the ecological relevance of these maps for local management and conservation applications (Serge et al., 2024).

In Réunion Island, hyperspectral airborne imagery has been used for over a decade to produce detailed reef maps and assess coral cover on fringing reefs (Bajjouk et al., 2019; Mouquet et al., 2025). However, such campaigns remain costly and difficult to repeat regularly. Recent advances include the availability of Very High Resolution (VHR) satellite imagery such as Pléiades Neo, which offers $30\,\mathrm{cm\,px^{-1}}$ resolution in the panchromatic band (a single broad-spectrum greyscale channel) and national airborne orthophotography programs like IGN *BD Ortho* that provide new opportunities for routine, cost-effective reef monitoring at broad scales (Mouquet et al., 2025). IGN provides high-resolution imagery for each French department (including Réunion Island), captured at intervals ranging between 2 and 4 years. The image resolution varies over time, ranging from $50\,\mathrm{cm\,px^{-1}}$ for most departments before 2014 to $15\,\mathrm{cm\,px^{-1}}$ in later years. In recent work, airborne images from the *BD Ortho* series were successfully used to detect and map marine aquaculture cages along the French Mediterranean coast using object detection techniques (Quaade et al., 2024). This result demonstrates the potential of IGN imagery to support automated, large-scale monitoring of marine structures using high-altitude imagery.

Beyond image resolution, a major source of classification error in satellite-based reef mapping is the absence of ecologically fine-grained training data. When coupled with underwater imagery, accuracy can be improved through data fusion techniques (Nguyen et al., 2021). Lutzenkirchen et al. (2024) conducted a systematic literature review comparing 199 approaches used to quantify benthic habitat. Among these, only 3% integrated both "in-situ" observations with broad-scale remote sensing. The authors proposed, as a promising direction for future research, a multi-scale framework that could leverage a hierarchy of sensors (from underwater data to UAV and satellite imagery) to bridge the gap between large spatial coverage and fine-scale resolution. This conceptual approach is illustrated in Figure 6.2.

Upscaling ecological information across spatial resolutions poses technical as well as

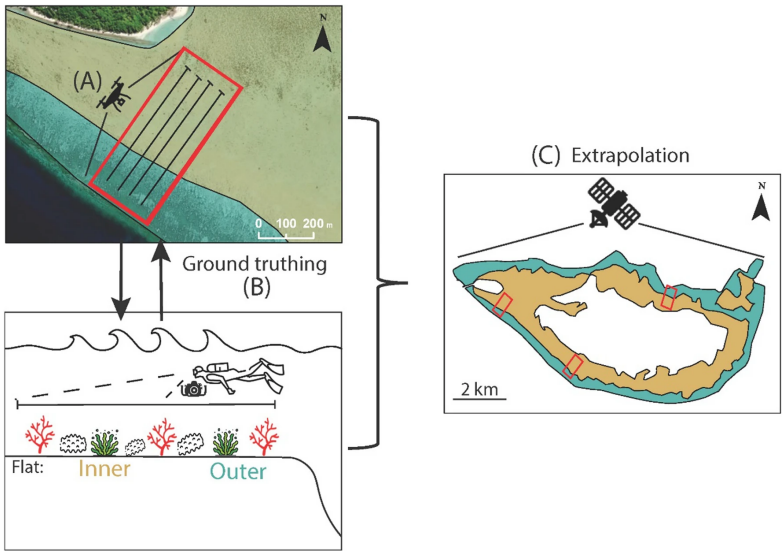


Figure 6.2: *Schematic summary outlining a potential approach using in-situ assessments, as well as multiple layers of sensors of varying resolution, to bridge scientific disciplines and upscale observations onto reef-scape scales. After multiple shallow reef areas are surveyed using a 'small-area-high resolution' sensor (A) and ground-truthed using in-situ assessments and sub-cm resolution in-water remote sensing technology (B), observations can be extrapolated and upscaled using a 'large-area-lower-resolution' sensor (C). Figure taken from (Lutzenkirchen et al., 2024).*

conceptual challenges. Differences in pixel size, spectral bands and acquisition conditions create domain shifts that hinder the direct transfer of models between platforms. Bridging these gaps requires strategies such as weak supervision, combined with accurate georeferencing and photogrammetric alignment, to ensure that fine-scale annotations can be meaningfully transferred across spatial scales.

In Chapter 5, we implemented a WSSS framework to transfer ecological information from underwater imagery to orthophotos using sparse annotations. In this chapter, we extend this framework to upscale coral reef classification from UAV-based orthophotos to high-altitude airborne imagery. After training a WSSS model on UAV orthophotos informed by underwater imagery, we use the resulting UAV-based predictions as high-resolution annotations for corresponding IGN *BD Ortho* tiles covering the same geographic areas. A model is then trained directly on the *BD Ortho* images using these weak labels. This approach enables the transfer of fine-scale ecological information onto large-scale aerial imagery, supporting high-resolution reef habitat mapping over entire reef systems.

We focus on the fringing reefs of Réunion Island, using the Hermitage reef as a case study. By combining UAV-derived information with the broad coverage of IGN *BD Ortho*, we demonstrate a scalable framework for continuous coral reef mapping that could be extended to VHR satellite imagery such as Pléiades Neo. To the best

of our knowledge, this is the first study to implement a full upscaling pipeline that transfers Artificial Intelligence (AI)-based underwater benthic classification to satellite-level imagery, using UAV data as an intermediate spatial bridge.

6.2 Materials and Methods

6.2.1 Aerial imagery data sources

Low-altitude aerial data

Aerial imagery was acquired using a *DJI Mavic 2 Pro* drone. A high-resolution UAV orthophoto was generated by mosaicking individual images through Structure from Motion (SfM) workflow. As the drone was not equipped with a differential GPS, georeferencing accuracy was improved by collecting Ground Control Points (GCPs) in the surveyed area with a differential GPS and applying them during orthophoto alignment. The resulting UAV orthophotos are accessible through an interactive web application[1], which allows users to explore the data spatially, filter datasets by date or platform, and initiate direct downloads.

Two UAV missions were conducted in the lagoons of Réunion Island: one in the *Saint-Leu* lagoon Contini et al. (2025a) and the other in the *Trou d'eau* lagoon Contini et al. (2025b), covering 204,748 m^2 and 189,682 m^2, respectively. The resulting orthophotos had a Ground Sampling Distance (GSD) of 0.9 cm for *Trou d'eau* and 1.6 cm for *Saint-Leu*, reflecting differences in flight altitude and mission configuration.

High-altitude aerial data

IGN produces the *BD Ortho* series, a nationwide aerial orthophotography database that covers the entire French territory (including overseas departments such as La Réunion), freely available for scientific and public use. These images are orthorectified to correct distortions from terrain relief and sensor perspective, ensuring accurate geospatial alignment. Acquisition is carried out for each French department every 2 to 4 years, with spatial resolution improving from 50 cm px^{-1} before 2014 to 15 cm px^{-1} in recent campaigns.

For this study, we used *BD Ortho* datasets from 2017 and 2022, both covering La Réunion at 20 cm px^{-1} resolution. These datasets provide consistent, high-quality imagery suitable for large-scale coral reef mapping. Their regular acquisition cycle and standardized processing make them valuable for temporal analyses of reef habitat change.

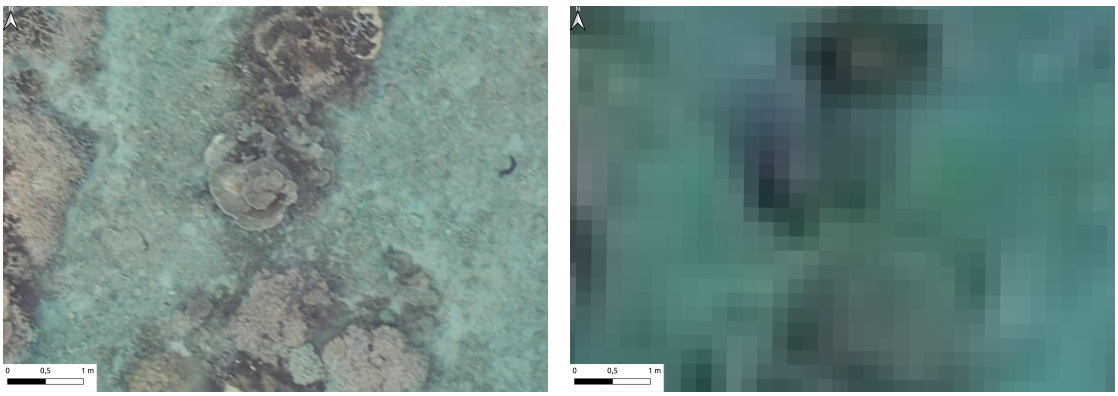
6.2.2 Semantic segmentation

Class selection

For the UAV-to-IGN upscaling step, the class set was adapted to account for the reduced spatial resolution and spectral characteristics of the *BD Ortho* aerial imagery compared

[1] https://seatizenmonitoring.ifremer.re/home Credit: Victor Illien

to UAV orthophotos. While the UAV teacher model was trained on five classes (see Section 5.2.4), not all of these classes could be reliably distinguished in the IGN imagery.



(a) *(b)*

Figure 6.3: *Example of an Acropora tabular colony at the centre of the images, observed in (a) a high-resolution UAV orthophoto and (b) the corresponding IGN aerial imagery at* $20\,\mathrm{cm\,px^{-1}}$*. The colony is clearly identifiable in UAV data but its structure is no longer distinguishable in IGN imagery.*

In particular, as shown in Figure 6.3, the *Acropora Tabular* class could not be reliably distinguished from *Other Corals* due to the limited spatial resolution and the loss of fine-scale colony structure in the $20\,\mathrm{cm\,px^{-1}}$ IGN images. As illustrated in Figure 6.3a, *Acropora Tabular* colonies are clearly identifiable in UAV orthophotos, whereas in the corresponding IGN imagery (Figure 6.3b) their structure cannot be distinguished. As a result, *Acropora Tabular* instances were merged into the *Other Corals* category for the upscaling step.

Visual inspection of the 2017 and 2022 *BD Ortho* datasets revealed extensive *Syringodium isoetifolium* seagrass meadows that are no longer visible in recent imagery of Réunion lagoons. These historically important habitats (Cuvillier et al., 2017) were systematically misclassified during inference, as the *Syringodium isoetifolium* class was absent from the training dataset. Its exclusion stems from the fact that this seagrass was not observed during the 2023 UAV surveys, due to its recent and drastic decline. Nonetheless, the class is clearly detectable at the IGN imagery scale on 2017 and 2022 orthophotos and represents a key benthic component that must be accounted for in temporal analyses.

To capture this ecologically important habitat, we manually annotated the seagrass patches in the 2017 and 2022 images. 2017 annotations were integrated into the training set by overlaying them onto the initial model's predictions, whereas 2022 annotations were reserved for testing to evaluate the model's performance on unseen data.

The final set of classes retained for the UAV-to-IGN upscaling is:

1. *Acropora Branching*
2. *Non-acropora Massive*
3. *Other Corals*
4. *Sand*
5. *Syringodium isoetifolium*

Segmentation model

The model was trained using UAV-based predictions as weak annotations for IGN tiles covering the same geographic zones, as illustrated in the thesis workflow diagram in Figure 6.1. Since UAV predictions were only available for surveys conducted in 2023, they were used as weak labels for IGN imagery from both 2017 and 2022. This choice was made to increase the diversity of training data and expose the model to different image acquisition conditions. To construct the training dataset, IGN tiles were extracted with a horizontal and vertical overlap of 75% between consecutive tiles. This procedure resulted in a total of 846 images of size 256×256 pixels.

The Segformer (Xie et al., 2021) model was trained with a batch size of 16 images. The learning rate was initially set to 10^{-5} and decayed by a factor of 10 whenever the validation loss plateaued for five consecutive epochs.

After the first training phase, predictions were generated for the 2017 and 2022 *BD Ortho* tiles and annotations for the *Syringodium isoetifolium* class were added on top of these predictions. Tiles from 2017 intersecting with the *Syringodium isoetifolium* patches were then extracted and used to retrain the model from scratch following a self-knowledge distillation strategy, where predictions are used to refine subsequent training. The final model was trained on 1,016 tiles, consisting of the original 846 tiles plus 170 tiles containing the *Syringodium isoetifolium* class.

6.3 Results

6.3.1 High-altitude model evaluation

To evaluate the performance of the model trained on IGN data, we employed two complementary evaluation approaches. The first involved comparing the model's predictions to those generated by the UAV model, using standard semantic segmentation metrics. The second evaluation used human-annotated ground truth in three test zones: the two introduced in Section 5.2.7 and a new 2022 zone with *Syringodium isoetifolium* annotations to assess performance on this added class.

Comparing the model predictions with the UAV model annotations on the test set, we obtained a global accuracy of 0.8303 and a mean Intersection over Union (IoU) of 0.5656. Global and per-class metrics, computed on the test set, are summarized in Table 6.1.

To assess the model's performance against human-labelled ground truth, we compared its predictions on the *Trou d'Eau*, *Saint-Leu* and *Hermitage* test zones. The resulting confusion matrix, presented in Figure 6.4, shows class-wise prediction accuracy and misclassification patterns between classes.

Table 6.1: *Overall accuracy, mean IoU and per-class IoU scores for the IGN model on the test set*

Accuracy	Mean IoU	Acropora B.	Non-acro M.	Other Corals	Sand	Seagrass
0.8303	**0.5656**	**0.7098**	**0.2178**	**0.2842**	**0.8294**	**0.7867**



Figure 6.4: *Normalized confusion matrix computed across the test zones: Trou d'Eau, Saint-Leu and Hermitage.*

6.3.2 Predicted habitat composition in the Hermitage / La Saline lagoon

To assess the spatial consistency and ecological relevance of the segmentation predictions across large reef areas, we applied the trained model to the *BD Ortho* imagery covering the Hermitage / La Saline lagoon, measuring approximately 2.8 km^2.

A broad view of the segmentation results is presented in Figure 6.5, where the model's predictions are shown for the entire lagoon area. In this figure the polygon boundary of one of the two UAV surveys presented in the previous chapter and used to train the model is shown in orange, illustrating the scale difference between low-altitude and high-altitude aerial platforms.

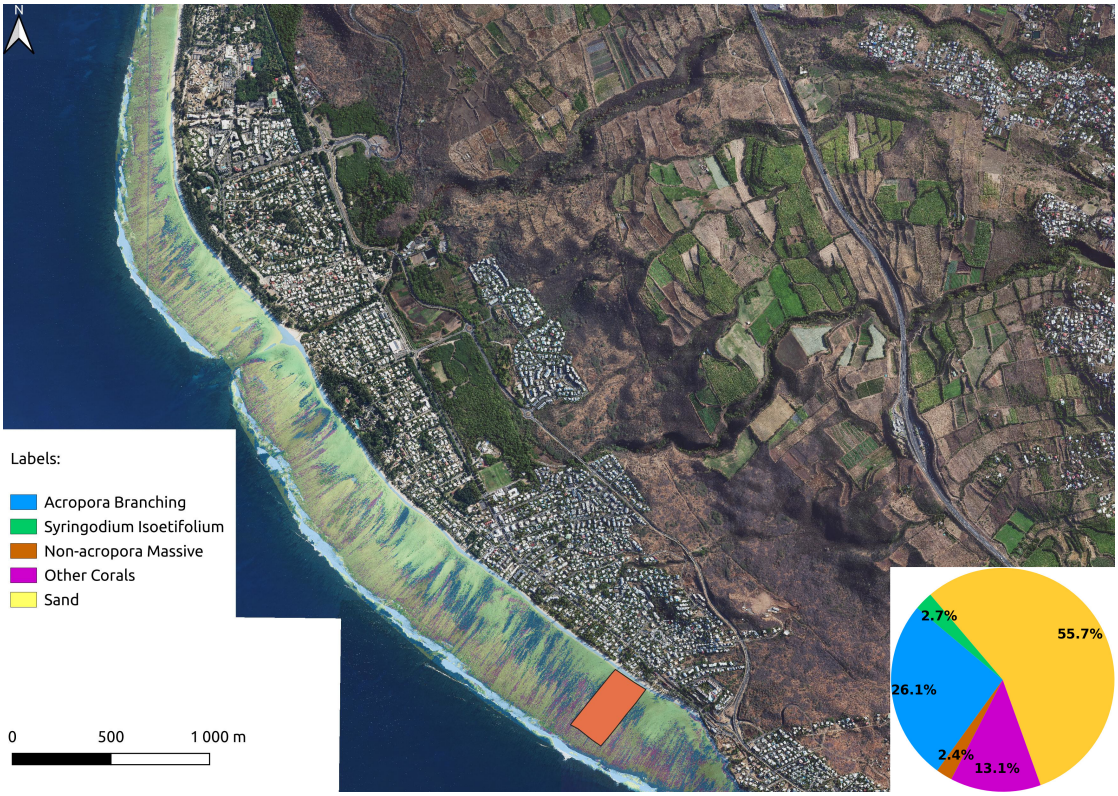
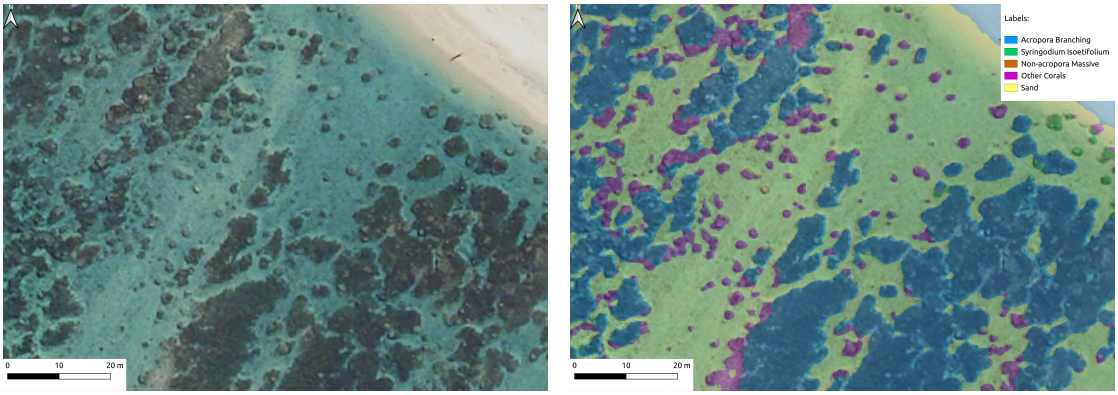


Figure 6.5: *Predicted segmentation mask for the entire Hermitage / La Saline lagoon area, covering approximately* $2.8\,\mathrm{km}^2$*. The orange polygon marks the extent of the Trou d'Eau UAV survey, highlighting the scale differences between low- and high-altitude platforms*

In Figure 6.6, we present a zoomed-in view of the segmentation results, showing the model's predictions for the five classes defined in Section 6.2.2 in the Hermitage lagoon.

6.4 Discussion

Our findings demonstrate the feasibility of leveraging UAV-derived annotations for training segmentation models on high-altitude aerial imagery. Despite some misclassification errors observed in Figure 6.6, the model shows a reasonable ability to distinguish between the simplified set of coral morphotypes considered in this study. Indeed, as shown in the aerial image in Figure 6.6a, even expert observers may struggle to confidently determine whether dark patches correspond to *Acropora Branching* or *Other Corals*, or whether isolated circular colonies belong to the *Non-acropora Massive* or *Other Corals* class, without additional context. Because the model was trained on UAV-derived annotations (where such distinctions are more clearly visible), it learned to extract and generalize relevant features for coral morphotype classification at the coarser IGN resolution.



(a) *(b)*

Figure 6.6: *Example of the segmentation output in the Hermitage lagoon. (a) Aerial view of the IGN orthophoto of the Hermitage lagoon. (b) Corresponding predicted segmentation mask.*

These initial observations are confirmed by quantitative evaluation on the test set, where the IGN model achieved a global accuracy of 0.8303 and a mean IoU of 0.5656, demonstrating its ability to capture the main patterns of UAV-based predictions. The confusion matrix in Figure 6.4 provides additional insight. As observed in the previous chapter, the best-recognized classes are *Acropora Branching* and *Sand*, which aligns with the fact that these were also the classes with highest performance in the IGN student model. This consistency confirms that the upscaling process preserves the relative performance trends between classes. Interestingly, 39.4% of *Non-acropora Massive* samples are correctly classified, which is noteworthy given the difficulty (even for human experts) of distinguishing *Non-acropora Massive* coral colonies from the massive corals belonging to the *Other Corals* category in aerial imagery. Finally, the model identified 81.5% of *Syringodium isoetifolium* instances correctly, a result that can be explained by its distinct colour and circular meadow structures, which stand out clearly in IGN imagery.

These results are particularly encouraging given that the model was trained using only two UAV surveys covering a total area of 394,430 m^2.

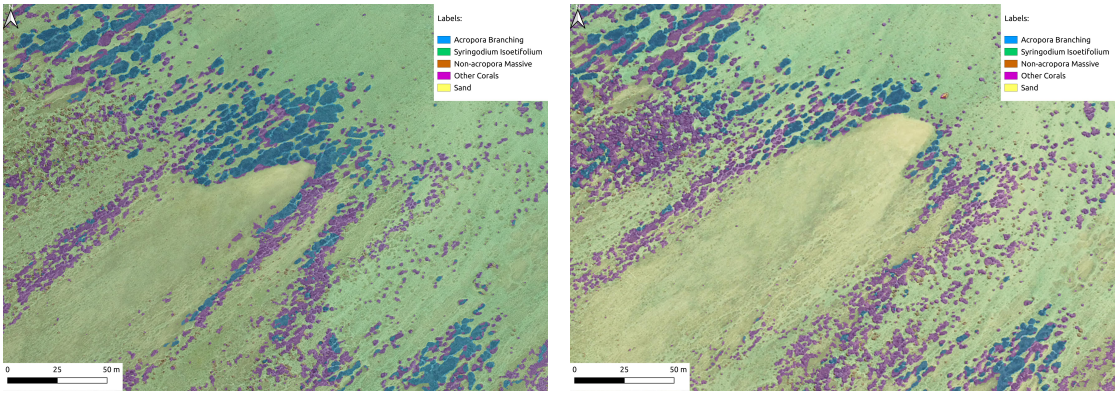
This approach could be incorporated into long-term coral cover estimation frameworks based on multitemporal VHR satellite archives, as demonstrated by Mouquet et al. (2025) in a recent work using SPOT 6/7 and Pléiades time series over the Hermitage reef. Moreover, while these methods provide valuable metrics of coral cover evolution, our segmentation-based approach offers additional structural detail and morphotype-level classification, which can further improve the ecological interpretation of these results.

While the results are encouraging, it is important to acknowledge the limitations of the current approach. UAV annotations used for weak supervision were acquired in 2023, while the IGN aerial images used for training come from 2017 and 2022. Although it is reasonable to assume that no major structural changes occurred in coral reef composition between 2022 and 2023 at the 20 cm px^{-1} resolution of the IGN imagery, this assumption

is less valid for the 2017 dataset. In that case, the annotations are likely to introduce temporal noise, as ecological shifts may have occurred over the six-year gap.

Despite this temporal mismatch, the 2017 imagery was deliberately included to enrich the training dataset with more diverse imaging conditions (such as variations in illumination and acquisition parameters). Moreover, as shown in the previous chapter, the SegFormer architecture combined with the Dice loss function has proven robust to weak and noisy annotations (including those affected by temporal uncertainty). This robustness allows us to benefit from a broader range of training images, even if the annotations are not perfectly time-aligned. That said, using IGN images and UAV-derived annotations acquired within the same year would certainly improve consistency and reduce label noise, which is expected to improve overall segmentation accuracy.

An illustrative example of ecological change detectable in the IGN imagery is the progression of a sand tongue on the Hermitage reef flat (Figure 6.7). In Figure 6.7a, the sand tongue was still relatively confined in 2017, whereas in Figure 6.7b from 2022, light-coloured coral debris deposits had advanced further across the platform replacing darker living coral habitats. Beyond visual interpretation, this case illustrates the possibility of quantifying fine-scale changes in reef habitats using high-altitude imagery. This observation is consistent with other studies reporting the progression of rubble tongues on Réunion's reef flats (Mouquet et al., 2025).



(a) ***(b)***

Figure 6.7: *Progression of a sand tongue (in yellow) on the Hermitage reef flat between 2017 (a) and 2022 (b). The imagery highlights the advance of coral debris deposits across the reef platform, occurring at the expense of living coral habitats (Acropora Branching, Non-acropora Massive, and Other Corals).*

Future work could explore the integration of VHR satellite imagery with frequent revisit rates (often ranging from once to twice per day) into the upscaling pipeline presented in this chapter. Such imagery offers two key advantages:

1. First, it would allow UAV surveys to be paired with satellite imagery acquired on the same day, eliminating temporal mismatches and reducing inconsistencies

between annotations and input images. This would directly improve the quality of training data and lead to more accurate prediction models.

2. Second, high-frequency VHR imagery would enable near-daily monitoring of coral reef ecosystems at very large scales. This capability could support monitoring rubble tongue progression in lagoons, assessing live coral cover decline and observing reef structural complexity loss (Mouquet et al., 2025).

Finally, brightness indices derived from blue and green spectral bands in VHR satellite imagery have been shown to correlate well with living coral cover in La Réunion, particularly when using Pléiades Neo data (Mouquet et al., 2025). Rather than relying solely on manually collected ground truth or on hyperspectral campaigns for coral health annotations (which are expensive and difficult to repeat), future work could leverage automated annotations of living coral cover generated from UAV and ASV surveys. These annotations could be used to train AI models directly on brightness indices in VHR satellite imagery, enabling scalable estimation of living coral cover.

Conclusion and continuity

This chapter extended the multi-scale WSSS framework by upscaling UAV-based coral predictions to IGN *BD Ortho* orthophotos. Building on outputs from the Chapter 5 model, we trained a semantic segmentation network capable of operating on 20 cm px^{-1} resolution imagery.

Despite the weak supervision and temporal inconsistencies between the UAV annotations (2023) and the aerial imagery (2017 and 2022), the model achieved promising results. It was able to accurately distinguish key benthic classes, including coral morphotypes and seagrass meadows and demonstrated strong generalization capabilities when evaluated on three test zones.

This upscaling approach enables the use of historical orthophoto archives for retrospective reef monitoring, thus facilitating long-term analyses of benthic habitat dynamics.

As with the approach described in the previous chapter, this method is generalizable and modular. Additional habitat types, such as seagrasses, can be integrated through targeted annotations. Integrating VHR satellite imagery, with its broader coverage and higher revisit frequency, could further extend this framework to enable coral reef monitoring at regional and global scales.

Bibliography

Bajjouk, T., Mouquet, P., Ropert, M., Quod, J.-P., Hoarau, L., Bigot, L., Le Dantec, N., Delacourt, C., and Populus, J. (2019). Detection of changes in shallow coral reefs status: Towards a spatial approach using hyperspectral and multispectral data. *Ecological Indicators*, 96:174–191.

Burns, C., Bollard, B., and Narayanan, A. (2022). Machine-learning for mapping and monitoring shallow coral reef habitats. *Remote Sensing*, 14(11):2666.

Contini, M., Bonhommeau, S., Bernard, S., and Barde, J. (2025a). Aerial images collected by an unmanned aerial vehicle in st-leu, réunion - 2023-12-08 (processed_data).

Contini, M., Bonhommeau, S., Bernard, S., and Barde, J. (2025b). Aerial images collected by an unmanned aerial vehicle in trou-deau, réunion - 2023-12-02 (processed_data).

Cuvillier, A., Villeneuve, N., Cordier, E., Kolasinski, J., Maurel, L., Farnier, N., and Frouin, P. (2017). Causes of seasonal and decadal variability in a tropical seagrass seascape (reunion island, south western indian ocean). *Estuarine, Coastal and Shelf Science*, 184:90–101.

Ienco, D., Interdonato, R., Gaetano, R., and Minh, D. H. T. (2019). Combining sentinel-1 and sentinel-2 satellite image time series for land cover mapping via a multi-source deep learning architecture. *ISPRS Journal of Photogrammetry and Remote Sensing*, 158:11–22.

Interdonato, R., Ienco, D., Gaetano, R., and Ose, K. (2019). Duplo: A dual view point deep learning architecture for time series classification. *ISPRS journal of photogrammetry and remote sensing*, 149:91–104.

Lazuardi, W., Wicaksono, P., and Marfai, M. (2021). Remote sensing for coral reef and seagrass cover mapping to support coastal management of small islands. In *IOP conference series: earth and environmental science*, volume 686, page 012031. IOP Publishing.

Lutzenkirchen, L. L., Duce, S. J., and Bellwood, D. R. (2024). Exploring benthic habitat assessments on coral reefs: a comparison of direct field measurements versus remote sensing. *Coral Reefs*, 43(2):265–280.

Mouquet, P., Schultz, M., Revillion, C., Cuet, P., and Pennober, G. (2025). Cartographie des coraux par imagerie satellite à très haute résolution. application aux données sentinel-2, spot6/7, pléiades et pléiades neo sur les récifs de la réunion. *Revue Française de Photogrammétrie et de Télédétection*, 226(1):45–61.

Nguyen, T., Liquet, B., Mengersen, K., and Sous, D. (2021). Mapping of coral reefs with multispectral satellites: A review of recent papers. *Remote Sensing*, 13(21):4470.

Quaade, S., Vallebueno, A., Alcabes, O. D., Rodolfa, K. T., and Ho, D. E. (2024). Remote sensing and computer vision for marine aquaculture. *Science Advances*, 10(42):eadn4944.

Serge, A., Maële, B., Stéphane, G., and Antoine, G. (2024). Evaluation of the allen coral atlas benthic habitat map product for new caledonia using representative habitat observations from a multi-species fishery assessment. *Coral Reefs*, 43(3):523–540.

Xie, E., Wang, W., Yu, Z., Anandkumar, A., Alvarez, J. M., and Luo, P. (2021). Segformer: Simple and efficient design for semantic segmentation with transformers. *Advances in neural information processing systems*, 34:12077–12090.

Chapter 7

Conclusion and perspectives



Figure 7.1: *Giant Tortoise (Aldabrachelys gigantea) on Aldabra Atoll, Seychelles, nicknamed "Biscuit". An emblematic example of fragile ecosystems that require long-term protection and monitoring.*
Image credit: Mervyn Ravitchandirane, October 2022.

7.1 Synthesis of results and contributions

Monitoring coral reef ecosystems relies on data acquisition across multiple spatial scales, from colonies observed by divers through Underwater Visual Census (UVC) to entire reef systems captured by satellites. However, a fundamental trade-off exists: platforms that cover larger areas generally provide images at lower spatial resolution, which limits the level of ecological detail and the classification accuracy achievable by Computer Vision (CV) methods.

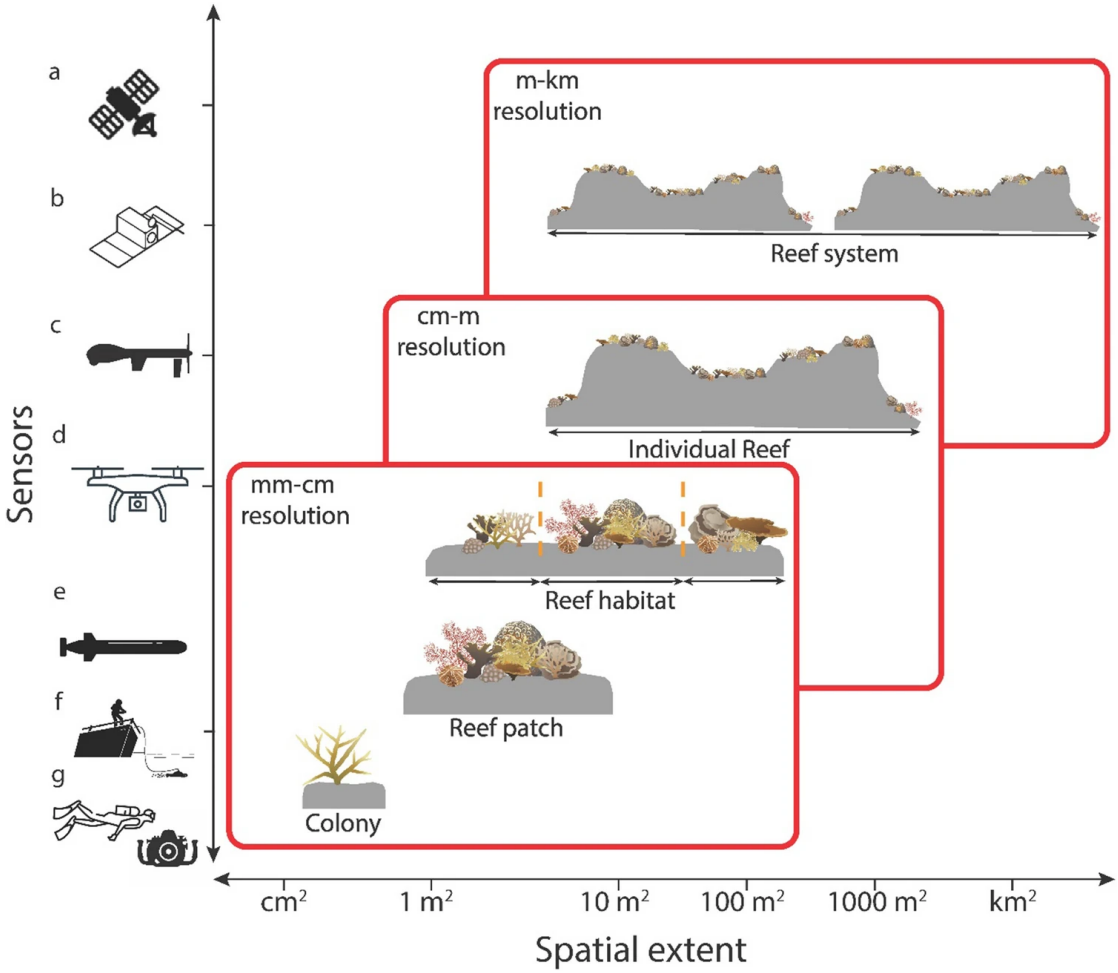


Figure 7.2: *Differences in spatial extent and resolution among technologies and approaches used in coral reef ecosystem research. Specifically, differences among satellite and aerial remote sensing platforms (i.e., (a) satellite, (b) aeroplane, (c) unoccupied aerial vehicles and (d) drones); and in-water remote sensing platforms, including (e) autonomous underwater vehicles, (f) underwater drones and (g) diver-operated camera systems. Figure taken from (Lutzenkirchen et al., 2024).*

Conversely, as illustrated in Figure 7.2, platforms operating at fine resolution such as diver-operated cameras or Autonomous Surface Vehicles (ASVs), offer detailed imagery but cover only limited areas. This trade-off also exists for the temporal aspect as field surveys are difficult to set up on a regular basis while satellite images can be collected frequently. This thesis tackled the challenge of developing a cost-effective multi-scale mapping framework that links underwater and aerial observations for the ecological assessment of coral reef lagoons. By integrating data collected at different resolutions and spatial extents, this work aims to bridge the gap between fine-scale benthic observations and large-scale habitat mapping, enabling more complete and reproducible monitoring strategies.

Throughout the work, several key scientific and technical barriers were encountered, each of which motivated novel methodological solutions and contributions to the scientific community.

A first major challenge was the scarcity of open, standardized and high-quality datasets that integrate heterogeneous data sources relevant for coral reef monitoring. To address this, we curated and published a new dataset (Chapter 2) that combines citizen-collected data with ASV and Unmanned Aerial Vehicle (UAV) data, all gathered under standardized protocols. This contribution provides the community with a reproducible and versatile dataset that goes beyond training and benchmarking Artificial Intelligence (AI) models. It enables ecological applications such as revisiting sites to assess habitat evolution, producing spatialized species distribution maps and serving as ground truth for remote sensing models. At the same time, it promotes best practices in open science and Findable Accessible Interoperable Reusable (FAIR)-compliant data management.

A second challenge was the absence of adapted and reproducible workflows for ecological monitoring of marine environments using autonomous platforms. In response, Chapter 3 developed a complete workflow that geotags underwater frames with precise positional information by preprocessing raw ASV data into imagery enriched with embedded metadata. This pipeline provides a reproducible and standardized data collection protocol that transform unstructured videos into geotagged frames, which can then be used for applications such as ecological assessment and photogrammetric reconstruction. To complement this, we introduced a multi-label classification model designed for underwater images, enabling automatic recognition of 31 benthic classes. By integrating classification outputs with positional metadata, the framework allows ecological information to be spatially upscaled, linking local benthic observations to broader monitoring scales.

The next challenge lay in transferring ecological information from underwater observations to the aerial scale. Directly using pixel-wise annotations was infeasible due to annotation costs, so in Chapter 4 we interpolated underwater predictions to generate aerial annotations, which were then used to train a multi-label classifier addressing 12 benthic classes. This work provided the first operational link between modalities, demonstrating how sparse underwater labels can inform broader spatial predictions. Building on this, Chapter 5 proposed a novel Weakly Supervised Semantic Segmentation

(WSSS) framework for aerial imagery, shifting from the classification-based interpolation to segmentation-based models exploiting only underwater multi-label annotations. The final upscaling was conducted on 5 classes as a proof of concept, generating detailed aerial habitat maps. This represents a significant methodological advance, reducing dependence on costly segmentation masks while enabling large-scale monitoring.

Finally, the extension to high-altitude aerial imagery (Chapter 6) addressed the challenge of scaling the framework beyond local UAV surveys to regional and potentially global observations. Here, we explored the potential and limits of transferring benthic information into coarser-resolution airborne-collected data, highlighting both the opportunities for large-scale mapping and the challenges imposed by reduced spatial resolution.

In summary, the contributions of this thesis can be grouped into three main axes:

(i) the creation and release of a standardized and heterogeneous dataset designed for coral reef monitoring at multiple scales

(ii) the design of a reproducible framework for underwater data collection and processing with autonomous platforms, integrating precise geotagging, metadata enrichment and automated multi-label classification of benthic habitats

(iii) the development of two novel methods to transfer ecological information across modalities and scales, from underwater to low-altitude and high-altitude aerial observations

Together, these contributions open new research directions at the intersection of marine ecology, photogrammetry, CV and remote sensing and provide concrete tools and data for the scientific community to build upon.

7.2 Limits and future directions

The multi-scale framework developed in this thesis embodies both the advantages and challenges of an interdisciplinary approach, combining elements from Deep Learning (DL), remote sensing, marine robotics, photogrammetry, Geographic Information System (GIS) and marine ecology. Building such an interdisciplinary framework requires careful consideration of a wide range of parameters during workflow design. Each component has its own requirements, which must be balanced to ensure that the needs of one domain do not negatively affect the others. Nonetheless, this same interdisciplinarity is also a major strength, since advances in any of these fields can directly enhance the overall performance of the framework, fostering continuous improvement as technology evolves.

These advancements and research perspectives can be grouped into three categories:

- Data advancements: improvements in the quality, diversity and richness of input data.

- Technique advancements: methodological progress in AI algorithms and photogrammetry workflows.
- Ecological advancements: enabling long-term monitoring through repeated surveys and temporal comparisons at the same sites.

Each of these categories is examined in the following subsections.

7.2.1 Data advancements

One of the most promising directions for improvement lies in the expansion and enrichment of the data sources used within the multi-scale framework. The models developed in this thesis are indeed constrained by site-specific training data (primarily from Réunion Island and Mauritius) and by class imbalance, as shown in Section 3.2.3. The quantity of data available for both model training and the actual monitoring of reef ecosystems remains limited due to the field constraints of scientific surveys. While RGB imagery has proven to be a critical source of information for this study, it represents only one component of the broader range of data that could be used to monitor reef communities and habitats. Future work could explore:

- Addressing dataset limitations: by collecting additional data from diverse reef systems and specifically targeting under-represented classes. Integrating these enriched datasets into the framework would improve model generalization across sites and the detection of rare classes.
- Citizen science integration for large-scale habitat mapping: the existence of a standardized collection protocol (as illustrated in Chapter 2) opens the possibility of incorporating opportunistic data collected by citizen scientists, thereby expanding spatial and temporal coverage of ecological surveys at minimal cost (Anton et al., 2021; Ierodiaconou et al., 2022). A dedicated, user-friendly online platform could guide contributors through the data collection protocol, allowing them to upload their data and, after automated processing on a central server, displaying results and highlighting individual contributions to the monitoring program. Such an approach would enable large-scale, long-term monitoring over vast areas, while also raising public awareness of ecological issues and fostering community engagement in marine conservation (Pucino et al., 2021).
- Integration of environmental and biological metadata: by enriching AI model inputs with additional geospatial, environmental and biological variables to improve predictive performance. These could include features such as latitude, longitude, depth, and distance to the coastline, as well as temporal information (e.g., season, time of day) relevant for tracking biological cycles (Masoudi et al., 2024). Additionally, Environmental DNA (eDNA) data collected via "in situ" samplers could be integrated into fine-scale underwater models to constrain predictions to species biologically confirmed at the same location and time (Cowart et al., 2024). Such

multimodal conditioning could enable hybrid AI–ecology models that integrate imagery with geospatial and biological evidence, moving beyond purely visual-based classification.

7.2.2 Technique advancements

From a technical standpoint, the framework developed in this thesis remains limited by the difficulties of producing reliable underwater orthophotos, the high annotation costs associated with semantic segmentation and the current dependence on multi-label classification rather than fully pixel-wise approaches. These constraints reduced the spatial precision of the annotations transferred from underwater predictions to the aerial scale and restricted the number of benthic classes that could be included in the analysis.

The initial objective of this PhD was to establish a multi-scale approach that integrated underwater photogrammetry with semantic segmentation of benthic habitats. After one year, this direction proved impractical for two main reasons: photogrammetry and semantic segmentation of benthic habitats. Indeed, the creation of reliable underwater orthophotos was hindered by difficulties in accurately geolocating ASV frames (see Section 3.2.2). Moreover, semantic segmentation of underwater images required labour-intensive annotation, which was not feasible at the required scale (and at the time the work was done). Thus, the research focus shifted to a more achievable path: multi-label classification of single underwater images, which allowed for the development of the first upscaling approach for combined underwater and aerial observations shown in Chapter 4. Nonetheless, the ambition of doing segmentation on aerial images remained an ongoing aspiration. Leveraging the rapid progress in CV, particularly WSSS, the method developed in Chapter 5 did not require underwater mask annotations to generate aerial segmentation predictions. Instead, it relied solely on underwater multi-label annotations, greatly simplifying data preparation and facilitating deployment.

As the thesis progressed, underwater photogrammetry matured into a feasible component of the framework. Once underwater geolocation issues were resolved and the environmental conditions favourable for underwater orthophoto generation were identified, we successfully produced 3D models covering tens of thousands of square meters using the ASV, as shown in Section 3.3.4. Moreover, recent advances in zero-shot segmentation models (such as Segment Anything Model (SAM)) now make the annotation of these orthophotos considerably more feasible and, as public underwater segmentation datasets continue to grow (Truong et al., 2025; Sauder et al., 2025), the availability of high-resolution underwater segmentation models increases (Remmers et al., 2025). This evolution reopens the possibility of transferring high-resolution semantic masks from underwater to aerial models, enabling two key improvements:

1. Spatially precise information transfer: as illustrated in Figure 7.3, correctly aligned underwater orthophotos can provide detailed information that can be transferred to aerial imagery. The upper part of the figure shows an underwater orthophoto reconstructed from ASV-acquired frames, revealing fine-scale habitat details. The lower part displays the corresponding aerial orthophoto obtained with an UAV,

which offers broader coverage but at lower resolution. By transferring underwater prediction masks to the aerial scale (rather than relying on centroid-based annotations for each frame) it becomes possible to generate full segmentation masks that capture the exact boundary of each instance in the underwater image, thereby preserving the spatial distribution of benthic classes.

2. Increased class diversity: pixel-wise underwater annotations allow overlapping or co-occurring classes to be distinguished within the same image. For example, instead of assigning a single dominant label such as *Sand* to an entire frame that also contains patches of *Rubble*, segmentation masks can isolate the specific pixels corresponding to each substrate. This refinement enables the transfer of richer and more precise ecological information to the aerial scale, making it possible to include classes such as *Algae*, *Rubble* or coral morphotypes that were previously excluded due to annotation noise.

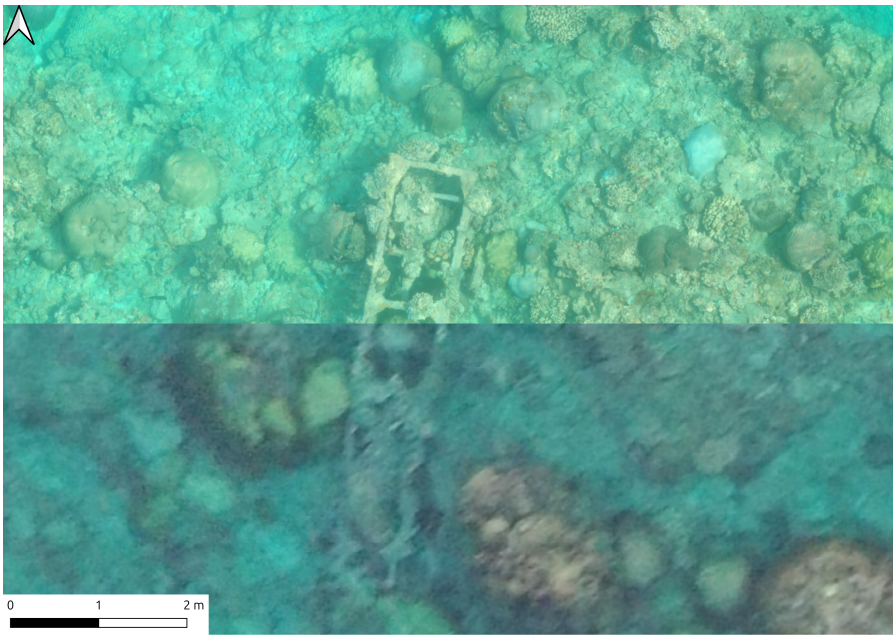


Figure 7.3: *Comparison between underwater (top) and aerial (bottom) orthophotos of the same reef zone at Saint-Leu (Réunion Island) illustrating the difference in level of detail.*

As the technological landscape continues to evolve, several parallel advancements further motivate and reinforce the relevance of this line of research.

Advancements in UAV hardware are among the most promising. Modern drones now feature higher-resolution sensors, enhanced optics and differential Global Navigation Satellite System (GNSS), enabling more precise georeferencing and faster acquisition of imagery over larger areas. The combination of better spatial resolution and accurate

positioning improves the ability to detect and distinguish ecologically meaningful classes in aerial imagery, facilitating finer-scale and more reliable large-area habitat mapping.

In parallel, AI models are progressing at a remarkable pace. For example, new foundation models in CV, such as DINOV3 (Siméoni et al., 2025), suggest that even by simply retraining with the same dataset and reapplying the whole pipeline, significant performance gains can be achieved. Such rapid methodological evolution gives confidence that the framework developed in this thesis will continue to benefit from future breakthroughs, without requiring fundamental changes in design.

Finally, State Of The Art (SOTA) models are now tackling increasingly complex tasks, including 3D point cloud segmentation and cross-domain multimodal fusion, which could allow direct analysis of combined aerial and underwater 3D datasets. These developments indicate that multi-scale, multi-platform frameworks like the one proposed in this thesis are likely to benefit substantially from upcoming AI-driven methods, ultimately enabling frequent, high-resolution, and cost-effective environmental monitoring at unprecedented spatial scales.

7.2.3 Ecological advancements

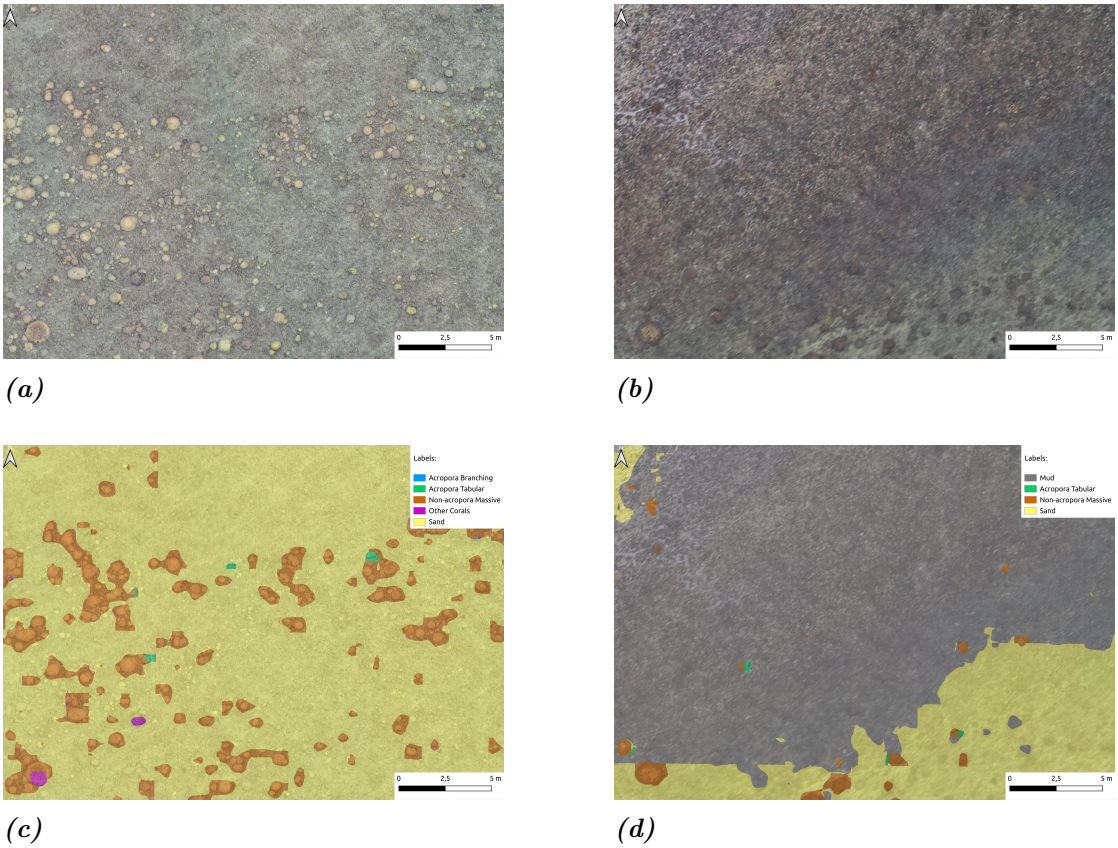
The ecological dimension of the framework is currently constrained by the absence of long-term standardized monitoring at the same sites. As a result, it is not yet possible to systematically quantify reef ecosystem dynamics or to rapidly detect ecological changes caused by disturbances such as cyclones, bleaching events or human impacts.

A key perspective is the long-term implementation of standardized monitoring protocols, enabling repeated surveys of the same areas. This temporal approach would facilitate the rapid detection of changes, the quantification of ecosystem trends and the provision of robust data for conservation management.

Recent advances in UAV technology have produced fully autonomous multi-rotor platforms capable of inspecting large sites without human intervention (Knitter et al., 2024). When paired with hangar systems, these drones can autonomously land, recharge, and relaunch, enabling continuous operation depending on local regulatory constraints.

As an example, in late February 2025, the powerful cyclone *Garance* struck Réunion Island, bringing winds of up to $234\,\mathrm{km\,h^{-1}}$ and extreme rainfall (140 mm in one hour and over 250 mm in three hours along the north coast (Lamy et al., 2025)). This extreme precipitation transported large amounts of sediment from the watershed into the lagoon, depositing a thick layer of mud on coral habitats. The full extent of the cyclone's ecological impact on corals has yet to be quantified.

Using the UAV-based monitoring technique described in Chapter 5, we compared two drone orthophotos acquired two years before and four months after the cyclone (Figure 7.4) to identify the areas most affected. In the Saint-Leu lagoon, in a zone directly in front of two ravines converging under a bridge, prior to cyclone *Garance* the area was occupied by numerous massive corals. The 2023 orthophoto, shown in Figure 7.4a and covering an area of $534.218\,\mathrm{m^2}$, depicts a part of this zone in which numerous massive coral colonies can be distinguished and individually identified (Figure 7.4c).



(a) *(b)*

(c) *(d)*

Figure 7.4: *UAV-based monitoring of the Saint-Leu lagoon before and after cyclone Garance: (a) Orthophoto from December 2023 showing a dense aggregation of massive corals. (b) Orthophoto from June 2025, four months after the cyclone, showing visible habitat changes. (c) Semantic segmentation of the 2023 orthophoto using the proposed WSSS framework, allowing identification and counting of massive corals. (d) Semantic segmentation of the 2025 orthophoto, highlighting the loss and degradation of coral cover in the same area (the grey area corresponds to mud cover).*

In the post-cyclone orthophoto (Figure 7.4b), a thick mud layer covers much of the surveyed area, burying a large number of massive corals. In the lower portion of the image, however, several massive corals remain visible. In the corresponding segmentation (Figure 7.4d), these remaining colonies are correctly recognised, while areas covered by mud are frequently misclassified. For illustration purposes, this misclassification has been coloured in grey as it reflects the absence of sediment-covered examples in the training dataset. This drawback could be addressed by adding a dedicated *Mud* class, as described in Section 5.4.3. Despite this limitation, the framework effectively identifies coral loss: 89.70% of *Non-acropora Massive* colonies present in 2023 were no longer detected in the 2025 survey. In marine conservation, these systems could be deployed for periodic lagoon surveys, providing high-frequency observations that enable the early

detection of ecological changes.

Looking ahead, we believe that the development of autonomous monitoring technologies will soon play a central role in implementing effective conservation strategies in response to the accelerating degradation of coral reef ecosystems.

Bibliography

Anton, V., Germishuys, J., Bergström, P., Lindegarth, M., and Obst, M. (2021). An open-source, citizen science and machine learning approach to analyse subsea movies. *Biodiversity data journal*, 9:e60548.

Cowart, D. A., Chevrier, T., Nieblas, A.-E., Chanut, J., Bernard, S., and Bonhommeau, S. (2024). Detecting local variations across metazoan communities in back-reef depressions of reunion island (mascarene archipelago) through environmental dna survey. *Frontiers in Marine Science*, 11:1423676.

Ierodiaconou, D., Kennedy, D. M., Pucino, N., Allan, B. M., McCarroll, R. J., Ferns, L. W., Carvalho, R. C., Sorrell, K., Leach, C., and Young, M. (2022). Citizen science unoccupied aerial vehicles: A technique for advancing coastal data acquisition for management and research. *Continental Shelf Research*, 244:104800.

Knitter, J., Lai, D. T., Baxter, L., Jayarathne, N., and De Vincentis, S. (2024). Survey of autonomous drone hangars-opportunities and challenges for maritime platforms. In *Proceedings of the International Conference on*, volume 2, pages 81–87.

Lamy, K., Poncet, N., Marpinard, É., Belon, R., Cammas, J.-P., Dubois, C., and Leroux, M.-D. (2025). *CLIMAAX-Deliverable Phase 1–Climate risk assessment*. PhD thesis, CLIMAAX; Région Réunion; CNRS; OSU Réunion.

Lutzenkirchen, L. L., Duce, S. J., and Bellwood, D. R. (2024). Exploring benthic habitat assessments on coral reefs: a comparison of direct field measurements versus remote sensing. *Coral Reefs*, 43(2):265–280.

Masoudi, M., Giering, S. L., Eftekhari, N., Massot-Campos, M., Irisson, J.-O., and Thornton, B. (2024). Optimizing plankton image classification with metadata-enhanced representation learning. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*.

Pucino, N., Kennedy, D. M., Carvalho, R. C., Allan, B., and Ierodiaconou, D. (2021). Citizen science for monitoring seasonal-scale beach erosion and behaviour with aerial drones. *Scientific reports*, 11(1):3935.

Remmers, T., Boutros, N., Wyatt, M., Gordon, S., Toor, M., Roelfsema, C., Fabricius, K., Grech, A., Lechene, M., and Ferrari, R. (2025). Rapidbenthos: Automated segmentation and multi-view classification of coral reef communities from photogrammetric reconstruction. *Methods in Ecology and Evolution*, 16(2):427–441.

Sauder, J., Domazetoski, V., Banc-Prandi, G., Perna, G., Meibom, A., and Tuia, D. (2025). The coralscapes dataset: Semantic scene understanding in coral reefs. *arXiv preprint arXiv:2503.20000*.

Siméoni, O., Vo, H. V., Seitzer, M., Baldassarre, F., Oquab, M., Jose, C., Khalidov, V., Szafraniec, M., Yi, S., Ramamonjisoa, M., et al. (2025). Dinov3. *arXiv preprint arXiv:2508.10104*.

Truong, Q.-T., Wong, Y.-K., Dang, V. H. K. T., Gotama, R., Nguyen, D. T., and Yeung, S.-K. (2025). Msc: A marine wildlife video dataset with grounded segmentation and clip-level captioning. *arXiv preprint arXiv:2508.04549*.

Declaration of Generative AI and AI-assisted technologies in the writing process

During the preparation of this work the author(s) used ChatGPT-4o, ChatGPT-4.1 and ChatGPT-5 in order to improve language and readability. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the published article. These tools were not involved in the design, implementation, data analysis, or manuscript preparation of the study.

List of papers

This thesis is based on four primary articles, listed below. These works were selected because they fall within the scope of the thesis and represent significant, progressive steps in my research.

Additional contributions related to the thesis topic are also presented, but they are not included in the core content as my involvement in these works was not major.

First author publications

1. **Contini, M.**, Illien, V., Julien, M., Ravitchandirane, M., Russias, V., Lazennec, A., Chevrier, T., Rintz, C.L., Carpentier, L., Gogendeau, P., *et al.* (2025). *Seatizen Atlas: a collaborative dataset of underwater and aerial marine imagery.* **Scientific Data**, 12(1), 67. DOI: 10.1038/s41597-024-04267-z.

2. **Contini, M.**, Julien, M., Poulain, S., Ravitchandirane, M., Illien, V., Russias, V., Lazennec, A., Hamylton, S., Spencer, T., Koester, A., Gogendeau, P., Chevrier, T., Talpaert Daudon, J., Constance, A., Jones, C., Fleischer-Dogley, F., Bernard, S., Joly, A., Barde, J., Bonhommeau, S. (2025). *Combining AI and drones for marine biodiversity mapping in remote islands.* **Deep Sea Research Part II**. *In revision.*

3. **Contini, M.**, Illien, V., Barde, J., Poulain, S., Bernard, S., Joly, A., Bonhommeau, S. (2025). *From underwater to drone: A novel multi-scale knowledge distillation approach for coral reef monitoring.* **Ecological Informatics**, 103149. DOI: 10.1016/j.ecoinf.2025.103149.

4. **Contini, M.**, Illien, V., Poulain, S., Bernard, S., Barde, J., Bonhommeau, S., Joly, A. (2025). *The Point is the Mask: Scaling Coral Reef Segmentation with Weak Supervision.* **Proceedings of the IEEE/CVF international conference on computer vision**, Workshop on *Computer Vision for Analysis of Underwater Imagery and Automated Analysis of Marine Visual Data for Environmental Monitoring (CVAUI-AAMVEM). Accepted.* DOI: abs/2508.18958.

Contributions to other publications

1. Gogendeau, P., Bonhommeau, S., Fourati, H., Julien, M., **Contini, M.**, Chevrier, T., Nieblas, A.E., Bernard, S. (2025). *An autonomous surface vehicle for acoustic tracking, bathymetric and photogrammetric surveys.* **Ocean Engineering**, 331, 121201. DOI: 10.1016/j.oceaneng.2025.121201.

2. Talpaert Daudon, J., **Contini, M.**, Urbina-Barreto, I., Elliott, B., Guilhaumon, F., Joly, A., Bonhommeau, S., Barde, J. (2023). *GeoAI for marine ecosystem monitoring: A complete workflow to generate maps from AI model predictions.* **The International Archives of the Photogrammetry, Remote Sensing and Spa-**

tial Information Sciences, XLVIII-4/W7-2023, 223–230. DOI: 10.5194/isprs-archives-XLVIII-4-W7-2023-223-2023.

Oral communications

I have had the opportunity to present my work in the following conferences:

- **BES**, Liverpool, United Kingdom (10–13 December 2024). Oral presentation at the British Ecological Society Annual Meeting. Link to presentation
- **ECCB**, Bologna, Italy (17–21 June 2024). Oral presentation at the European Congress of Conservation Biology. Link to presentation
- **GeoHab**, La Réunion, France (8–12 May 2023). Oral presentation at the Marine Geological and Biological Habitat Mapping conference. Link to presentation

Field work

During my PhD, I participated in several field campaigns:

- Indian Ocean Expedition (October–November 2022). Collected bathymetry and images using ASVs at Aldabra (32 missions), Saya de Malha (3 missions) and Saint-Brandon (9 missions).
- Data collection campaign on the west coast of Madagascar (29 April–5 May 2023). Collected imagery using ASVs during a one-week campaign comprising 13 missions.
- Four Ground Control Point (GCP) data collection campaigns in Réunion Island (2024–2025).
- Prior to this thesis, I completed a two-year Civic Service at Ifremer in Réunion Island in the same team, contributing to 73 data collection campaigns using paddleboards, snorkelling and ASVs.

Appendices



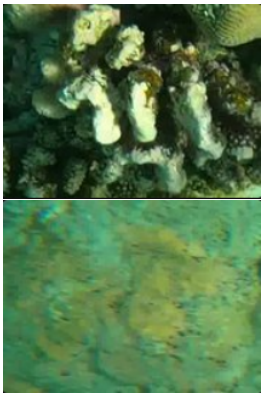
Figure 7.5: *A sea cucumber camouflaged on a sandy and algal substrate in shallow coastal waters of Nosy Ve (Madagascar).*
Image credit: Matteo Contini, April 2023.

1 Supplementary material Chapter 2: Seatizen atlas: a collaborative dataset of underwater and aerial marine imagery

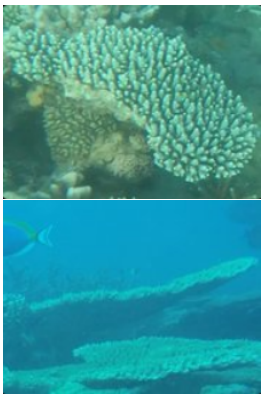
1.1 Annotated classes

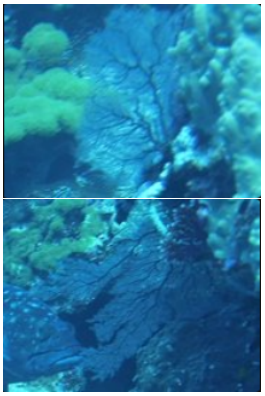
In the following table, we present a complete list of the labels that we used for annotating images. These labels do not necessarily represent the phylogenetics of the species. While some labels correspond to monophyletic groups, others are based on morphological similarities that often reflect similar ecological roles. For example, the label *Non-acropora Massive* is not a clade (it does not represent a common ancestor and all its descendants) and includes polyphyletic species. The **Group** column specifies the group to which each label belongs, and the **Label** column indicates the name of the label in annotation files. The **Notes** column provides a brief description justifying annotator choices. Finally the column **Level of Expertise** indicates the taxonomic rank. Thus, a Level of Expertise of 1 represents labels differentiated by close morphology, while Levels of Expertise 2 and 3 correspond to clades at the genus and species levels, respectively. Depending on the task, annotations can of course be used at their finest ecological level or can be degraded, as in Figure 7.5[1], where only the *Sea cucumber* class is annotated without going to the species level.

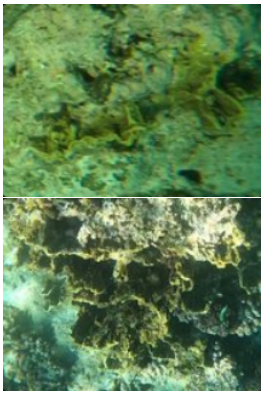
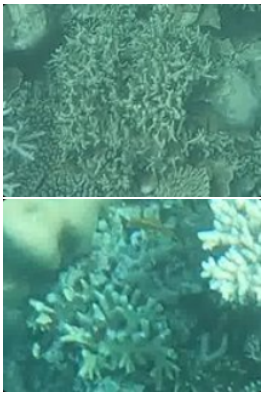
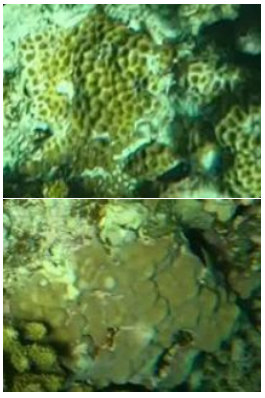
[1]The image can be found in Contini et al. (2025) under the relative path:
`20230430_MDG-NOSY-VE_SCUBA-1_02/PROCESSED_DATA/FRAMES/20230430_MDG-NOSY-VE_SCUBA-1_02_1_009.jpeg`

Group	**Label**	**Notes**	**Level of Expertise**	**Example Image**
Algae	Algal Assemblage	Algal turfs containing several species of algae.	1	
	Algae Halimeda	Upright algae $\geq$ 10mm, rigid and with complex structures. Green, brown or red algae (e.g., *Halimeda*, *Turbinaria*).	1	
	Algae Coralline	Often encrusting, of the corallinaceous type. The thallus can be encrusting and strongly adherent to the substrate (*Hydrolithon*), lamellar (*Mesophyllum*) or branched (*Cheilosporum*, *Lithophyllum*).	1	

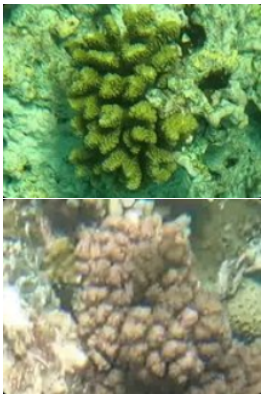
Group	**Label**	**Notes**	**Level of Expertise**	**Example Image**
	Algae Turf	Short filamentous algae. An assemblage of filamentous algae, often $\leq$ 10mm in height, which are fast-growing and highly productive. (e.g., Turf algal).	1	
Coral	Acropora Branching	With secondary branches (e.g., *A. formosa*, *A. palmata*).	2	
	Acropora Digitate	No secondary branches (e.g., *A. digitifera*, *A. humilis*).	2	

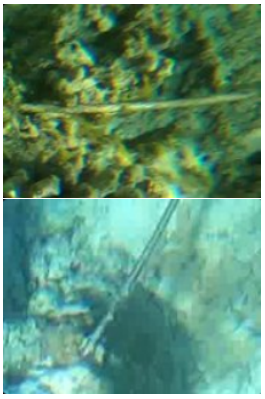
Group	Label	Notes	Level of Expertise	Example Image
	Acropora Submassive	Robust with few digitized branches (e.g., *A. monticulosa*).	2	
	Acropora Tabular	Large horizontal plates (e.g., *A. hyacinthus*, *A. cytherea*).	2	
	Bleached Coral	Coral that has turned white due to the loss of its symbiont zooxanthellae, typically as a result of stress factors.	1	

Group	Label	Notes	Level of Expertise	Example Image
	Dead Coral	Recent mortality (different bleached coral), white to light brown coral.	1	
	Gorgonian	A type of soft coral characterized by a flexible, tree-like structure.	1	
	Living Coral	Used only when the coral cannot be distinguished and not for each living coral.	1	

Group	Label	Notes	Level of Expertise	Example Image
	Non-acropora Millepora	Fire coral (e.g., *Millepora platyphylla*).	2	
	Non-acropora Branching	With secondary branches (e.g., *Seriatopora hystrix*).	1	
	Non-acropora Encrusting	A large part is attached to the substrate (e.g., *Porites vaughani*, *Montipora undata*).	1	

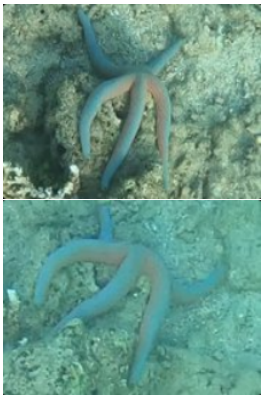
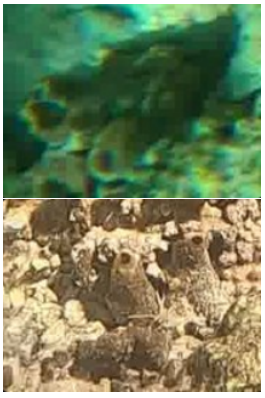
Group	Label	Notes	Level of Expertise	Example Image
	Non-acropora Foliose	Attached to the substrate by one or more points, leaf-like appearance (e.g., *Echinopora mammiformis*, *Pavona cactus*).	1	
	Non-acropora Massive	Massive form resembling a large rock (e.g., *Porites lutea*, *Platygyra daedalea*).	1	
	Non-acropora Coral free	Free-living solitary coral (e.g., *Fungia*).	1	

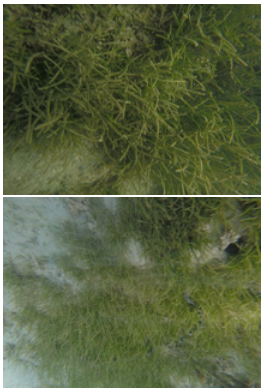
Group	**Label**	**Notes**	**Level of Expertise**	**Example Image**
	Non-acropora Submassive	Very large group. Corals that tend to form small colonies without digitization (e.g., *Porites nigrescens*, *Pocillopora verrucosa*).	1	
Custom Classes	Blurred	Part of the image that is blurred or has air bubbles in it but does not render the image useless	-	
	Homo Sapiens	Part of human body	3	

Group	Label	Notes	Level of Expertise	Example Image
	Human Object	Human object that is not waste (e.g., snorkelling fins, surfboard fins)	-	
	Trample	All pieces of coral removed mechanically	-	
	Useless	Images that cannot be used because they were taken out of the water or show great depths with a lot of blue	-	
	Waste	Human waste	-	

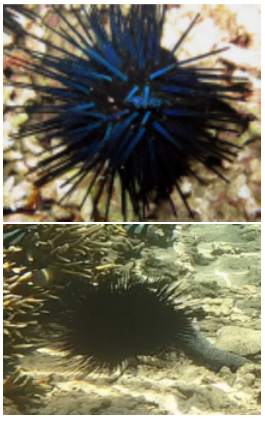
Group	**Label**	**Notes**	**Level of Expertise**	**Example Image**
Habitat	Rock	Basaltic, granitic or other nature	-	
	Rubble	Scrap, debris, particularly coral	-	
	Sand	Sand of coral or basaltic nature, etc.	-	
Other	Thorny starfish	Coral-eating starfish (e.g., *Acanthaster planci*)	2	

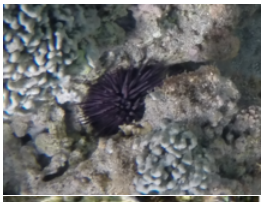
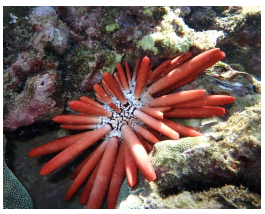
Group	Label	Notes	Level of Expertise	Example Image
	Sea anemone	Marine invertebrates that belong to the phylum *Cnidaria*, typically characterized by a columnar body topped with a ring of tentacles.	1	
	Ascidians	Also known as sea squirts, these are sessile filter-feeding tunicates, often found attached to substrates in the ocean.	1	
	Giant clam	Bivalve mollusks with two hinged shells, commonly found burrowed in sand or mud in marine environments.	1	

Group	Label	Notes	Level of Expertise	Example Image
	Fish	Generic class for all fishes	1	
	Other starfish	Starfish other than *Acanthaster*	1	
	Sponges	Filter-feeding organism through their aquifer system, e.g., *Cliona* sp.	1	

Group	**Label**	**Notes**	**Level of Expertise**	**Example Image**
	Turtle	Generic class for all marine turtles	1	
Seagrass	*Syringodium isoetifolium*	A species of seagrass known for its long, cylindrical leaves.	3	
	Thalassodendron ciliatum	A species of seagrass with flattened, strap-like leaves	3	

Group	Label	Notes	Level of Expertise	Example Image
Sea cucumber	*Actinopyga echinites*	Color ranging from yellowish beige to orange-brown	3	
	Actinopyga mauritiana	Prefers hard substrates, dark brown back and grayish sides	3	
	Bohadschia vitiensis	Yellowish color with a black spot at the base of each portion	3	
	Holothuria atra and *Holothuria leucospilota*	Group of species with bodies sometimes often covered with a thin layer of sand, variable in color and size up to 30 to 60 cm long and 10 cm wide	3	

Group	Label	Notes	Level of Expertise	Example Image
	Stichopus chloronotus	Quadrangular section body, dark green and covered with radii	3	
	Synapta maculata	Up to 3m long and 5cm wide, beige body with brown bands	3	
Sea urchin	*Diadema savignyi*, *Stomopneustes variolaris*, *Echinothrix calamaris* and *Echinothrix diadema*	Group of species with black color with gray or blue reflections and generally numerous and long radioles	3	

Group	Label	Notes	Level of Expertise	Example Image
	Echinometra mathaei	Uniform color (e.g., beige, violet, gray)	3	 
	Echinostrephus molaris	Color ranging from violet to black through brown, long and thin radioles	3	
	Heterocentrotus mamillatus	Large diameter radioles, body and radioles tending towards brown/red	3	
	Heterocentrotus trigonarius	Triangular section radioles, body and radioles tending towards dark brown/red	3	 

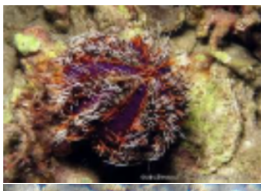
Group	Label	Notes	Level of Expertise	Example Image
	Toxopneustes pileolus	Slightly flattened body, beige, short radioles and flower-shaped	3	
	Tripneustes gratilla	10 bands dotted with short radioles separated by bands without radioles	3	 

Table 1: Classes Table

Echinostrephus molaris (Blainville, 1825) observed in Maldives by John Hepburn (licensed under CC BY 4.0), source: iNaturalist. *Heterocentrotus mamillatus* (Linnaeus, 1758) observed in United States of America by blitzen1986 (licensed under CC BY-NC 4.0), source: iNaturalist. *Toxopneustes pileolus* (Lamarck, 1816) observed in Kenya by thehousebunting (licensed under CC BY-NC 4.0), source: iNaturalist.

2 Supplementary material Chapter 3: Underwater marine biodiversity mapping in remote islands

2.1 ASV specifications

The specifications of the Autonomous Surface Vehicle (ASV) are outlined in Table 7.1.

Table 7.1: *ASV specifications*

Category	Specifications
Transport	1 to 2 people
Deployment	From the shore or a boat
Environment	Tropical and temperate (10°C - 35°C)
Breaking Waves	Max 30 cm
Non-Breaking Waves	Max 2 m
Wind	Max 25kts
Current	Max 2 $m\,s^{-1}$
Payload	10 kg
Range (Connected)	1 km
Range (Disconnected)	Visible limit
Dimensions (LxWxH)	100cm x 55cm x 50cm
Weight	>5kg
Mission Duration (Imaging + Bathymetry)	1 hour
Mission Duration (Bathymetry Only)	2.5 hours
Speed	Between 0.5 $m\,s^{-1}$ and 1.2 $m\,s^{-1}$
Navigation (Automatic)	On flight plan
Navigation (Manual)	Manual control
Echosounder (Model)	ECT400 or CERULEAN S500
Camera (Waterproof)	GoPro Hero 8 black
Echosounder (Frequency)	400 kHz
Echosounder (Range)	100 m

For a visual representation of the electronic part of the ASV, refer to the interconnection diagram in Figure 7.6.

2.2 ASV camera settings

GoPro camera setting can be found in Table 7.2.

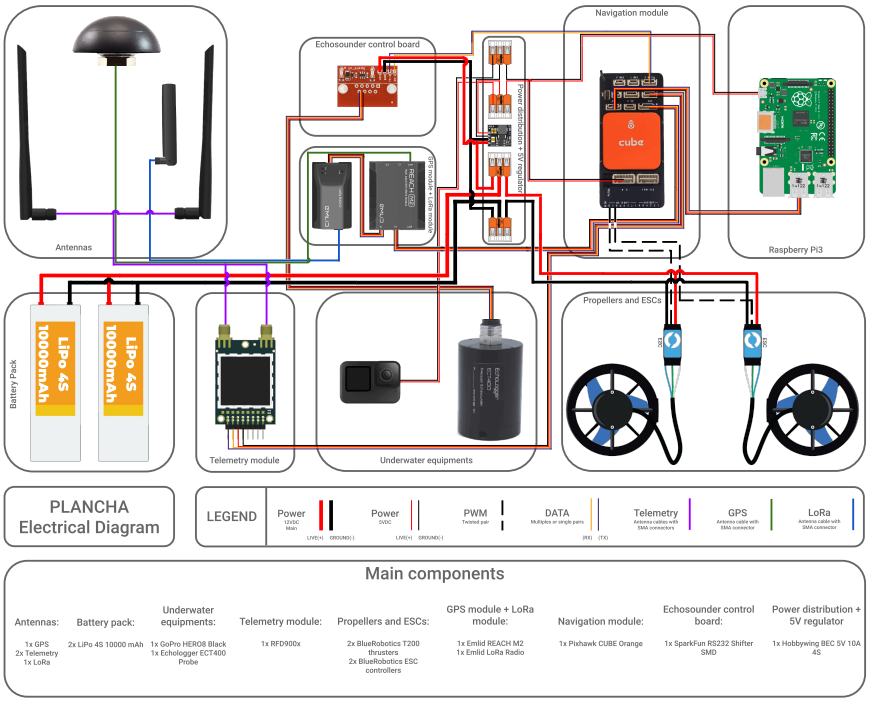


Figure 7.6: *Interconnection diagram of the autonomous surface vehicle realised by Mervyn Ravitchandirane*

2.3 ASV bathymetry

The bathymetric surveys conducted with the ASV using a single-beam echosounder follow the criteria specified in the *IHO Standards for Hydrographic Surveys, Ed. 6.0.0 Sept 2020*. These surveys were designed to meet the requirements of the *order 1a classification*, specifically suited for data acquisition in ports, port approach channels, coastal regions, or inland navigation channels, with a depth limit of less than 100 meters.

The collected data, including measurements from the echosounder, GPS and Inertial Measurement Unit (IMU), were extracted from the log file of the ASV's autopilot. It is of the highest importance to ensure the accuracy and fidelity of the seabed representation through the implementation of a rigorous post-processing methodology. This necessitates the execution of a series of steps, including filtering, compensation and geolocation.

The processing of raw ASV data to obtain a properly georeferenced seabed model was facilitated by a Python workflow. Following this, a series of processing algorithms were implemented in Python, utilising `Numpy`, `Scipy` and `Pyproj` libraries.

The main processing steps were:

1. The analysis of the autopilot log files and the pre-filtering of the data.

2. The construction of a data table containing depth, GPS corrected with Post-Processing Kinematic (PPK) and attitude (Roll, Pitch and Yaw) data through

Table 7.2: *Video file parameters*

Parameter	Value
File Type Extension	MP4
MIME Type	video/mp4
Time Scale	60000
Preferred Rate	1
Preferred Volume	1
Firmware Version	HD8.01.02.51.00
Camera Model Name	HERO8 Black
Auto Rotation	U
Digital Zoom	Y
Pro Tune	Y
White Balance	6500K
Sharpness	HIGH
Color Mode	FLAT
Auto ISO Max	400
Auto ISO Min	100
Rate	2_1SEC
Field Of View	L
Sensor Readout Time	7.9200005531311
Electronic Image Stabilization	N/A
Image Width	1920
Image Height	1080
Graphics Mode	0
X Resolution	72
Y Resolution	72
Compressor Name	GoPro AVC encoder
Bit Depth	24
Video Frame Rate	59.9400599400599
Avg Bitrate	45266194

the use of smoothing filters and KD-tree search techniques.

3. The correction of GPS and depth data by transforming the geometric reference in accordance with the ASV's attitude and local geoid parameters (if available).

4. The creation of a regular X,Y grid and interpolation of depth data over the grid.

5. The removal of outlying depth values based on prior knowledge about the seabed.

6. The generation of raster and shapefile outputs based on re-interpolated data for integration into Geographic Information System (GIS) softwares.

All the codes are available on Plancha Workflow Github.

2.4 Remote data georeferencing

The ASV is equipped with a dual-frequency Global Navigation Satellite System (GNSS) module that provides an initial position accuracy of about 1 meter. Upon the team's arrival at Aldabra island, a temporary GNSS base station was installed near the research station. The position of this station was then used to obtain centimetre-level accuracy in position measurements. To achieve centimetre-level precision for echo sounder and

camera data, the GNSS data from both the ASV and the reference base station (itself corrected through Precise Point Positioning (PPP) methodology[2]) were combined using PPK techniques.

Table 7.3 shows the final position of the base calculated by averaging the estimated positions over six days.

Table 7.3: *Geographical coordinates of the base station*

Latitude (deg)	Longitude (deg)	Altitude (m)
-9.40188139	46.20632859	3.79

2.5 Per-class F1 scores

Table 7.4 reports the F1 score for each class described in Section 3.2.3. These results complement the global metrics presented in Section 3.3.1 by providing a more detailed view of class-wise performance, highlighting which categories are better or more poorly predicted by *DinoVdeau* model.

Table 7.4: *Per-class F1 scores on the test dataset*

Class	F1 score
Acropora Branching	0.909
Acropora Digitate	0.608
Acropora Submassive	0.356
Acropora Tabular	0.934
Algal Assemblage	0.769
Algae Halimeda	0.460
Algae Coralline	0.736
Algae Turf	0.865
Atra/Leucospilota	0.852
Bleached coral	0.671
Blurred	0.661
Dead coral	0.751
Fish	0.765
Homo Sapiens	0.829
Human object	0.821
Living coral	0.411
Non-acropora Millepora	0.797
Non-acropora Encrusting	0.657
Non-acropora Foliose	0.819
Non-acropora Massive	0.749
Non-acropora Coral free	0.476
Non-acropora Submassive	0.722
Rock	0.885
Sand	0.928
Rubble	0.774
Sea cucumber	0.843
Sea urchin	0.784
Sponges	0.461
Syringodium isoetifolium	0.979
Thalassodendron ciliatum	0.996
Useless	0.979

[2]PPP is a method employed for securing precise positioning over extended observations. We executed PPP on the reference base station using the CSRS-PPP online service provided by the Geodetic Survey Division of Natural Resources Canada.

3 Supplementary material Chapter 4: From underwater to drone: a novel multi-scale knowledge distillation approach for coral reef monitoring

3.1 UAV mission planning

Imaging the seabed, even in tropical environments with very clear waters, is often complicated by reflections of sunlight from the water surface. Direct sun rays reflections can be extremely bright and cause oversaturated areas on aerial images. These reflections overexpose the image, making it difficult to see the seafloor structure, as shown in Figure 1.6.

If the seafloor is not visible in the image, Structure from Motion (SfM) algorithms cannot map these areas effectively. It is thus necessary to avoid these reflections in the images. A solution is to take images when the sun is at a low angle, which reduces the reflection of sunlight on the water surface (Slocum et al., 2019). This can be achieved by surveying the desired area early in the morning or late in the afternoon, or on a cloudy day. In our case, images were taken between daybreak and sunrise, which is the best time to avoid reflections on the water surface on the west part of Réunion Island.

3.2 UAV mission execution

In SfM surveys, overlap and sidelap refer to the percentage to which each consecutive image overlaps with the preceding image and the images on adjacent flight lines. In order to obtain a good 3D model, it is important to have a high overlap and sidelap between images. The overlap and sidelap should be at least 70% for a good 3D model (Westoby et al., 2012). In our case, the overlap and sidelap were both set to 80%.

For what concerns the flight altitude, a lower one implies a higher resolution point cloud in the SfM process since more details of the seabed will be visible but on the other hand the wave induced refraction is more visible. On the contrary, setting a too high flying altitude will reduce the resolution of the point cloud, making it difficult to distinguish characteristics of study objects (e.g., coral colonies, habitats, etc.). In our case, the flight altitude was set between 40 m and 60 m.

Bibliography

Contini, M., Illien, V., Bonhommeau, S., Barde, J., and Joly, A. (2025). Underwater images collected by scuba diving in nosy-ve, madagascar - 2023-04-30.

Slocum, R. K., Wright, W., Parrish, C., Costa, B., Sharr, M., and Battista, T. (2019). Guidelines for bathymetric mapping and orthoimage generation using suas and sfm, an approach for conducting nearshore coastal mapping.

Westoby, M. J., Brasington, J., Glasser, N. F., Hambrey, M. J., and Reynolds, J. M. (2012). 'structure-from-motion'photogrammetry: A low-cost, effective tool for geo-science applications. *Geomorphology*, 179:300–314.