THÈSE

POUR L'OBTENTION DU GRADE DE

DOCTEUR DE L'UNIVERSITÉ DE POITIERS

FACULTÉ DES SCIENCES FONDAMENTALES ET APPLIQUÉES

DIPLÔME NATIONAL - ARRÊTÉ DU 25 MAI 2016

École Doctorale : Sciences et Ingénierie des Systèmes, Mathématiques, Informatique - SISMI
Secteur de Recherche : Traitement du Signal et des Images

Présentée par:

Quang Huy Do

Integration and non-intrusive human clinic machine learning for kidney assessment

Directeurs de thèse:

Rémy GUILLEVIN

JURY

Rémy Guillevin, PU-PH, CHU de Poitiers ..... Directeur

Pascal Bourdon, Maître de conférence, Université de Poitiers ..... Co-directeur, Encadrant

David Helbert, Maître de conférence, Université de Poitiers ..... Encadrant

Marie Beurton-Aimar, Maître de conférences, Université de Bordeaux ..... Rapporteur

Jean-Noël Vallee, PU-PH, CHNO des 15-20 ..... Rapporteur

Olivier Alata, Professeur, Université Jean-Monnet Saint-Étienne ..... Examinateur

Contents

Acknowledgements **X**

Abstract **X**

Introductions **X**

Context and motivations . . . X

Objective . . . X

Thesis overview . . . X

Contributions . . . X

1 Medical imaging for kidney disease diagnosis **9**

1.1 Renal anatomy and function . . . 9

1.1.1 Kidney anatomy and function . . . 9

1.1.2 Kidney dissease and transplantation . . . 11

1.2 Magnetic Resonance Imaging for non-invasive diagnosis . . . 12

1.2.1 Nuclear magnetism . . . 13

1.2.2 Image reconstruction . . . 17

1.2.3 Image quality and artifacts . . . 19

1.3 Ultra-high field MRI: the next generation . . . 21

1.4 Computer-aided MRI analysis: the augmented pathologist . . . 24

1.5 Conclusions . . . 26

2 Deep learning and Neural network **27**

2.1 Machine learning overview . . . 27

2.1.1 Supervised learning . . . 28

2.1.2 Unsupervised learning . . . 29

2.1.3 Other concepts . . . 30

2.2 Deep Neural Network . . . 32

2.2.1 Very first neural network . . . 33
2.2.2 Neural network training . . . 33
2.3 Concepts and terminology . . . 34
2.3.1 Activation function . . . 34
2.3.2 Loss function . . . 35
2.3.3 Back-propagation . . . 35
2.3.4 Optimization . . . 37
2.3.5 Learning rate . . . 38
2.3.6 Batch Normalization . . . 38
2.3.7 Dropout . . . 39
2.4 Convolutional Neural Network . . . 40
2.5 Representation learning . . . 41
2.5.1 Pre-train . . . 42
2.5.2 Transfer Learning . . . 43
2.6 Conclusion . . . 43

3 Generative models **45**
3.1 Autoencoders . . . 46
3.1.1 Architecture . . . 46
3.1.2 Regularized AE . . . 47
3.1.3 Variational autoencoders . . . 49
3.2 Generative adversarial network . . . 49
3.2.1 Adversarial architecture . . . 50
3.2.2 Loss function . . . 52
3.2.3 Training GAN difficulty . . . 52
3.2.4 Deep Convolutional GAN . . . 54
3.3 GAN variations . . . 55
3.3.1 Wasserstein GAN . . . 56
3.3.2 Conditional GAN . . . 57
3.4 CycleGAN . . . 59
3.4.1 General architecture . . . 60
3.4.2 Loss function . . . 61
3.4.3 Training procedure . . . 62
3.4.4 Generator architecture . . . 62
3.4.5 Discriminator architecture . . . 65

3.5 Conclusion . . . 66

4 MRI super-resolution **67**

4.1 Introduction . . . 67

4.2 Related work . . . 68

4.2.1 Image super-resolution . . . 68

4.2.2 Learning-based methods for super-resolution . . . 70

4.2.3 Generative adversarial networks for super-resolution . . . 71

4.2.4 MRI super-resolution . . . 72

4.3 Methodology . . . 73

4.3.1 Network architecture . . . 74

4.4 Experiments . . . 77

4.4.1 Dataset . . . 77

4.4.2 Training setup . . . 77

4.4.3 Evaluation metrics . . . 78

4.5 Results . . . 80

4.6 Discussion . . . 81

4.7 Conclusion . . . 84

5 Ultra-high field MRI synthesis **85**

5.1 Introduction . . . 85

5.2 Related work . . . 86

5.2.1 Common network architectures . . . 86

5.2.2 MRI synthesis . . . 89

5.3 Network architecture . . . 90

5.3.1 Adversarial network architecture . . . 90

5.3.2 Generative network architecture . . . 91

5.4 Experiments . . . 92

5.4.1 Dataset . . . 92

5.4.2 Pre-processing . . . 94

5.4.3 Training setup . . . 95

5.4.4 Evaluation metrics . . . 95

5.5 Results . . . 97

5.6 Discussion . . . 97

5.6.1 Overview . . . 97

5.6.2 Network characterization . . . 100
5.6.3 Perspective . . . 101
5.7 Conclusion . . . 101

6 MRI Cross modality translation **103**
6.1 Introduction . . . 103
6.2 Related work . . . 105
6.3 Network architecture . . . 106
6.3.1 starGAN . . . 107
6.4 Experiments . . . 110
6.4.1 Dataset . . . 110
6.4.2 Pre-processing . . . 111
6.4.3 Training setup . . . 111
6.4.4 Evaluation metrics . . . 112
6.5 Results . . . 112
6.6 Discussion . . . 117
6.7 Conclusion . . . 120

7 Perspectives **125**

List of Figures

1.1 Kidney anatomy. [22] 10

1.2 Hydrogen spins around its axis, producing an angular momentum. Spins are randomly oriented until the application of an external magnetic field B0. Spins align with B0 in a parallel and anti-parallel direction with an excess of spins oriented anti-parallely, giving a net magnetization vector (NMV) aligned in the opposite direction with B0. 13

1.3 The net magnetization vector aligned with the external magnetic field B0. After the excitation by a radio-frequency pulse, the NMV flips by a certain angle following a spiral trajectory and imposes transversal and longitudinal components. 14

1.4 Examples of transversal (a) and longitudinal (b) magnetization relaxation for different tissues 15

1.5 FID and T2* decay obtained by fitting the curve of FID 16

1.6 Diagram of a spin echo pulse sequence 17

1.7 Illustration of k-space 19

2.1 An example of overfitting, underfitting and usual performance. 31

2.2 Cross validation. 32

2.3 Dropout layer. 39

2.4 Convolutional operation. 40

2.5 Convolutional neural netowrk architecture. 41

3.1 General autoencoder architecture. 47

3.2 Variational autoencoders architecture. 48

3.3 General GAN architecture. 50

3.4 Illustration of DCGAN generator architecture. 55

3.5 Illustration of DCGAN discriminator architecture. 55

3.6 WGAN generator training. The generator produce images during this training. Synthesized data is pretended to be real with label=1. The discriminator weights are frozen but gradients propagate back to the generator 57

3.7 WGAN discriminator training. 57

3.8 Illustration of CGAN workflow. 58

3.9 CycleGAN generator workflow. 60

3.10 The original U-net architecture [170]. It contains a contraction path and expanding path The contraction and expanding paths are sometimes referred to as encoder and decoder, respectively. 63

3.11 A single residual block. (1) generic residual block (2) residual block with removal of normalization 64

3.12 U-net generator architecture.It contains a contraction path and expanding path The contraction and expanding paths are sometimes referred to as encoder and decoder, respectively. 65

4.1 Examples of low and high-resolution MRI. From left to right: low-resolution MRI and high-resolution images, from top to bottom: full view and zoom-in view. . . . 69

4.2 SRDenseNet architecture. 70

4.3 Residual blocks and dense blocks with skip connection. 71

4.4 Residual dense block architecture. 75

4.5 Generator architecture. 75

4.6 Architecture of discriminator. It contains several convolutional layers followed by Instance Normalization and Leaky ReLU. 76

4.7 ESRGAN architecture. 79

4.8 Comparison of model performance on the different ground-truth MRI for quantitative image similarity metrics using SSIM: (a) 3T MRI with 0.9 mm slice thickness, (b) 3T MRI with 0.6 mm slice thickness, (c) 7T MRI with 0.75 mm slice thickness, (d) 7T MRI with 0.5 mm slice thickness 82

4.9 Comparison of model performance on the different ground-truth MRI for quantitative image similarity metrics using PSNR: (a) 3T MRI with 0.9 mm slice thickness, (b) 3T MRI with 0.6 mm slice thickness, (c) 7T MRI with 0.75 mm slice thickness, (d) 7T MRI with 0.5 mm slice thickness 82

4.10 Visualization of 3D model performance on CHU 3T and 7T MRI: (a),(e) low-resolution MRI, (b),(f) ground-truth MRI, (c),(g) SRCycleGAN, (d),(h) SRCycleGAN with deconvolutional layers output. On randomly selected sample, zoom-ins are shown in the red box. The vessel in the yellow circle is blurred out LR MRI (a),(e) and partially recovered in (d),(h) and preserves more details in (c),(g) . . . 83

5.1 Architecture of generator. Here, we use 3 RDBs for feature extraction. The number of blocks can change the size and complexity of whole model. 91

5.2 Architecture of discriminator. It contains several convolutional layers followed by Instance Normalization and Leaky ReLU for 3D data. 92

5.3 Illustration pair 3T and 7T MRI before preprocessiing. 3T samples are inclined with 7T samples, lead to a mismatch of voxel position. 93

5.4 Illustration pair 3T and 7T MRI after alignement. Voxel position of 3T and 7T samples are similar 94

5.5 Axial, sagittal and coronal views of synthetic 7T MRI. From left to right: 3T MRI (input), ground-truth 7T (expected output), synthetic 7T generated by WATNet, synthetic 7T generated by our model. Zoomed-in areas appear as red rectangles. . 96

5.6 Visualization in axial, sagittal and coronal views of 3T synthesized MRI from 7T. From left to right: 7T MRI input , real 3T MRI, and CycleGAN output. 98

5.7 Illustration mismatch between pair 3T and 7T. 99

5.8 Illustration comparison of 7T synthesized MRI for two CycleGAN model. From left to right: 3T MRI input, ground-truth 7T MRI, 7T synthesized using SRCycleGAN generator and 7T synthesized of current CycleGAN. 101

6.1 Comparison between cross-domain models using CycleGAN and starGAN. To handle multiple domains, there are six CycleGAN model needed for each pair of images, while StarGAN only uses a single generator. 104

6.2 Illustration of starGAN schematic, consisting of one generator and one discriminator 108

6.3 The architecture of the generator. 110

6.4 Architecture of discriminator. 110

6.5 Visualization of T1-T1c translation on HGG object in BraTS dataset. In each image, from left to right: T1 input, ground-truth T1c, CycleGAN and starGAN output. 113

6.6 Visualization of T1-T1c translation on LGG object in BraTS dataset. In each image, from left to right: T1 input, ground-truth T1c, CycleGAN and starGAN output. 114

6.7 Visualization of model performance on BraTS dataset on T1c-T1 conversion. In each image, from left to right: T1 input, ground-truth T1c, CycleGAN and starGAN output. 115

6.8 Visualization of model performance on synCHU dataset. In each image, from left to right, top to bottom: input/ground-truth T1/T1c MRI, synthesized T1c/T1c generated by CycleGAN and starGAN. 116

6.9 Failed recontruction of model on glioma subjects from T1 to T1c. From left to right: T1 MRI input, T1c ground-truth MRI, synthesized output by CycleGAN and starGAN. 116

6.10 Visualization of model performance on subset dataset. From left to right: T1 MRI input, T1c MRI, CycleGAN and starGAN output. 118

List of Tables

4.1 Average value of PSNR (dB) and SSIM for scale factors ×2 and ×4 on BraTS and CHU MRI dataset. Resolution-enhancement methods (tricubic interpolation, ESRGAN, 2D SRCycleGAN, 3D SRCycleGAN and 3D SRCycleGAN+DC) were compared with ground-truth images for quantitative evaluation. 80

5.1 MRI synthesis quality assessment in terms of PSNR (dB) and SSIM values. Synthetic 7T data generated out of 3T data with our method and WATNet is compared to its corresponding ground-truth samples for quantitative evaluation. 97

5.2 Average value of PSNR (dB) and SSIM between SRCycleGAN and the current CycleGAN model for MRI synthesis. Synthesized 7T volumes by two CycleGAN models are compared with corresponding ground-truth MRI for quantitative evaluation. 101

6.1 Average value of PSNR (dB) and SSIM for T1-T1c translation on BraTS dataset. Synthesized images generated by CycleGAN and starGAN are compared with corresponding ground-truth MRI for quantitative evaluation. 113

6.2 Average value of PSNR (dB) and SSIM for T1-T1c translation on synCHU dataset. Synthesized images generated by CycleGAN and starGAN are compared with corresponding ground-truth MRI for quantitative evaluation. 114

6.3 Average value of PSNR (dB) and SSIM for T1-T1c translation on subset dataset. Synthesized images generated by CycleGAN and starGAN are compared with corresponding ground-truth MRI for quantitative evaluation. 117

Asbtract

Chronic kidney disease and kidney failure are a major public health issue. In the context of a constantly increasing number of cases, kidney transplantation is considered an optimal management strategy. It has the advantage of increasing the chances of survival together with a higher quality of life and a reduced cost.

For kidneys intended for transplantation, it is essential to quickly determine the functional status and the optimal method of preservation, both of which remain open problems at this time. Surgeons may use suboptimal kidneys or exclude potentially better transplants. More generally, the choice of imaging techniques to analyze the kidney in different clinical tasks, including transplantation, is a research subject in its own right. Magnetic resonance imaging (MRI), in particular, has great potential as a non-invasive method for retrieving structural and functional information. However, the quantity and complexity of the data it generates remains an important obstacle to its full exploitation.

Machine learning in general, and deep learning in particular, are widely studied scientific items that can be found in many applications and research fields. Learning-based methods allow a computer to build complex concepts from simpler ones. Recent advances in medical imaging and machine learning have prompted many researchers to pursue the idea of augmented anatomical and functional imaging to aid in diagnosis. By augmented imaging, we mean artificial intelligence (AI) models designed to assist radiologists and allow them to make an optimal diagnosis.

Our PhD thesis work aims to improve the quality assessment of kidney grafts using MRI and machine learning techniques. This work includes three applications belonging to two main tasks: super-resolution and ultra-high field MRI synthesis for image quality improvement; and cross-modality translation. Note that for practical reasons explained in the document, a significant part of our work was carried out using human brain data.

In the first application, we develop a method based on self-supervised models to solve super-resolution tasks on routine 3T MRI through learning on paired and unpaired data. The evaluation of our results shows that the proposed methods can produce high resolution output from low resolution input with low distortion. Furthermore, the explored solution overcomes the limitation of existing methods requiring aligned sample pairs.

In the second part, we aim to synthesize ultra-high field (7 Tesla, or 7T) MRI data from 3T volumes. The proposed model obtains convincing results on both objective and subjective criteria. The final models can work stably on 3D MRI volumes, which is very promising.

In the last work, we focus on MRI cross-modality translation task. The models are designed to generate high precision volumes among different modalities such as T1↔T2, T1↔T1c or T1↔T2-Flair. Current work focuses on the translation of T1 MRI to its enhanced contrast version T1c, this scenario presenting a very strong potential with respect to the precautionary principle with regard

to gadolinium injections for obtaining T1c sequences. A comparative study between the methods of the literature and our methods from previous work is presented. The results demonstrate that our methods obtain a stable result on the research dataset and promising results on the practical dataset. Moreover, experiments have shown that the results of the models are optimizable.

Introduction

The introduction of this thesis manuscript starts with the motivations behind this Ph.D. project. The joint laboratory I3M is presented, along with its research teams and objective. Next is the introduction in terms of context and main challenges. Then, it will be thesis objective and the overview of thesis structure.

Motivation

The work presented in this thesis is hosted by I3M Joint Laboratory - a collaboration between Siemens Healthineers France, the XLIM Research Institute UMR CNRS 7252, and the Laboratory of Mathematics and Applications (LMA) UMR CNRS 7348 from the University Poitiers. The I3M laboratory is a consortium space for basic and applied research and publications.

In November 2019, the CHU Poitiers was equiped an ultra-high-field magnetic resonance imaging (MRI) at 7 Tesla MAGNETOM Terra, which provides access to high resolution molecular and metabolic imaging, making it possible to measure the structure and function of organs distinctively. This imaging can give hope for significant progress in the in vivo study, in a non-invasive way, of many pathologies, in their clinical management and the monitoring of their evolution, under treatment, possibly after surgery, in real-time. The goal is to perpetuate its scientific research by broadly investigating its benefits on several public health issues.

Teams in I3M laboratory are composed of DACTIM-MIS from the LMA and ICONES from the ASALI axis from XLIM. Hosted at Center Hospital University of Poitiers (CHU), members will thus have access to a high-field MRI platform for research and clinical use here. The DACTIM-MIS team within the LMA is composed of CHU staff and mathematicians specializing in bio-statistical and realistic models for modeling brain metabolism and tumor metabolism. ICONES team from XLIM has internationally recognized expertise in multi-variate images, particularly textured ones, using vector and bio-inspired approaches, and is developing complementary skills for machine learning algorithms, particularly in the context of medical imaging. Teams have known and worked together since 2012 as part of the MIRES federation. The collaboration is, therefore, quite natural, with the CHU, within the I3M laboratory, whose objective is to implement innovative Artificial Intelligence (AI) techniques for the processing and automatic analysis of multi-modality images to aid in the diagnosis and therapeutic monitoring of pathologies:

1. Brain (neuro-oncology, psychiatry, studies of neurodegenerative diseases, research in cognitive sciences, pharmacology with studies of the central diffusion of antibiotics by intracerebral micro-dialysis)

2. Heart (accidents and cardiovascular diseases)

3. Kidney (study of the graft before removal, follow-up of the graft in the transplant recipient)

Context and issue

Kidneys are vital organs that filtrate blood from waste and extra water caused by the normal functioning of the human body. When the functions of the kidneys are not working correctly, the latter is gradually poisoned by this waste, which is kidney failure. It is considered to be chronic when the loss of function is progressive, and the lesions present in the kidneys have an irreversible character. Following the French Renal Epidemiology and Information Network (REIN), it affects more than 80,000 people treated for chronic end-stage renal failure, 2/3 being treated with dialysis and 1/3 by transplantation. The number of affected patients increases by 2% each year. Moreover, the number of people with kidney disease presenting no symptoms would be around 10% of the French population, according to estimates current.

Chronic kidney disease (CKD) describes the gradual loss of renal function. It is a long-term condition where the kidneys cannot work the way they should. CKD reaches end-stage renal disease (ESRD), where renal dialysis or transplant is required.

From the clinical point of view, ESRD is the terminal of CKD which affects approximately 10,000 patients per year and requires replacement therapy; this number increases by around 4% per year. This evolution of CKD represents a problem of significant public health, intimately linked to arterial hypertension, diabetes and/or the syndrome metabolism, and cardiovascular pathologies. At the social and economic level, CKD is associated with severe health, personal and professional consequences, and very high costs [17].

In the context of the increasing cost of management of ESDR, reports of the High Authority for Health and the Biomedicine Agency have modeled the possibilities of change in the trajectory of patient care and assessed the consequences from clinical and economic aspects. Report confirming that the development of kidney transplantation in all age groups is an efficient strategy compared to all the strategies evaluated [26]. Patients who receive dialysis have an expected remaining lifetime of 6.8 years, compared with 17.8 years for transplant recipients. In addition, kidney transplantation is effective for medical and health care savings compared with dialysis by approximately 100,000 euros per year from the second year. It also offers improved quality of life [107]. Moreover, dialysis patients are known to experience accelerated atherosclerosis, and there are several inflammatories and atherogenic factors due to the increased cardiovascular risk proportional to the increase in serum creatinine, suggesting that renal failure correlates with, if not causes, accelerated vascular and metabolic defects that predispose patients to cardiovascular death. The better outcomes of patients with preemptive transplants and with a shorter time on dialysis underscore the importance of early referral and evaluation for renal transplantation. Patients with ESRD benefit from transplantation as early as possible to maximize their potential for extended survival after transplantation.

The major challenge in organ transplantation is the preservation of grafts. Kidneys from extended criteria donors (ECD) and donations after circulatory death (DCD) donors are increasingly used worldwide, providing good post-transplantation outcomes [197]. ECDs have risk factors for poor function after transplantation, such as higher donor age, a history of hypertension, increased serum creatinine, and death from a cerebrovascular accident. For a kidney intended for DCD transplantation, the quality is usually linked to the ischemia-reperfusion injury (IRI), causing dysfunction and/or loss of the graft [176, 204]. The extent of this damage is related to the situ-

ation of the donor. The simplest method of dealing with the destructive cellular processes after donor death has been to cool the kidneys. It slows down the metabolism but does not completely stop it, leading to ongoing damage in cold-preserved organs, hence the critical importance of limiting cold ischemic times. Assessment of graft quality in these marginal graft categories is complex, and exact criteria predicting graft injury and function after transplantation are lacking [47].

Grafts lead more frequently to a delayed resumption of function, such as the delayed graft function (DGF), or even to the loss of the organ and a return of the patient to dialysis. Hence, surgeons usually use sub-optimal kidneys (reduced graft and even patient lifespan) or exclude potentially better grafts. Consequently, high graft discard rates are reported for all types of marginal renal grafts, even for those that might be appropriate for transplantation.

Before transplantation, the determination of the functional status is critical. Kidney function has been commonly evaluated either by estimating the Glomerular Filtration Rate (GFR), which is based on the blood serum creatinine level or by invasive biopsy. However, blood serum creatinine level is a late indicator of renal impairment. Moreover, although the biopsy represents the gold-standard method for assessing renal structure, it is still limited to its invasiveness and risk to the patient (e.g., bleeding, pain), especially for CKD patients who are subjected to multiple biopsies. Moreover, the optimal method of preserving the kidney remains an entire problem. If a visual examination is essential, the criteria used to reach the right decision are still vague and challenging to explain.

On the other hand, radiology was reported to help with follow-up renal diseases. Medical images such as magnetic resonance imaging (MRI) and computed tomography (CT) play an increasingly more critical role in assessing renal function. They have provided structural, functional, and molecular information that can detect the alteration in renal tissue properties and functionality and help predict and diagnose renal function. The use of imaging techniques to analyze the kidney in different clinical tasks, including that transplantation, is becoming potential research topics. Before, ultrasound and CT were modalities of the first choice in renal imaging. Ardakani et al. [7] proposed a pipeline for analyzing images from ultrasounds to monitor and evaluate possible complications in the follow-up of transplant patients by automating the assessment process through statistical analysis of image textures. Studies in [164, 54] target the characterization of renal masses by computed tomography to diagnose solid renal tumors. The work focuses on kidney transplantation, especially the evaluation of the quality of a kidney before its transplantation. In fact, there are few studies that have been proposed in this field. Fananapazir et al. [57] introduce experiments to determine whether MRI could more confidently characterize indeterminate small renal lesions (< 15 mm) previously seen on CT scans of potential renal donor patients and whether such characterization could impact surgical management and donor candidate status.

Along with the development of medical imaging techniques, MRI becomes more popular in clinical and research center. It provides a non-invasive assessment of body anatomy and physiology for health and disease examination, while maintaining superior contrast resolution on soft tissues. At this moment, MRI has mainly been used as a problem-solving technique. Currently, clinical magnetic resonance examinations are mainly deployed at 1.5 and 3 Tesla magnetic fields. Ultra-high field (UHF) MRI *e.g.* 7-Tesla (7T) or higher devices were recently introduced, allowing better sensitivity of signal-to-noise ratio (SNR) and higher spatial resolution compared to 3-Tesla (3T) or 1.5T MRI [149]. Ultra-high field MRIs can be used to visualize physiological/pathophysiological consequences and to resolve structures more precisely that would be difficult to detect at lower field strengths. MRI with high quality is preferred in the clinical and research domains because it can provide structural details critical for identifying and determining biomarkers through accurate

image analysis.

For the last two decades, machine learning has been coming into its own, with a growing recognition that it can play a key role in a wide range of critical domains, such as computer vision, natural language processing, data mining, and expert systems. Machine learning can be defined as a set of algorithms that have the ability to learn and improve from experience without being explicitly programmed for a specific task. Later, deep learning with neural networks was applied to solve machine learning tasks with more advantages by introducing more straightforward and meaningful representations. Deep learning enables the computer to build complex concepts out of simpler concepts. The use of learning-based methods has been increasing rapidly in the medical imaging field, including computer-aided diagnosis, radiomics, and medical image analysis. Popular tasks of machine learning for in computer-aided diagnosis can be mentioned, such as classification, segmentation, synthesis, etc. By performing quantitative analysis on conventional medical images, machine learning in general and deep learning in particular hold the potential to turn them into a fully-automated tool that can assist radiologists and clinicians in prognostic and diagnostic.

Objective

Recent medical imaging and machine learning advantages inspired many innovative researchers to use anatomical and functional imaging for diagnostic assistance. In this context, the requirements for the development of a non-invasive assistant have appeared to allow practitioners to carry out their diagnosis of the quality of a kidney intended for transplantation in optimal conditions. Currently, the use of prediction models for kidney disease is still in its infancy, and further evidence is needed to identify its relative value. AI models are not approved to replace radiologists medical decision-making; instead, they can assist them in providing optimal diagnoses for their patients.

The primary objective of the thesis is to improve the assessment of renal grafts using medical imaging techniques, especially MRI. With the support of the Siemens Healthineers MRI system from I3M laboratory at CHU Poitiers, we expect to collect practical MRIs to implement the research tasks.

In the scope of the project, the thesis mainly focuses on MRI instead of ultrasound or CT. The work aims at supporting radiologist diagnosis on MRI by solving two missions: medical image quality enhancement and cross-modality translation. For each task, we propose a specific method to solve the problem and make a comparison with other baseline methods in order to evaluate method performance.

In terms of MRI quality enhancement, the goal is to improve the quality of routine 3T MRI by doing two main different tasks:

- Super-resolution (SR) is a topic that aims to reconstruct HR images from given LR images. In the past decade, various super-resolution methods have been widely applied to medical images to increase the spatial resolution of scans after acquisition has been performed. In this thesis, SR is considered a tool to enhance the spatial resolution of routine MRI.

- MRI synthesis is based on a paradigm shift where a transformation model can learn to regenerate images from a given input domain into another desired domain. Applications of image synthesis in the field of MRI can range from cross-modality translation within single

types (*i.e.* MRI T1 $\leftrightarrow$ T2) or between different types of medical images (*i.e.* CT $\leftrightarrow$ MRI) to field-strength conversion (*i.e.* 3T $\leftrightarrow$7T) [219]. In this thesis, the topic of work is ultra-high field MRI rendering, in which we try to produce realistic UHF MRI from routine MRI.

Each modality provides a unique view of intrinsic MR parameters for the cross-modality translation, but different factors limit the existence of complete multi-modality MR images. For example, in brain MRI, T1-weighted images observe structures on different white and grey matters, and contrast-enhanced T1 (T1c or T1-gado) can be used for assessment of the change of tumor shape with its enhanced demarcation around the tumor and T2-weighted images are utilized for locating tumors from cortical tissue. At the same time, contours of the lesion can be delineated clearly on fluid-attenuated inversion recovery (Flair) images. Hence, integrating the strengths of each modality can help explore rich underlying information of tissue that facilitate diagnosis and treatment management. Moreover, deep neural networks provide a generic solution for producing cross-modality images. Gadolinium-enhanced T1 MR imaging provides excellent delineation of the renal structures and boundaries that increase the segmentation of the kidney from the surrounding anatomical structure and the measurement of the renal and renal cyst volume. However, gadolinium-containing contrast agents may increase the risk of a rare but serious disease called nephrogenic systemic fibrosis in people with severe kidney failure. In this work, at this moment, we first focus on multi-contrast cross-modality MRI translation, which aims to transform T1-gado MRI from T1 MRI. It will be valuable research if we can produce realistic gadolinium-enhanced T1 MRI without invasion.

However, the thesis encountered specific difficulties. First, The COVID-19 epidemic happened at the beginning of the thesis process and became very serious in 2020-2021. It made it challenging to access CHU due to the priority of COVID quarantine and isolation. Moreover, the UHF MRI scanners have been set up since November 2019, and the parametric testing took time to complete, while acquiring kidney MRIs is not the most priority for community benefits. Due to all that reasons, the available resource for kidney MRIs necessary for research purposes was insufficient. On the other hand, brain MRI resources are abundant and diverse, many available for the 3T MRI scanner and some for the new 7T scanner.

Besides, the objective of the thesis is to improve the kidney assessment to support radiologists making decisions by enhancing the quality of MRI in terms of spatial resolution, field strength, and multi-contrast. With the advantage of superior contrast MRI on soft tissues, we believe that improving the general quality of MRI can also help improve the quality of kidney MRI. Hence, we decided to use brain MRI instead of kidney MRI to achieve the most optimized results.

Thesis overview

This manuscript is organized into three parts:

1. Basic medical imaging concepts and deep learning fundamentals (chapters 1, 2, and 3).
2. Model selection and implementation (chapters 4, 5, and 6).
3. Applications and Future work (chapters 7).

Chapter 1: Medical imaging for kidney disease diagnosis

The first chapter of the thesis briefly introduces anatomy and function of kidney within the context of CKD. Then magnetic resonance imaging comes as an optimal method for non-invasive diagnosis. Additionally, the development of ultra-high field MRI and computer-aided MRI analysis are also presented as the next step for medical image analysis.

Chapter 2: Learning-based for MRI synthesis

The second chapter firstly presents the fundamental knowledge of machine learning. Then, concept and terminology of deep learning/deep neural networks are demonstrated, including neural network components and short explanation of network training.

Chapter 3: Generative model

The third chapter focuses on generative models based on deep neural networks to learn the representations of complex data. We start with autoencoders - primary generative models, then extend to adversarial models and their variations.

Chapter 4: MRI Super-resolution

This chapter demonstrates MRI quality enhancement in terms of super-resolution. After addressing the current context and challenges in practice, we propose a method to solve the super-resolution on either paired/unpaired MRI as the first phase of research. Details of the model architecture, and training procedure, are presented, along with the study of its performance.

Chapter 5: Ultra-high field MRI synthesis

The advantage of 7T MRI is promising compared to the current 3T MRI and the research topic challenge. This chapter tackles the development and implementation of MRI synthesis. in this context, the MRI synthesis is ultra-high-field MRIs rendering from routine data. Details of the model architecture and training procedure are presented, along with their performance.

Chapter 6: MRI cross-modality translation

This chapter presents multimodal cross translation on routine MRI dataset. We provide a comparative study between methods of the literature and our methods to synthesize MRI between T1 and T1-gado.

Contributions

Published papers

- Do, H., Bourdon, P., Helbert, D., Naudin, M., Guillevin, R. (2021). 7T MRI super-resolution with Generative Adversarial Network. Electronic Imaging, 2021(18), 106-1.
- Do, H., Helbert, D., Bourdon, P., Naudin, M., Guillevin, C., Guillevin, R. (2021, October). MRI super-resolution using 3D cycle-consistent generative adversarial network. In 2021 Sixth International Conference on Advances in Biomedical Engineering (ICABME) (pp. 85-88). IEEE.

Submitted papers

- Do, H., Bourdon, P., Helbert, D., Naudin, M., Guillevin, R. (2022). Realistic ultra-high field MRI rendering using cycle-consistent generative adversarial networks.

Chapter 1

Medical imaging for kidney disease diagnosis

1.1 Renal anatomy and function

The kidney is essential to the urinary system, purifying blood and eliminating waste. These functions are impaired or lost when the kidneys fail, affecting homeostasis. Affected individuals may exhibit weakness, lethargy, shortness of breath, anemia, widespread edema, metabolic acidosis, elevated potassium levels, and cardiac arrhythmias. The urinary system, which is controlled by the nervous system, stores urine until the proper time for elimination and then provides the anatomical structures necessary to transport this waste fluid out of the body. Incontinence is caused by the failure of nervous control or anatomical structures, resulting in loss of control over urination.

This section provides basic concepts of kidney anatomy and function. Then, kidney disease and common symptoms of kidney failure are presented along with the necessity of kidney transplantation in treatment. Magnetic resonance imaging (MRI), a non-invasive method, is becoming increasingly popular for medical diagnosis. Besides, ultra-high field MRI and computer-aided applications are rapidly developing, bringing the massive potential for medical image analysis.

1.1.1 Kidney anatomy and function

Kidneys are located on both sides of the spine between the parietal peritoneum and the posterior abdominal wall in the retroperitoneal space. The right kidney is lower due to a slight liver displacement, whereas the left kidney is located roughly between the T12 and L3 vertebrae. The eleventh and twelfth ribs provide some protection for the upper kidneys. The typical size of the kidney is 11-14 cm in length, 6 cm in width, and 4 cm in depth. The kidney weighs approximately 125-175 grams in males and 115-155 grams in females [22].

Kidneys are encased by a fibrous capsule composed of irregular connective and dense tissue to maintain shape and protection. A layer of shock-absorbing fatty tissue covers the capsule, which is surrounded by a tough renal fascia. In a retroperitoneal position, the fascia and, to a lesser extent, the overlying peritoneum serve to firmly anchor the kidneys to the posterior abdominal wall. The kidneys are well-vascularized and maintain approximately 25 percent of the resting cardiac output. The adrenal gland is located on the superior surface of each kidney. The adrenal cortex influences kidney function directly by producing aldosterone, a hormone that stimulates

sodium reabsorption.

Figure 1.1 it demonstrates the internal anatomy of a kidney. The outer region of the kidney is the renal cortex, while the inner region is the medulla. The renal columns are connective tissue processes that radiate downward from the cortex through the medulla, separating the medulla's most distinctive features, the renal pyramids, and renal papillae. The papillae are collections of collecting ducts that transport nephron-produced urine to the renal calyces for excretion. In addition to dividing the kidney into lobes, the renal columns provide a framework for the vessels that enter and exit the renal cortex. Pyramids and renal columns form the renal lobules collectively [22].

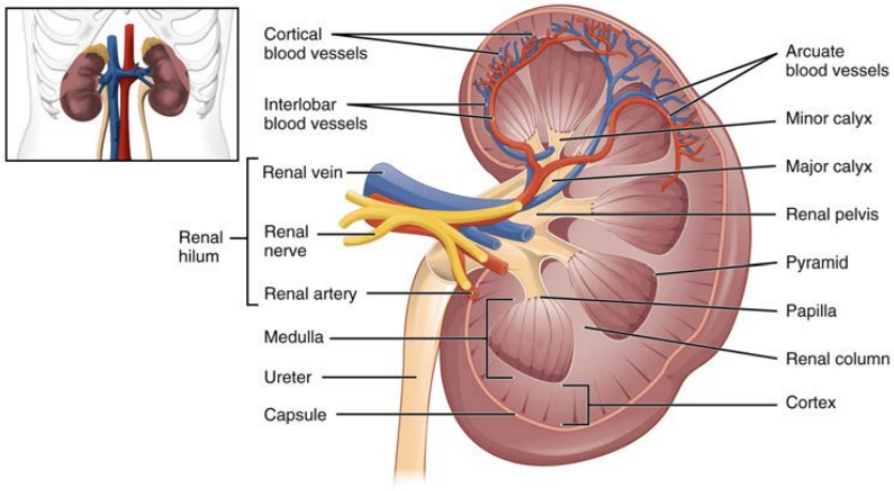


Figure 1.1: Kidney anatomy. [22]

Approximately 2 to 3 million tufts of specialized capillaries - the glomeruli - are distributed more or less equally between the two kidneys in their ability to filter blood. Blood cells, platelets, antibodies, and protein are excluded because the glomeruli primarily filter the blood based on particle size. The glomerulus is the initial component of the nephron, which then transforms into a highly specialized tubular structure responsible for the urine final composition. All other solutes, including ions, amino acids, vitamins, and waste products, are filtered to produce a filtrate whose composition resembles that of plasma. The glomeruli produce approximately 200 liters of this filtrate per day, while less than two liters of waste are excreted as urine [22].

In addition, kidneys share with other organs to do additional critical functions, such as regulating pH or blood pressure. Kidneys are essential for determining the solute concentration in red blood cells. In response to low oxygen levels, the kidneys produce 85 percent of the hormone erythropoietin (EPO). EPO increases oxygen delivery to tissues by stimulating the production of red blood cells in the bone marrow. In addition, the kidneys are involved in several intricate endocrine pathways and produce specific hormones. In addition to EPO, a decrease in blood flow to the kidneys stimulates them to release the enzyme renin, activating the renin-angiotensin-aldosterone (RAAS) system and promoting sodium and water reabsorption. The reabsorption increases both blood pressure and blood flow. The kidneys also contribute to vitamin D synthesis by converting calcidiol to calcitriol. It also regulates blood calcium levels by converting vitamin D3 into calcitriol, which is then secreted in response to the release of parathyroid hormone.

1.1.2 Kidney dissease and transplantation

Chronic Kidney Disease (CKD) is characterized by progressive and irreversible deterioration of renal function. The glomerular filtration rate (GFR) determines the volume of filtrate formed by both kidneys per minute. The heart pumps about 5 liters of blood per min under resting conditions and even more when exercising. Approximately 20 percent enters the kidneys to be filtered. Then, ninety-nine percent of the filtrate is returned to the circulation by reabsorption, producing only about 1-2 liters of urine per day. CKD is defined as decreased kidney function by GFR of less than 60 ml/min/$1.73m^2$, or markers of kidney damage [196].

Determination of GFR is one of the tools used to assess the excretory function of the kidney. GFR can be accurately determined by intravenous administration of inulin or by measuring naturally occurring creatinine. Inulin is a plant polysaccharide that is neither absorbed nor excreted by the kidney. Its occurrence in urine is directly proportional to the rate at which the renal corpuscle filters it. However, because measurement of inulin clearance is cumbersome in clinical practice, GFR is usually estimated by measuring naturally occurring creatinine, a molecule composed of proteins that is produced by muscle metabolism, is not reabsorbed, and is only marginally excreted by the nephron [196].

A patient reaches end-stage renal disease (ESRD) when GFR is less than 15 ml/min/$1.73m^2$, at which point kidney function is no longer able to sustain life over the long term. Options for patients with ESRD are kidney replacement therapy which includes renal dialysis and transplantation, or conservative care (or non-dialytic care) [222]. Failure of the renal anatomy and physiology can lead gradually or suddenly to renal failure. Symptoms of kidney failure can be mentioned as weakness, lethargy, widespread edema, anemia, metabolic acidosis and alkalosis, heart arrhythmias, uremia, or oliguria.

Stage of CKD	Description	GFR
1	Kidney function remains normal or increase GFR	>90
2	Mild decrease in kidney function	60-89
3a	Mild to moderate decrease in kidney function	45-59
3b	moderate decrease in kidney function	30-44
4	Severe decrease	15-29
5	Kidney failure	<15

Some conditions of the kidneys that may result in ESRD include: repeating urinary infections; kidney failure caused by diabetes or high blood pressure; polycystic kidney disease or other inherited disorders; glomerulonephritis, which is inflammation of the kidney filtering units; hemolytic uremic syndrome, a rare disorder that causes kidney failure; lupus and other diseases of the immune system; obstructions. Renal failure also increases mortality from cardiovascular disease and causes directly resulting from renal failure, including fluid and electrolyte imbalance and uremia.

There are two treatment options for kidney failure: dialysis (hemodialysis or peritoneal dialysis) and kidney transplantation. Although dialysis addresses the immediately life-threatening complications of renal failure, it does not provide fluid and electrolyte homeostasis comparable to that of a well-functioning kidney. Several metabolic functions of the kidney, such as vitamin D synthesis and erythropoietin synthesis, are also not regulated appropriately in the absence of a well-functioning kidney. On the other hand, kidney transplantation offers several advantages for

patients suffering from EDSR when compared with dialysis [205]

However, the major limitation is the preservation of kidney grafts for transplantation [138]. The delayed graft function (DGF), and early graft loss are increased in transplantation [167]. Various studies indicate that ischemia, especially with increased age and prolonged cold ischemia times, are independent risk factors for primary nonfunction, DGF, and graft failure in kidney transplantation [154]. Assessment of graft quality in these marginal graft categories is complex, and exact criteria predicting graft injury and function after transplantation are lacking.

1.2 Magnetic Resonance Imaging for non-invasive diagnosis

The phenomenon of nuclear magnetic resonance (NMR), introduced by Bloch F. and Purcell E. in 1946 [25, 159], is the principle of Magnetic Resonance Imaging (MRI). The founding principle of imaging techniques was introduced in 1949. However, the first NMR images were successfully demonstrated in 1973 [114].

MRI is now a mature analytical method that is widely used as a diagnostic tool in clinical medicine and research. It is non-invasive, and does not use ionizing radiation. MRI works in almost all cases based on the sensitive interaction with hydrogen, the main component of any biological organ. Therefore, there are almost no limitations on the samples of biological origin that can be imaged. On the other hand, only bone tissue, which contains less and more tightly bound hydrogen than most other body parts, provides an inherently low amplitude signal.

The contrast in an MRI image is due to the fact that the hydrogen atoms in different tissues and compounds have slightly different chemical and magnetic environments. Therefore, they respond somewhat differently to radio waves in the form of short radio frequency pulses (RF) sent into the object under study. This makes it possible to detect pathological changes deep inside an object. The greatest advantage of NMR over similar diagnostic imaging techniques such as X-ray, computed tomography (CT) and ultrasound (US) is therefore the contrast it provides between diseases and tissues.

In addition, MRI techniques are becoming more advanced. In the beginning, it took several hours to produce an image, a process that can now be done in minutes or even seconds. The faster and more reliable equipment now commercially available has given physicians a valuable diagnostic tool. New pulse sequences are constantly being invented for specific tasks and improved contrast. Faster scanning routines allow real-time imaging of dynamic processes such as blood flow and drug metabolism in the body. Magnets are also improving, but at a slower pace. There is a trend toward lighter and better-shielded magnets that make the stray field near the device weaker, despite an increased working field.

However, there are still some disadvantages of MRI. Because MRI uses radio waves and strong magnetic fields, imaging is not possible in patients with pacemakers or various types of metallic implants in the body. In addition, the machines are still heavy and expensive and require a lot of space. They are usually housed in separate buildings. In addition, professional judgment and skill are still required when evaluating the images and choosing imaging parameters to achieve the desired contrast.

In this chapter, we explain how the phenomenon of magnetic resonance enables the acquisition of a physical signal starting from the basic structure of matter. This signal is then converted into an image whose properties. Finally, we explain how the addition of different NMR acquisition

modalities transforms MRI into a multivariate and information-rich exploration tool.

1.2.1 Nuclear magnetism

Nuclear magnetism refers to the magnetic moment property of an atom. A nucleus with an even number of protons and neutrons will have no magnetic force, but a nucleus with an odd number of protons or neutrons carries charges and magnetic resonance.

The hydrogen nucleus comprises a solitary proton with very high intrinsic magnetic properties. It is considered an active nucleus, in which the proton and neutron spins do not cancel out each other, having a net spin on itself. This spin can take two values. Depending on its direction, it can be positive or negative. The charge of the proton, coupled with the rotation of the nucleus, contributes to creating a magnetic moment specific to the proton. This magnetic moment is a vector quantity, collinear with the axis of rotation of the proton. Thus, the proton is comparable to a magnetic dipole due to its charge, rotation, and magnetic moment. Subjected to an electromagnetic field, it orients itself like a magnet. MRIs use hydrogen because the human body contains around 70% of hydrogen, which is its potential for nuclear magnetization and relative abundance in the body in water and fat.

Nuclear magnetic resonance

Based on the quantum-mechanical phenomenon, each nucleus rotates around its axis, inducing an angular momentum. Nuclear magnetic resonance studies the variations of magnetization of nuclei in electromagnetic fields.

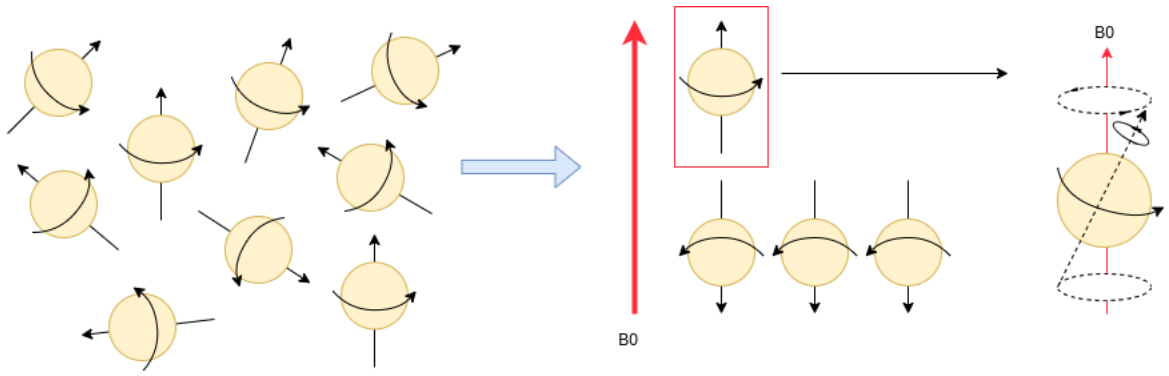


Figure 1.2: Hydrogen spins around its axis, producing an angular momentum. Spins are randomly oriented until the application of an external magnetic field B0. Spins align with B0 in a parallel and anti-parallel direction with an excess of spins oriented anti-parallely, giving a net magnetization vector (NMV) aligned in the opposite direction with B0.

A tiny piece of biological tissue contains billions of billions of hydrogen atoms. The nuclei of the hydrogen atoms - the protons - act like tiny compass magnets, generally having random orientation and equal energy. For every magnet pointing in any direction, there is another pointing in the opposite direction. In this way, they balance each other's magnetic moments, and there is no external magnetic effect.

However, if one places a sample containing hydrogen in a magnetic field B_0, the magnetic field defines a direction in space, which by convention is the longitudinal z-axis. The magnetic moments

of all protons aligned along this axis add up to give a macroscopic magnetization vector. Half of it is parallel to the field, the other half is antiparallel to the field. The protons whose magnetization vectors are parallel to the field have slightly lower energy than the others, and therefore there is a small but significant difference in population between the two energy levels. Thus, the magnets no longer exactly cancel each other out. This effect is called the net magnetization vector (NMV), denoted by Mz. In the xy-plane, the magnetic moments of the protons still cancel. The magnitude of Mz, and correspondingly the amplitude of the MR signal, is proportional to the proton density, the external magnetic field, and the square of the gyromagnetic ratio. This has led to moving to higher and higher fields.

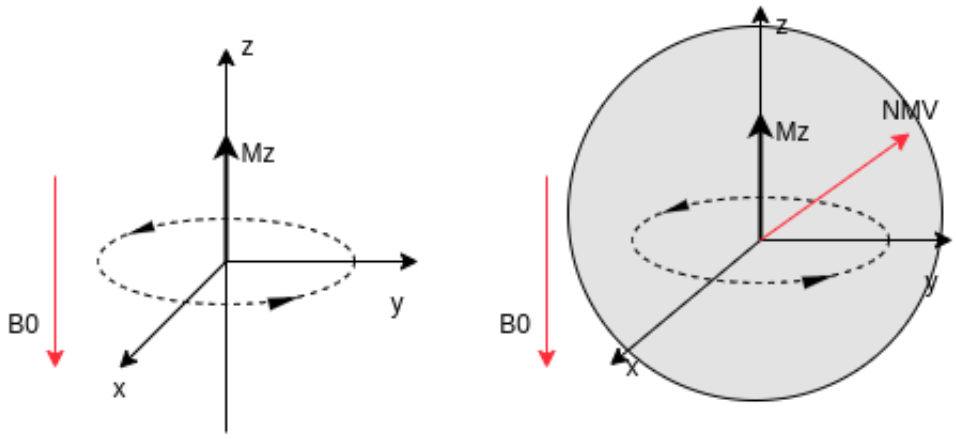


Figure 1.3: The net magnetization vector aligned with the external magnetic field B0. After the excitation by a radio-frequency pulse, the NMV flips by a certain angle following a spiral trajectory and imposes transversal and longitudinal components.

The motion of simple precession refers to the rate of precession of the magnetic moment of the proton around the external magnetic field, which is described by the Larmor equation:

$$\omega_0 = B_0 \times \gamma \tag{1.1}$$

where ω_0 represents the precessional frequency, and γ is the atomic gyromagnetic ratio.

When the nuclei start to process, their protons are immersed and subjected to electromagnetic fields B_0 until the net magnetization has reached a steady state of equilibrium. The precessional path around the magnetic field is circular, like a spinning top. While rotating on their axis, the protons turn around B_0 in the shape of a double cone, oriented positively or negatively according to the nuclear spin [23, 59]. The next stage is to apply an electromagnetic radio frequency (RF) field at the resonance Larmor frequency, which will transfer energy to the protons. As a result, the net magnetization vector is tilted away from the z-axis, with changes in directions described using a rotation frame with three axes.

The effects of B0 disappear, and the precession of NMV is now locked to RF pulses at Larmor frequency, which makes it tilted and consequently tipped within the X-Y plane. Depending on the duration of the RF excitation pulses, the NMV goes more or less precess around its transverse axis. In practice, one uses impulses causing rockers of 90° or 180°. The flip angle, often represented by α, is determined by both the strength and duration of the RF field/pulse. [30]. When the RF pulse stops, the magnetization vector will gradually return to the state of equilibrium. The energy

accumulated during the resonance will be restored in the form of a wave: this is the phenomenon of relaxation.

Relaxation

The key of MRI is to flip the NMV from the parallel alignment by applying excitation, and then the vector begins to drive toward its initial state and realign with the external magnetic field during the relaxation process. When the RF pulse is turned off, spins lose their phase coherence. Nuclei return to a state of equilibrium and the restitution of the energy accumulated during nuclear magnetic resonance.

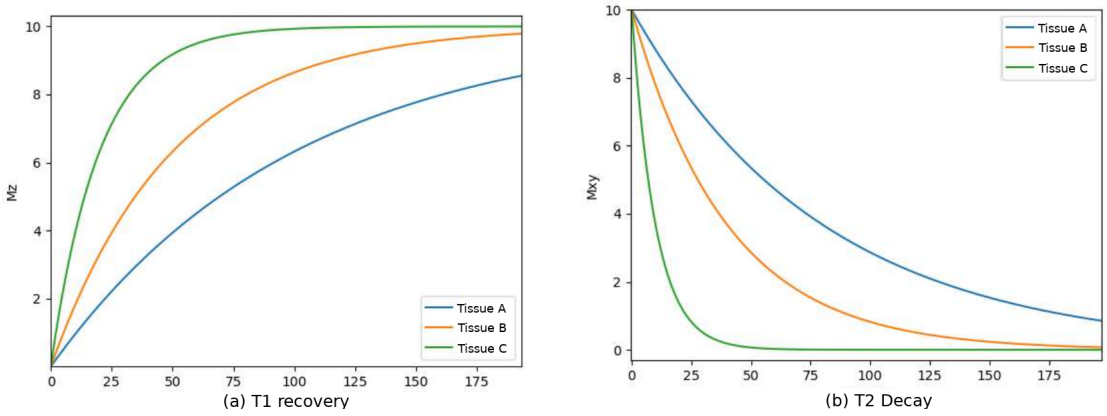


Figure 1.4: Examples of transversal (a) and longitudinal (b) magnetization relaxation for different tissues

There are two types of relaxation, occurring simultaneously but independently of each other: T1 and T2. Both are time constants, with T1 describing the return to equilibrium as vector realigns along the z-axis, whereas T2 characterises the decay of the signal as the excited protons begin to dephase, with changes occurring in the x-y planes. Follow exponential laws, the T1 is the time necessary for 63% of the longitudinal regrowth to have taken place, while T2 is the time necessary for 63% of the transverse decrease to take place.

T1-relaxation time, which is unique to each tissue, is determined by how protons are bound and is generally longer at higher field strengths. For example, in adipose tissues, the protons are tightly bound and will release the energy in their surroundings much faster than loosely bound protons. In other words, the speed at which tissues release the energy/relax will determine the value of T1. The time constant T1, modelled as an exponential growth curve, corresponds to the time needed after the excitation pulse for M to reach 63% of its initial value. Finally, the fact that T1 values vary with tissue types is also the rationale behind the good contrast resolution in MRI scans. The physical process behind longitudinal relaxation is the loss of energy from spins to lattice. T1 is related to the amount of lattice and the probability of interaction between spins and lattice. Thus, fat has a shorter T1 than water. This exponential of t1 relaxation is given by:

$$M_z(t) = M_z(0) \times (1 - e^{-\frac{t}{T1}}) \tag{1.2}$$

On the other hand, the T2 relaxation also occurs after the RF pulse has been applied, but the spins have been tilted in the x-y plane, and all proton spins are synchronized and precessing

at the same frequency. This is the stage where protons are in phase. As mentioned above, it is the loss of this synchronization or dephasing. The T2 relaxation is also called spin-spin relaxation because interactions between spins cause dephasing. There will be the transfer of energy from excited protons to nearby non-excited ones. This affects the speed at which each proton spins and causes progressive inhomogeneity that leads to signal decay. Similar to T1 relaxation, the signal decay occurring during T2 relaxation can be modelled as an exponential curve, similar to radioactive decay. The time constant T2 corresponds to the length of time elapsing between the excitation and the point at which the signal has been reduced to ∼36.8% of its original value. Moreover, fluids (e.g. water) have a long T2 since the probability of energy exchange between spins is relatively low compared to tissues that have a greater degree of binding spins within a matrix, which increases the probability of spins interaction. This exponential decay is given by:

$$M_{xy}(t) = M_{xy}(0) \times e^{-\frac{t}{T2}} \tag{1.3}$$

MR signal is linked to T2 relaxation. This signal is called free induction decay (FID). Its initial amplitude is determined by the degree to which NMV has been flipped on the xy plane, with the highest signal obtained when the vector has been flipped to 90°. The signal is modelled with a decay curve containing the actual signal. The signal itself is oscillating at the resonance frequency in the MHz range. However, in practice, the magnet is likely to have some flaws in its manufacture, and tissue variability means that each tissue has a different magnetic susceptibility, which causes field distortions at tissue borders. As a consequence, the signal decays faster than the T2 relaxation would predict, and the actual signal is called T2*.

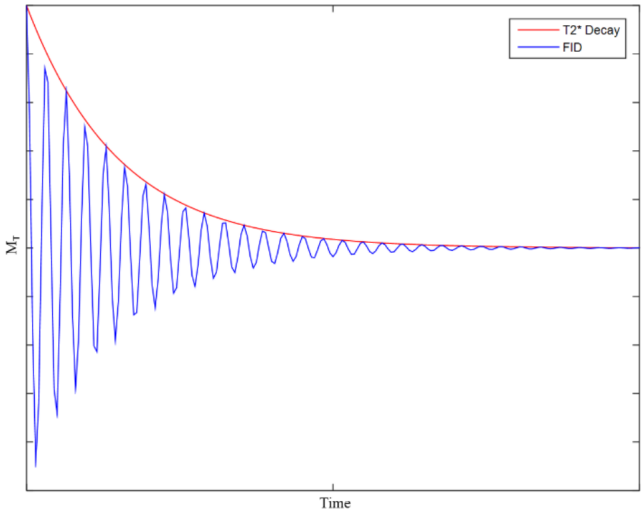


Figure 1.5: FID and T2* decay obtained by fitting the curve of FID

Spin echo

A Spin Echo (SE) is generated when a second 180° RF-pulse is applied a short time after the 90° one. The effect of this pulse is to rotate the entire system upside-down, causing the spins to rephase and thus producing a large signal: the Spin Echo. In order to have the optimal effect, the 180º pulse has to be applied at a specific time in the sequence. Corresponding to the middle time point between the first RF pulse and the peak of the spin-echo. This time interval is called

echo time (TE). The time at which the 180° pulse must be applied is defined as half of echo time TE/2.

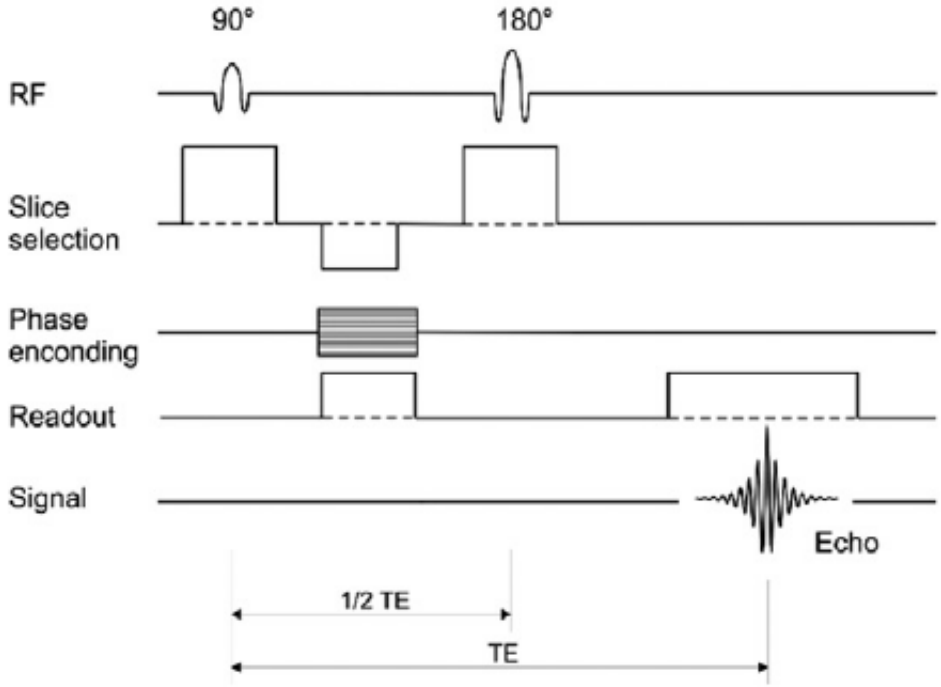


Figure 1.6: Diagram of a spin echo pulse sequence

In terms of protons, when the system has been turned on its head, it is faster precessing spins. A good analogy that is often used is one of a race in which participants have a great variety of speeds. At the start of the race, they are all aligned, in MRI terms, in phase. Once the race starts (at t = 0), the contestants start moving at their fastest pace and soon find themselves at different spots. At a given point (TE/2 in the MRI sequence), It turns around without losing speed, which means the starting line is now the finish. As the fastest contestants will be furthest away from the starting/finishing line and the slowest ones the closer, if they all keep going at the same speed as they did before, they should all reach the finishing line simultaneously. For the protons, this will be another TE/2, after which they will be in phase once more.

Once all the spins are back into phase, they immediately start to dephase again. However, a second 180° pulse can be applied, using the same TE, to generate a second echo. The process can be repeated until the time at which T2 relaxation has caused the signal to decay completely. It should be noted here that while the 90°-180° sequence gives the strongest signal, spin echoes can be generated with other flip angles.

On the other hand, the repetition time (TR) is the time elapsing between the repetitions of the sequence. So for the spin-echo sequence, it will be between the 90° pulses. Different TR and TE are also used to determine the contrast when reconstructing MRI.

1.2.2 Image reconstruction

So far, signal generation has been described based on NMR theory. However, to get solid volume imaging, this signal has to be transformed into an image.

We explained the nature of the NMR signal and the mechanism by which T2 relaxation releases stored energy. However, this signal needs to be spatially quantified and encoded to form images. A particular coding using three gradients and the inverse Fourier transform is applied. The slice selection gradient (GS) allows to define the thickness of the slice. Then, in the section or the selected volume, two gradients will be successively applied to distinguish coding the protons in line and in columns. A matrix is thus obtained, the rows and columns of which will be identified by the following two gradients. The phase encoding gradient (GP) is used to code the lines by assigning each proton in the volume a different phase per line. The frequency coding gradient (GF) is used to make it possible to code the columns by assigning a frequency to each column of protons.

Slice selection

It is done using a slice encoding or slice selection gradient, GS, together with a simultaneous RF pulse at the Larmor frequency in Equation 1.1 determined by the selected slice, as seen in Figure 1.7. Consequently, only the protons in the chosen slice will be excited. Since each slice contains a range of frequencies or bandwidth, the RF pulse transmitted needs to comprise the whole range. The thickness of the slice itself is determined by combining the strength/steepness of the gradient and the range of frequencies/bandwidth in the RF pulse. When the signal source has been located in a specific area, further encoding is necessary to know its position within the slice in the phase encoding gradient. The total magnetic field strength becomes a function of z and can be formulated as shown in Equation 1.5.

Phase encoding

The phase encoding gradient is applied for a specific period in the vertical direction, causing the protons to rotate at different frequencies depending on their position along the gradient. Precession will increase where the gradient increases and similarly decrease in the part of the slice where the gradient causes a decrease in the magnetic field. In other words, the protons will have different phases depending on their position, and this persists after the gradient is switched off. So now all the protons are precessing with the same frequency but have different phases. To obtain an image, multiple repetitions with different encoding gradients, which are progressively incremented, are necessary, as can be seen in Figure 1.7.

Frequency encoding

Last, the frequency encoding gradient is applied at the right angle from the previous gradient. It affects protons frequencies according to their positions along the gradient direction and therefore modifies the Larmor frequencies in the remaining direction for the duration of its application, while the phase changes from GP remain after the pulse is turned off.

Fourier transform and k-space

The reconstruction process involves the Fourier transformation of k-space in order to get an image from the signal. k-space is represented by a square, with k_x and k_y axes corresponding to horizontal and vertical axes in the actual image, respectively. However, the axes in k-space

represent spatial frequencies rather than positions, with each k-space point containing both phase and spatial frequency information about each pixel in the final image. The k_x axis in the k-space represents the time component, while the k_y axis represents the phase encoding direction. k-space data is converted into a grey-level image using the inverse Fourier transform in 2D space.

The k-space comprises signals with low spatial frequencies, which is essential for the delineation of contrast of the image, so the fact that different tissues are going to have different signal intensities; and the high-spatial frequencies signals are essential for the ability to define borders or edges of the image. The ability of the image to demonstrate contrast and edge definition depends on having more or less information in the k-space. The low frequencies are located in the center of the space, and the high frequencies are in the periphery. Each of these regions contributes to different aspects of the image. Thus, it is necessary to have a filled k-space prior to image reconstruction.

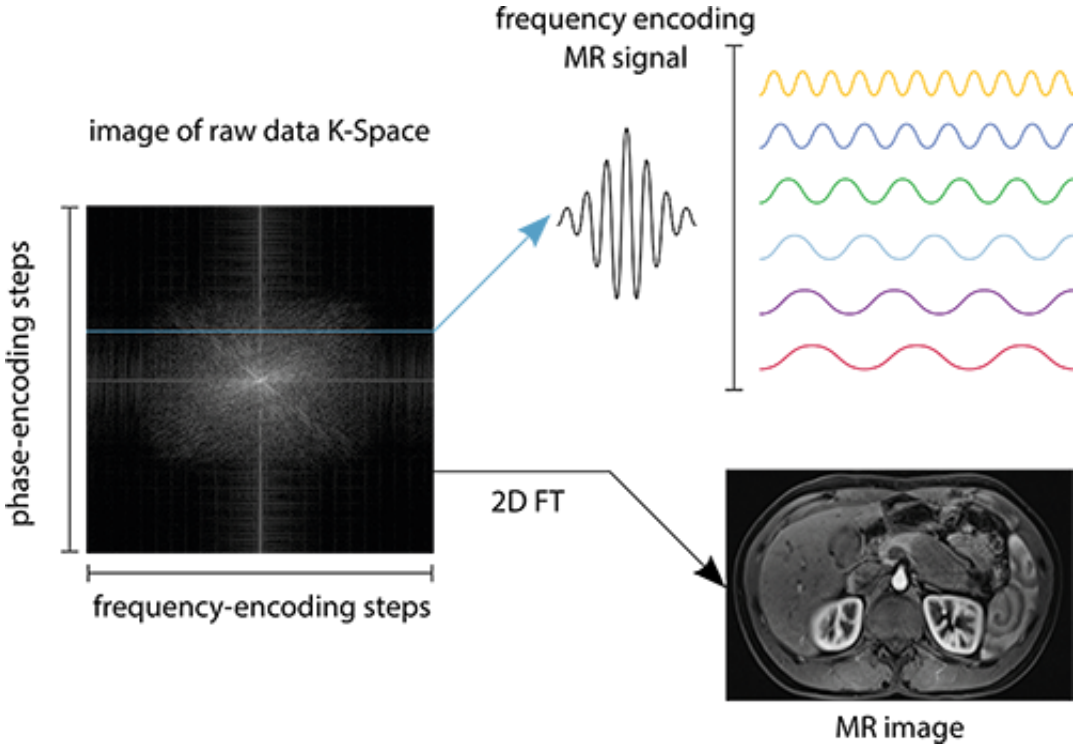


Figure 1.7: Illustration of k-space

When it is received, the MNR signal, encoded in phase and frequency, will be assigned to a value of the Fourier plane. Then during a spin-echo acquisition, each TR fills a line of the Fourier plane. For a linear course, it can start filling the central lines before filling the periphery of the space. It is also possible to traverse the space of k non-linearly, for example, following a spiral pattern starting from the center. k-space filling techniques can shorten the acquisition time [134].

1.2.3 Image quality and artifacts

Having obtained an image, we must now discuss image quality factors. Indeed, these notions are essential when setting up a machine learning system. For a system to be efficient, it must be able to provide it with good images, which requires defining what constitutes a good quality image.

Different MRI sequences present different functions of the object. The anatomical MRI sequences make it possible to obtain the topological structural information. In general, the T1 and T2 weighted are primary anatomical sequences. Besides, T1-weighted images are also acquired after gadolinium injection to enhance the contrast. The infusion of gadolinium is used as a contrast product. It alters the T1 and T2 times of the tissues in which the gadolinium spreads, which makes it possible to reveal the vascular system. Moreover, the fluid attenuated inversion recovery (FLAIR) sequence uses the principle of inversion recovery to suppress the signal from free water [49]. The sequences using recovery inversion completely switch the longitudinal magnetization by a 180° RF pulse. The FLAIR sequence suppresses free water, but other settings allow the fat signal to be suppressed.

After reconstructing an image from the signal, several factors influence digital image quality. These properties are important to have high-quality images for building an efficient system. Reconstructing an image from a physical signal requires spatial sampling. The 3D acquisition volume is then discretized into a finite number of voxels, which has notions of image resolution and contrast.

In general, the resolution of an image often refers to the equivalent of pixel count in digital imaging. However, in terms of medical images, resolution refers to the subjective visual quality of the image. A large image with objects on a few voxels usually has a good numerical resolution but poor qualitative resolution. Thus, it is actually desired that the voxels be as small as possible physically in order to best describe the variations of anatomical structures. On the other hand, increasing the spatial resolution impacts the quantity of signal present in each voxel. A small voxel might contain fewer protons than a significant voxel. In the case of ultra-high field MRI, it is possible to increase the resolution of the images. The initial external magnetic field B_0 is more robust, and the quantity of signal available increases, which in turn makes it possible to increase the resolution of the images [64].

In terms of image contrast, at a constant magnetic field, the amount of signal available for each voxel decreases as the size of the voxels decreases. For a given voxel, this decrease in the quantity of signal collected impacts the intensity of the measured signal. In general, the contrast of images refers to the intensity dynamics of the image. When the contrast is good, its direct effect is to make it possible to differentiate two different structures of similar intensities visually. The MRI contrast depends on many parameters, including the proton density, the T1 and T2 relaxation times of the tissues, the variations of the magnetic field, or the TE and TR acquisition parameters. For example, the image contrast depends on the combination of TE and TR values, with three combinations used: the short TR/short TE define the T1-weighted, the long TR/short TE define proton density (PD-weighted), and the long TR/long TE define the T2-weighted. The short TR/long TE combination is not used because it produces a poor contrast.

In addition, noise is also a property of digital image, especially medical images. On the metrological level, noise is parasitic information resulting from physical constraints. These disturbances are added to the expected theoretical signal and alter the resulting image. The signal-to-noise ratio (SNR) is used to evaluate the signal disturbance. At a constant level, if the signal is weak, SNR becomes high, and the visual quality of the image is degraded. Thus, one of the acquisition goals is to find the optimal parameters for reducing the signal-to-noise ratio while maintaining high radiological resolution.

The image artifact refers to features that appear in an image that is not present in the original image object. There are several types of artifacts in MR imaging. Each type consists of a characteristic signature in the image.

- Metallic artifact: This is a distortion of the magnetic field caused by the presence of ferromagnetic materials in the acquisition volume. This causes an absence of a signal at the heart of the disturbance and a non-linear deformation at the periphery of the disturbance. Some sequences specifically fight metal artifacts.

- Movement artifact: MRI acquisitions require the motionless as possible for the whole process. However, movements such as breathing heartbeat, are irreducible. Motions disrupt phase encoding more than frequency encoding since the phase encoding is done before reading the signal, while the frequency encoding is applied with the signal acquisition. Some motions can be corrected by an appropriate acquisition time of the actions. The motion artifact causes blurring in the affected area and the appearance of ghost images in the sense of phase encoding.

- Folding artifact: It appears when the studied object is larger than the field of view. We then observe a folding of the off-screen part of the contra-lateral side of the image. At a constant field of view, it is possible to make these artifacts disappear by changing the size of the voxels, to the detriment of the spatial resolution, or by applying an anti-aliasing filter.

- Chemical shift artifact: It might cause by variations in the resonance frequency encoding between different tissues. This concerns the interfaces between tissues, such as a transition between fat and water. Fat molecules have a lower precessing frequency compared to water. This leads to the presence of a black border on the water/fat interface. This artifact can be eliminated by increasing the bandwidth to the detriment of the signal-to-noise ratio.

- Truncation artifact: It occurs at the level of intensity between two different structures after sudden change in the acquisition, such as the transition between bone and fat. The truncation artifact produces alternating dark and light bands in the direction of the phase encoding.

- Magnetic susceptibility artifact: it appears at the interfaces of two juxtaposed structures with different magnetic sensitivities, such as the transition of hemoglobin/tissue or bone/tissue. At these interfaces, an intrinsic magnetic field gradient exists, which causes hypo-intense areas that are all the more marked as the voxel is large.

- The cross-excitation phenomenon: The excitation of protons on a slice is not limited to the edges of the slice itself. The RF pulse also affects adjacent protons. If the acquisition slices are stuck to each other, the protons find themselves excited several times, distorting the signal. Therefore, spacing the slices is necessary to avoid this cross excitation.

1.3 Ultra-high field MRI: the next generation

In recent years, clinical MR examinations are mainly conducted in static magnetic fields between 1.5 and 3T. However, several studies have also been performed at magnetic fields higher than 3T [216]. The ultra-high fields (UHF) refers to static magnetic field strengths of $B_0 \geq 7T$. It provides a significantly higher signal-to-noise (SNR) ratio, and several magnetic resonance applications benefit from higher contrast-to-noise ratios. Moreover, UHF can be utilized to resolve structures more precisely or to visualize physiological/pathophysiological impacts that would be challenging or impossible to detect at lower field strengths.

The first commercial 7T MR system for clinical imaging of the neuro- and musculoskeletal system was approved as a medical device in 2017. Since then, the number of UHF MRI devices has significantly expanded. It has allowed fundamental research and clinical diagnosis with MRI scanners. Besides, it helps to simplify doing research with more significant cohorts as part of clinical assessments with patient consent.

In general, the measurement noise in MR applications is estimated by the samples. The procession resonance frequency for hydrogen atoms is at least 64 MHz, which equates to a frequency of 1.5T, and can linearly increase. In the case of 7T and beyond, the frequency rises to at least 300 MHz. However, at UHF strengths and higher proton resonance frequencies, an even greater gain in SNR for hydrogen in MR applications has been recorded. Consequently, MRI at UHF delivers a considerable increase in MR signal compared to 1.5T and 3T, the usual field strengths in clinical MRI. This gain in signal with increasing magnetic field strength B_0 can be utilized, on the one hand, to achieve higher spatial resolution in the same measurement time or, on the other hand, to achieve comparable image quality in a shorter measurement time, which also permits higher temporal resolution in dynamic MRI techniques.

On the other hand, the contrast-to-noise ratio (CNR) also plays an essential role in MRI techniques. The contrast in MRI is generally determined by the interactions between nuclear spins and the changing environment. Different contrasts depend on various factors, such as T1 and T1-weighted, which alter differently as B_0 increases. When applications require complete relaxation before the subsequent stimulation, a longer T1 time causes a longer measurement time at a higher magnetic field. The higher field strength in spectroscopic applications results in a more significant splitting of resonant frequencies in the MR spectrum, which can be useful. In addition, it also enhances susceptibility sensitivity, which benefits susceptibility-weighted imaging and quantitative susceptibility mapping [111].

However, this can also lead to increased susceptibility artifacts, including geometric distortions and signal dropouts in the images. In DWI, higher SNR at the ultra-high field can provide the opportunity for higher-resolution imaging. However, it also brings challenges such as increased magnetic field and RF inhomogeneity of the transmission field, shorter T2 relaxation, and higher specific absorption rates. However, it can also result in susceptibility artifacts, including geometric distortions and signal dropouts in the images. In DWI, a higher SNR at an ultra-high field can enable imaging with a higher resolution. Nonetheless, it presents obstacles such as increased magnetic field and RF inhomogeneity of the transmission field, shorter T2 relaxation, and specific absorption rates [226].

As field strength and radio frequency increase, the wavelength for different body parts results in standing wave effects. In the case of UHF and higher, strong inhomogeneities in the transmit and receive fields can occur, leading to cancellations in the MR images, degraded contrasts, and regional peaks in the specific absorption rate distribution [111]. Consequently, coil designs at lower B_0 may display deleterious behavior at UHF for medium to large excitation volumes. Address the issues of inhomogeneous RF, advancements in excitation hardware and techniques such as parallel RF transmission (pTx) have been proposed to reduce the inhomogeneity of RF excitations and RF energy deposition by employing multiple-transmission RF coils that operate independently and concurrently [216].

Parallel transmission plays a fundamental role in UHF MRI, turning the potential applications of UHF into reality. It can optimize the magnetic field distribution or adjust the spin magnetization in the desired region by using the amplitudes and phases of the multiple channels as additional degrees of freedom. For pTx, all RF pulse shapes and gradient trajectories are typically optimized

for each transmits pulse element. In addition, magnetic field maps in 2D or 3D for each channel are necessary for determining pulse shapes and gradient trajectories. The latest 7T MR system enables static RF to be performed in a short amount of time without significantly delaying the clinical workflow, despite the fact that the required preparation steps were complex and time-consuming. In addition to inhomogeneities in the magnetic and transmission fields, cancellations and hot spots can occur in the local specific absorption rate distribution, which is dependent on the RF electric field pattern during safety evaluation [58].

Application

High-resolution MRI is still the most important clinical application in UHF topic. MRI at 7T field strengths provides higher SNR and, in several applications, higher CNR compared to MRI at lower field strengths, resulting in higher resolution and improved tissue differentiation. There are numerous important clinical applications for UHF MRI high-resolution imaging.

In recent years, the rapidly growing of ultra-high MRI devices hold the potential to improve clinical diagnostics. The high-resolution MRI improves the visualization of anatomical substructures [193]. For example, study in [63] has been demonstrated using a magnetization-prepared fast acquisition gradient echo, which could aid in diagnosing cranial nerve disease.

High-resolution morphological imaging has also been shown to be feasible for the complex anatomical structures of the brainstem, including nonquantitative techniques and quantitative approaches. In general, magnetic susceptibility MRI-based methods benefit significantly from strong magnetic fields. Post-processed phase images sensitive to magnetic susceptibility enhance the contrast between gray and white matter. Magnetic resonance angiography (MRA) can improve imaging of small intracranial vessels and has the potential to characterize vessel walls better [77]. MRI is the most sensitive imaging modality for detecting acute cerebral infarcts in patients with stroke [48]. In addition, cortical microinfarcts have been described using 7T magnetization transfer, limited to 3T [211]. In the detection of brain tumors, MRI is a cornerstone of diagnosis. Higher spatial resolution at 7 T could help distinguish infiltrating tumors from adjacent tissues better or reduce the administered contrast dose because of the higher contrast-to-noise ratio [166]. In patients with multiple sclerosis, better visualization of white matter lesions and visualization of a central vein and iron deposition within MS lesions have been described [108].

In terms of abdominal MRI, these promising results of high-quality ultra-high field imaging in neuroradiology form the theoretical basis for investigations in the abdomen. However, other physical effects associated with higher magnetic field strengths may impair diagnostic imaging. An increase in the magnetic field strength can provoke artifacts because of changes in tissue susceptibility, chemical shifts, or signal heterogeneities due to RF wavelength effects. Furthermore, the energy deposited by RF waves is restrictive at high-field imaging, as the RF absorption increases proportionally to the static magnetic field. Despite these limitations, 3T renal MRI has been nearing readiness for implementation into clinical standards.

Initial approaches to 7T whole-body MRI have highlighted the potential of UHF MRI and the need for further investigation in terms of RF technology and sequence optimization [215]. In addition, recent studies [110] have shown that 7T MRI showed partially comparable strength and drawback of MRI in the abdomen compared to lower field strengths, with significant differences in T1- and T2-weighted MRI. Higher SNR and CNR were observed in several abdominal organs, possibly allowing the detection of smaller pathologies that might be missed at lower field strengths.

Specially for kidney, the study in [210] has presented the feasibility of a dedicated 7T MRI of the kidneys utilizing a custom-built multiple transmit/receive RF body coil on a 7T whole-body MR system. 7T MRI of the kidneys is feasible, providing good overall image quality of the region of interest. T1-weighted imaging showed outstanding results regarding the differentiation of small anatomical structures and even allowed for a robust assessment of the renal vasculature without intravenous contrast medium administration, while T2-weighted MRI was limited at 7T because of artifacts and specific absorption rate restrictions. Furthermore, T1 MRI with fast gradient echo-based sequences allowed for a robust depiction of the renal vessels without requiring intravenous gadolinium administration. The benefit of intravenous gadolinium administration for renal MRI at 7T has not yet been evaluated. Contrast medium administration will be the focus of ongoing studies if dynamic T1-weighted imaging and perfusion analysis of renal tissue can be successfully implemented [210].

The initial imaging results demonstrated the successful transformation of the increased SNR into a high spatiotemporal resolution, yielding highly defined non-enhanced anatomical images while maintaining data acquisition within the window of a breath-hold with parallel imaging. Further optimization of RF technology and dedicated coil concepts can be expected to surmount better the physical effects linked to high magnetic field strength and enable the acquisition of even greater image quality with corresponding clinical diagnostic value.

In the future, further increases in field strength will enhance the possibilities in clinical research and will certainly lead to significant advances in these MRI techniques. The initial imaging results demonstrate success in converting the increased SNR into high spatial and temporal resolution, providing highly defined, unenhanced anatomical images while maintaining data acquisition within the window of a breath-hold with parallel transmission. 7T MRI systems are currently being installed. Systems beyond 7T are all geared towards basic research, either to study brain function or to understand healthy human physiology and ageing or the pathophysiology of various diseases. Given the challenges involved, field strengths greater than 7T are not currently being considered for diagnosis in individual patients. It is expected that further optimisation of RF technology and special coil designs will better overcome the physical effects associated with high magnetic field strength and enable the acquisition of even better image quality with corresponding clinical diagnostic value. With regard to renal imaging, it is possible to investigate the advantages of dynamic imaging and perfusion analysis in 7T MRI overall.

1.4 Computer-aided MRI analysis: the augmented pathologist

Machine learning in general and deep learning, in particular, have recently received enormous attention. Deep neural networks have outperformed other established models on several important benchmarks. Deep Learning methods are now state-of-the-art machine learning models in various domains, from image analysis to natural language processing, and are widely used in academia and industry. These developments hold enormous potential for medical imaging technology, data analytics, diagnostics, and healthcare in general, which is slowly being realized.

Healthcare providers generate and hold enormous amounts of data containing extremely valuable signals and information at a pace far surpassing what traditional analysis methods can process. Deep learning in medical data analysis is here to stay as a new research topic to integrate, analyze and make predictions based on large, heterogeneous datasets. Applications of deep learning can

range from analyze and the prediction of medical events, e.g. seizures [15] and cardiac arrests [215], to computer-aided detection [155] and diagnosis [24], supporting clinical decision making and survival analysis [43].

Deep learning methods are increasingly used to improve clinical practice, and the list of examples is long and growing daily. Even though there are many challenges associated with introducing deep learning in clinical settings, the methods produce results that are too valuable to discard. Beyond the application of machine learning in medical imaging, the attention in the medical community can also be leveraged to strengthen the general computational mindset among medical researchers and practitioners, mainstreaming the field of computational medicine.

Deep neural networks such as convolutional neural networks can be used for efficiency improvement in radiology practices through protocol determination based on classification [120]. They can also be used to reduce the gadolinium dose in contrast-enhanced brain MRI by order of magnitude [66] without significant reduction in image quality. Deep learning is applied in radio-therapy [141], in PET-MRI correction [130, 6] and for theranostics in neurosurgical imaging, combining confocal laser endomicroscopy with deep learning models for automatic detection of intraoperative CLE images on-the-fly [86]. Another important application area is advanced deformable image registration, enabling quantitative analysis across different physical imaging modalities and time. For example, fast deformable image registration of brain MR image pairs by patch-wise prediction of the large deformation diffeomorphic metric mapping model [232]; deep learning-based 2D/3D registration frame-work for registration of preoperative 3D data and intraoperative 2D X-ray images in image-guided therapy [248]. Neural network-based methods rely heavily on the support of big data.

There are several thorough reviews and overviews of the field to consult for more information, across modalities and organs, and with different points of view and levels of technical details. For example, the comprehensive review [41] covers both medicine and biology and spanning from imaging applications in healthcare; to key concepts of deep learning for radiologists [119, 33, 139, 203], deep learning in neuroimaging and neuroradiology [237]; brain segmentation [2]; and more technical surveys of deep learning in medical image analysis [184, 199, 32].

Computer-aided diagnosis (CAD) refers to systems that can detect, mark, and assess potential pathologies for radiologists to help improve identification accuracy in the case of data overload and human resource limitation. The analysis, quantification, and categorization of images with these methods is an important technique that can improve patient safety and care. There are many advantages to using machine learning techniques in CAD systems. The first advantage of machine learning is its accurate and robust performance in many radiology studies. Moreover, CAD systems are expected to perform consistently and produce robust results with large amounts of data at any time and space. In contrast, manual diagnosis results may be affected by fatigue, reading time, and emotion on the part of the practitioner. The second advantage is that the diagnosis can be finalized in a brief time. Radiological analyses might be complex and require experienced radiologists, while a learning-based model only takes a few seconds to analyze results from an image. With the help of a CAD system, radiologists can have the support of automatic systems to speed up the diagnosis process less cost-effectively.

Although the performance of computer-aided diagnosis systems is improved every day to tackle the most common clinical problems, current contributions need further investigations before being widely applied in practice. First, most current diagnosis contributions mainly focus on predicting one type of disease, which may not meet the clinical demands. There may be one or more diseases existing in one radiological image. Besides, the current model training is mainly based on one

type of measurement. However, most disease decisions in clinical practice rely on multiple domain measurements. Information from multi-measurement may increase model accuracies. Moreover, current medical datasets mainly cover common diseases. Only a limited number of rare diseases are exposed to human clinicians, and many contributions may not consider these individual cases during their model training. More comprehensive systems that can detect various diseases and report rare cases are expected to be seen in the future.

1.5 Conclusions

In this chapter, we have presented basic concepts of kidney anatomy, functions, and diseases. As a central part of the urinal system, kidneys are responsible for filtrating blood from waste and extra water, preserving the inner body equilibrium, and maintaining the acid-base balanced, synthesizing vitamins and hormones. The gradual loss of renal function defines chronic kidney disease. It is a long-term condition where the kidneys cannot work as they should. GFR metrics with invasive methods are used to define kidney diseases, which mention the end-point of the kidney where either renal dialysis or transplant.

Accurate assessment of renal function and structure non-invasively is important in diagnosing and predicting kidney diseases. Non-invasive methods using medical images such as MRI play an increasingly critical role in assessing renal function. Moreover, advanced techniques of MRI recently developed, such as ultra-high field MRI, which allow images able to provide more structural, functional, and molecular details that can detect the alteration in renal tissue properties and functionality and help predict and diagnose renal function.

In the MRI field, deep learning holds a huge potential at each step of entire workflows, from acquisition to image retrieval, segmentation to disease prediction. The following chapters provide fundamental knowledge of deep learning and neural networks, as well as state-of-the-art for several tasks of MRI analysis that the thesis targets on.

Chapter 2

Deep learning and Neural network

Statistical learning or machine learning is a scientific research domain combining applied mathematics, statistics and computer science. It is essentially a form of applied statistics with increased emphasis on using computers to statistically estimate complicated functions and a decreased emphasis on proving confidence intervals around these functions.

Deep learning is an exciting sub-field of machine learning based on artificial neural networks and representation learning. It uses lots of data to teach computers how to do tasks that only humans were capable of before, such as recognizing an image, translating, etc. Types of learning can be supervised, semi-supervised or unsupervised. Deep learning has emerged as a central tool for solving perceptual problems and has become state-of-the-art in many fields. These include computer vision, speech recognition, natural language processing, medical image analysis, climate science, etc., where algorithms have produced results comparable to and, in some cases surpassing those of human experts.

The following sections present the fundamental principles and terminology of deep learning, neural networks, and network training. Usage of multiple network layers is referred to as the term "deep". Deep learning is a modern variant that works with several layers of limited size, enabling practical application and optimal implementation while preserving theoretical universality under moderate conditions. In deep learning, layers may also be diverse and strongly vary from biologically informed connectionist models for efficiency, trainability, and interoperability.

2.1 Machine learning overview

The general objective of machine learning algorithms is to solve tasks by learning from sets of samples. Unlike traditional computer programming, models automatically extracts significant patterns from data and completes a task without being explicitly programmed. In this context, tasks are presented in terms of how the system processes a sample to generate output. A sample is a collection of characteristics that have been objectively measured from an object or event. Machine learning can be used to handle a variety of problems, including classification, regression, transcription, machine translation, synthesis, and sampling. At this moment, machine learning has been applied in various domains ranging from computer vision, natural language processing, speech processing, signal processing, robotics, healthcare, biology, manufacturing, economics, advertising, etc.

Machine learning algorithms can be separated into supervised learning and unsupervised learning by arranging samples during the learning process. The following paragraphs provide an overview of supervised and unsupervised learning, along with basic concepts of machine learning that will be used in the thesis.

2.1.1 Supervised learning

Supervised learning algorithms refer to a system that learns from each sample feature associated with a label. In other words, it aims at predicting an outcome variable by giving a set of descriptive variables that are assumed to influence the outcome. Based on the type of observation, supervised learning include two types: regression for the continuous outcomes; and classification for discrete outcomes.

Theoretically, supervised algorithms learn to seek a hypothesis function that describes the relationship between input and output space. A probability from a joint distribution is applied on input and output to reduce the dependence of variables, or random noise [68]. The loss functions measure the accuracy of the hypothesis. The goal is to minimizes the risk of generalization error through a learning algorithm or learning rule. Learning rules aim at selecting the function that minimizes the average loss on the observed data [214]. The empirical risk is also commonly known as training error. A learning system objective is to generalize well to unobserved examples. A low training error is not a guarantee of a low generalization error. A predictor often performs well on the training samples, but might fails to generalize over test samples.

Model complexity is another term for the expressiveness of the learning process. Other learning rules that assign different weights to sets of hypotheses are structural risk minimization and minimum description length. There are several supervised learning from statistical methods such as random forest, K-nearest neighbours, decision tree and support vector machines to deep learning methods.

K-Nearest Neighbors (KNN) [4] is a non-probabilistic supervised learning methods. In general, KNN can be used for both classification and regression. KNN works based on the majority vote between similar instances. The similarity is defined using distance metrics between data points. When the model predicts output for a new data point, it finds the nearest neighbours to that point in the training data by returning the average of the corresponding values in the training set. As a non-parametric learning algorithm, KNNs are not restricted to a fixed number of parameters. There is not even really a training stage or learning process. Moreover, KNN is a simple and powerful method with any distance measure. It can make a model obtain high accuracy given a large training set. However, it may generalize very badly given a small finite training set.

Decision Tree [29] is another supervised learning algorithm that is also based on separating input space into regions with separate parameters. Nodes are associated with regions, while internal nodes break that region into a small non-overlapping region for each child of the node, with leaf nodes and input regions having a one-to-one correspondence. However, similar to KNNs, the performance of a single decision tree is limited to unseen data. Thus, Random Forest (RF) [28] is created to increase the ability to generalization of the model. It involves several decision trees from training data, then combined to form the final output.

Support Vector Machine is one of the most robust supervised learning algorithms that analyses data for classification and regression. The idea of SVM is to find hyperplanes that

classify data points into classes. One key innovation associated with SVM is the kernel trick to learn non-linear models as a function using convex optimization techniques that are guaranteed to converge efficiently.

2.1.2 Unsupervised learning

In many cases, labelled data are not readily available because labeling the training data is costly, time-consuming, and impractical on a large scale. Additionally, it is frequently less concerned with predicting a specific target variable than with comprehending the data. Unsupervised learning algorithms experience a dataset containing many features and then learn valuable properties of the structure of this dataset. As opposed to predicting a target variable, unsupervised learning involves observing random vectors and attempting to learn the probability distribution or interesting properties.

Unsupervised learning generally refers to the majority of attempts to extract information from a distribution that does not require an annotation procedure. The term is usually associated with learning to draw samples from a distribution, denoise data from some distribution, density estimation, find a manifold that the data lies near, or cluster the data into groups of related examples. In deep learning, the objective is to learn the entire probability distribution over the dataset, either directly, as in density estimation, or implicitly, for tasks such as synthesis or denoising.

A classic unsupervised learning task is to find the best representation of the data. Lower-dimensional representations, sparse representations, and independent representations are three of the most common methods. Low-dimensional representations aim to condense as much input data as possible into a smaller representation.

Sparse representations [14] encapsulate the dataset in a representation in which the majority of entries are zeros. Sparse representations often require an increase in the dimensionality of the representation, so that the loss of information caused by the representation consisting primarily of zeros is minimized. This leads in an overall structure that tends to distribute data along the axes of the representation space. Independent representations aim to separate the sources of variation underlying the data distribution so that the representation dimensions are statistically independent.

Dimensionality reduction methods also can be considered as an unsupervised learning algorithm that learns a representation of data [68]. It consists of reducing the dimension, or the number of variables to improve a downstream task. Reducing the dimension has several benefits: reducing the computational burden by compressing features, improving the behaviour of learning algorithms by finding useful variables, or enabling direct visualization and interpretation. These techniques can be divided into feature extraction and feature selection. Approaches identify a suitable subset of the original variables to represent the data. In contrast, feature extraction constructs new variables that include a substantial amount of global information.

Dimensionality reduction is a problem involving data approximation in high-dimensional vector spaces. The simplest method is to reduce all data to a single representative value, such as the mean or standard deviation. Then, data can be approximated more finely using projections on hyperplanes. These methods are called linear dimensionality reduction with principal component analysis (PCA) as a typical representative. This technique is based on linear algebra to learn the orthogonal projection of data with lower dimensionality than the original input. Other linear

methods can be mentioned as non-negative matrix factorization, random projections, compressed sensing, and multi-dimensional scaling. Non-linear dimension reduction [118] or manifold learning are techniques to approximate data by learning low-dimensional embeddings to maintains global and local structures of the data. Such methods include locally linear embedding [171], isomap [201], spectral embedding [180], laplacian eigenmap [16], kernel PCA [182].

Neural networks are another important aspect of nonlinear dimension reduction. Deep learning [116] aims to learn effective representations using multi-layer neural networks, whether it is supervised or unsupervised. In the next chapter, concepts of neural networks will be presented.

Clustering is one of the most common tasks in unsupervised learning. It consists in finding meaningful groups of individuals in an unlabeled data set. In exploratory data analysis, this provides insight into the structure of a dataset and can also be used for classification, where data points are coded with a clustered index. An object can be described by a set of features or by its relationship to other objects, such as a distance or affinity matrix. Popular clustering methods range from simple methods such as k-means clustering and mean-shift clustering to complex methods such as Gaussian mixture models or hierarchical clustering. However, clustering is outside the scope of this work and will not be used throughout the thesis, so we will not present them in detail.

2.1.3 Other concepts

2.1.3.1 Training, testing, validation sets

Machine learning algorithms usually consist of many hyperparameters and settings that we can use to control the behavior of the algorithm. The learning algorithm does not adapt to the values of hyperparameters. Sometimes a setting is chosen as a hyperparameter that the learning algorithm does not learn because the setting is difficult to optimize. More often, the setting must be a hyperparameter because it does not make sense to learn that hyperparameter on the training set. This is true for all hyperparameters that control model capacity. When learning on the training set, such hyperparameters would always choose the maximum possible model capacity, which would lead to overfitting.

A data set is usually divided into training, testing, and validation sets. While the role of the training and testing sets is completely significant, the validation set aim to support the tranining process to avoid overfitting or underfitting problem. The validation set consists of examples also from the same distribution as the training set and can be used to estimate the generalization error of the model after the learning process is complete.

The test examples may not be used to make decisions about the model, including its hyperparameters. Therefore, no example from the test set can be used in the validation set. The training set is used to learn the parameters, while the validation set is used to estimate the generalization error during or after training and update the hyperparameters accordingly. Since the validation set is used to train the hyperparameters, the error of the validation set underestimates the generalization error, but usually by a smaller amount than the training error. After the optimization of the hyperparameters is complete, the generalization error can be estimated using the test set.

2.1.3.2 Overfitting and underfitting

The key challenge in machine learning is that algorithms must also perform well on new, previously unknown inputs, not just those on which our model has been trained. Generalization is the capacity to perform well on previously unobserved inputs. During the training phase, typically only the training set is accessed. The training error, is computed and reduced during measurement on the training set. So far, this has been considered simply as an optimization problem. Machine learning differs from optimization in that we seek to minimize the generalization error. The generalization error defines the expected value of the error given on new inputs. Here the expectation is taken across different possible inputs, drawn from the distribution of inputs we expect the system to encounter in practice.The generalization error of a machine learning model is estimated by measuring its performance on a separate test set than the training set.

One immediate connection that can be observed between the training error and the test error is the standard training error of a randomly selected model equal to the expected test error of that model. The factors determine how successfully an machine learning algorithm will be able to reduce the training error and maintain the smallest gap between the training error and the test error. These two factors remain to the two primary challenges in machine learning: underfitting and overfitting.

Underfitting occurs when the model cannot obtain a suitably low error value on the training set. The model cannot generalize over the dataset and then provide low performance.

Overfitting occurs when the difference between training error and test error is excessively high. In practice, it is observed when a model achieves high accuracy on the training set but low performance on the test set. It can happen when the model is too complex or the unbalance of samples between classes in the dataset. Figure 2.1 illustrates examples of overfitting, underfitting and usual models.

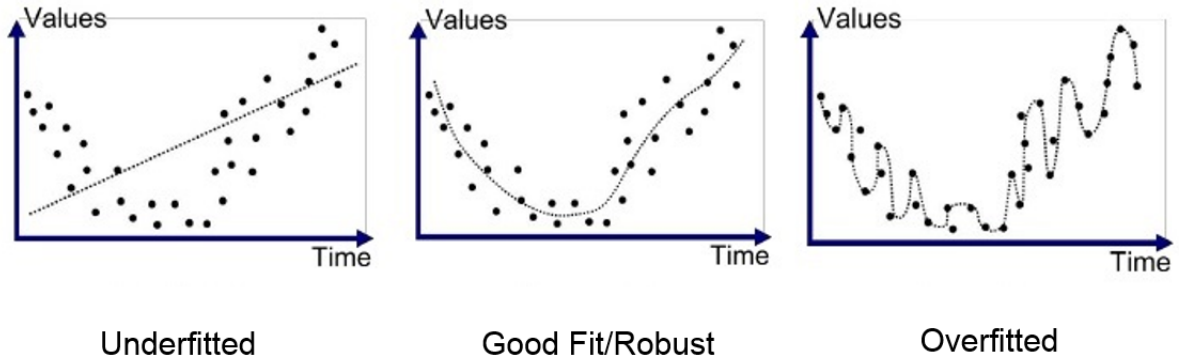


Figure 2.1: An example of overfitting, underfitting and usual performance.

It is possible to control whether a model is likely underfit or overfit by controling its capacity. Capacity of a model is its adaptability to fit a wide variety of functions. Low-capacity models may have difficulty to fit the training set. High-capacity models may overfit by memorizing properties of the training set that are ineffective on the test set. Choosing the hypothesis space and the set of functions is refered as an effective solution for regulating the capacity of a learning algorithm. For example, the linear regression algorithm has the set of all linear functions of its input as its hypothesis space.

2.1.3.3 Cross-Validation

Dividing the data set into training and test set only once can become a problem without considering of ratio between these sets. A small test set implies statistical uncertainty around the estimated average test error, making it challenging to assert that algorithm, particularly when the dataset is small. In this situation, alternative training procedures based on repeating the training and testing computations on randomly selected subsets or splits of the original dataset are typically used. The k-fold cross-validation procedure is the most common technique, in which a partition of the data set is formed by splitting it into non-overlapping subsets. The test error is estimated by finding the average test error over a few steps. A subset of the data is used as the validation set, while the rest of the data is used as the training set. In this way, the model can take all the examples to estimate the average test error, but at the cost of more computational cost.

Cross-validation is a very efficient technique to overcome the limitations of training on small datasets. It can improve the performance of the model in terms of representation and generalization. On the other hand, cross-validation still faces the problem that there are no unbiased estimators for the variance of such estimators of the mean error, so it is usually approximated [20]. Figure 2.2 demonstrates the training phase using k-fold cross validation.

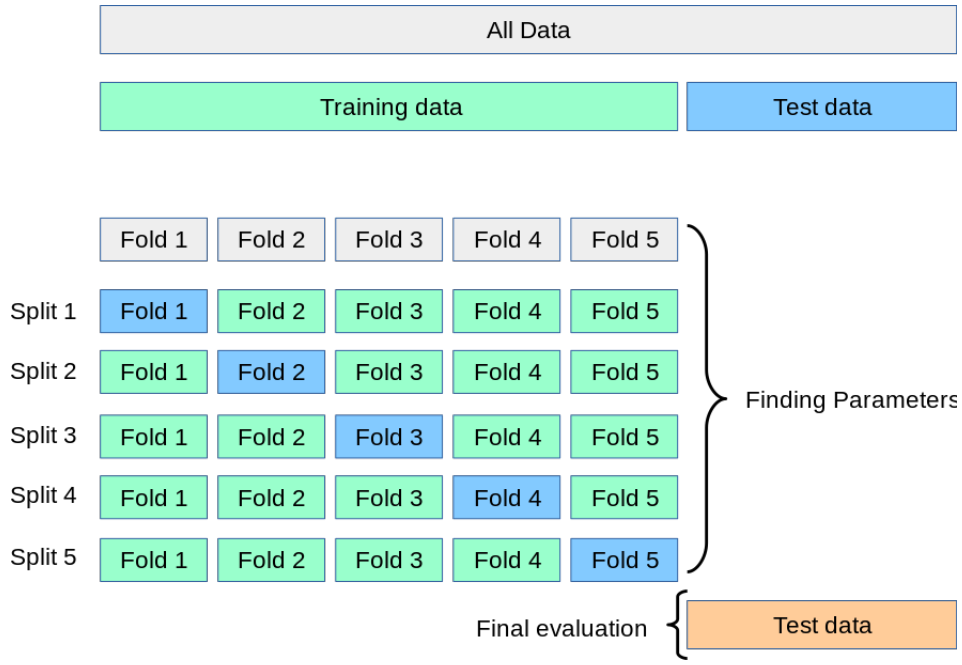


Figure 2.2: Cross validation.

2.2 Deep Neural Network

Artificial neural networks, usually called neural networks, are computing systems based on a collection of connected units or nodes called neurons, which loosely model the neurons in a biological brain. Like the synapses in a biological brain, each artificial neuron in the network can receive a signal, and then transmit it to neurons connected. In this case, the signal at a connection is a real number, and the output of each neuron is computed using a nonlinear function of the sum of its inputs. The interconnections are known as edges. Typically, each neuron and synapse contains a

weight to learn during training procedure. The value of weights reflects the signal strength at a connection. Nodes may have a threshold such that a signal is only transmitted if the aggregate signal exceeds it. Neurons are typically arranged into layers. Different layers can apply different transformations to inputs. Signals transmits from the input layer to the output layer, traversing the layers multiple times if necessary.

The term deep neural networks in deep learning refer to a computing system that contains a series of stacked layers. Each layer is composed of different types of components from basic units such as nodes to complex ones such as operators or blocks. A layer is connected to other layers through a set of weights. The computation happens in each layer, where it combines input with a set of parameters or weights that either amplify or dampen that input. The weight products are then summed, and the sum is passed through the activation function to determine to what extent the value should progress through the network to affect the final prediction, which depends on task purposes.

2.2.1 Very first neural network

Feedforward networks, or multilayer perceptrons, are the classic deep learning model. They are the most basic and extreme importance in the field of machine learning in general or deep learning in particular. The goal of this network is to learn the approximation of a function that maps an input to a category. The model is called feedforward because information flows through the function being evaluated from input, through the intermediate computations used to define mapping function, and finally to the output. There are no feedback connections in which model outputs are fed back into themselves. They form the basis or are a part of many critical neural networks, such as the convolutional networks for object recognition; or the conceptual stepping stone on the path to recurrent networks

In a feedforward network, units are typically connected together by many different functions, while the number of layers defines the depth of the network. The final layer of a feedforward network is called the output layer. The behavior of the other layers is not directly specified by the training data. Because the training data does not show the desired output for each of these layers, they are referred as hidden layers. Typically, each hidden layer is vector-valued. The size of hidden layers determines the network width. Because each unit receives input from numerous other units and computes its activation value, it resembles a neuron.

2.2.2 Neural network training

Every neural network has to be trained. The goal of the training network is to find the best configuration of parameters that can produce results with the highest accuracy possible. In the beginning, a neural network may start with the random initialization of weights. That means all neurons in a given layer produce an output, and they have different weights for the next neural layer. The values of weights are adjusted to fit with data during the training phase. A significant weight means that the input is important, while a small weight means that it has less impact. The goal is to find another value of parameters that performs better than the initial one. In training, using layers to produce the output is called the forward pass. However, in neural network design, the weights of units cannot be updated by themselves because there are no feedback connections where outputs of the model are fed back into itself. Hence, once the output has been calculated, the system will re-propagate the evaluation error using the back-propagation algorithm [174]. The

system will adjust the weights of the different inputs into each neuron with a given step which is called the learning rate.

To evaluate the training phase, the objective is to minimize the error between the predicted output made by the neural network and the actual data. This difference is represented by a function called the loss function, cost function, or error function. The process to minimize or maximize a function is called optimization. Gradient Descent (GD) is an iterative optimization algorithm for finding the local minimum of a function. To find the local minimum of a function using gradient descent, we must take steps proportional to the negative of the gradient that moves away from the gradient of the function at the current point. Besides, there are a lot of algorithms that optimize functions. These algorithms can be gradient-based or not, in the sense that they are not only using the information provided by the function but also by its gradient. The details of neural network components and training concepts are presented in the following sections.

2.3 Concepts and terminology

2.3.1 Activation function

An activation function decides whether a neuron should be activated or not. It means that it will decide whether the neuron input to the network is important or not in the process of prediction using simpler mathematical operations. The primary function of the activation function is to transform the summed weighted input from the node into an output value to be fed to the next hidden layer or as output.

In general, each neuron performs a linear transformation on the input using weights and biases without considering the number of hidden layers attached to the network. If there is no activation function, all layers will conduct in the same way because the composition of two linear functions is a linear function itself. The purpose of an activation function is to add non-linearity to the neural network.

Different neural networks use different non-linear activation functions such as sigmoid, tanh, or rectified linear unit (ReLU) activation function, etc. Among that, ReLU and its variations have become very popular in the last few years. The formula of ReLU is defined as:

$$f(x) = \max(0, x) \tag{2.1}$$

where x is the output from a hidden unit.

Although it gives the impression of a linear function, ReLU has a derivative function and allows for back-propagation while simultaneously making it computationally efficient. The main catch here is that the ReLU function does not activate all the neurons simultaneously. The neurons will be activated if the output of the linear transformation exceeds 0. However, units after ReLU can be fragile during the back-propagation process, called the dying ReLU problem, when a ReLU neuron causes the weights to update so that the neuron will never activate on any data point again. The gradient flowing through the unit will always equal zero from that point on. Therefore, ReLU variants such as leaky ReLU, parametric ReLU, and exponential linear unit (ELU) are used to handle the slope of the negative part in the dying ReLU issue.

The sigmoid activation function is also very popular in neural networks. It produces an output

within the range [0, 1], which is a probability. The sigmoid function is defined as:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{2.2}$$

where x is the output from a hidden unit.

In fact, it is commonly used for models to predict the probability of an output. Since the probability of anything exists only between the range [0, 1], the sigmoid solves it well because of its range. However, when the output requires probabilities for multiple classes, the sigmoid cannot return the correct values. Thus, the softmax function is used to calculate the relative probabilities that return each class's probability. It is most commonly used as an activation function for the last layer of the neural network in the case of multi-class classification. The softmax function is defined as:

$$f_i(x) = \frac{e^{x_i}}{\sum_{j=1}^{K} e^{x_j}} \tag{2.3}$$

where x_i are the elements of the input vector with size K to the softmax function, e^{x_i} is standard exponential function applied to each element of the input vector and $\sum_{j=1}^{K} e^{x_j}$ are the sum of of all exponential function of all input. It ensures that value of output will be in range [0, 1] and sum equal to 1, thus constituting a valid probability distribution

2.3.2 Loss function

In machine learning, a loss function presents the cost paid for the difference between the expected and predicted outcomes produced by the model. The training of neural networks is similar to an optimization problem that seeks to minimize a loss function.

A neural network attempts to learn the probability distribution underlying the given data observations. In machine learning, the common term of the statistical framework of maximum likelihood estimation is used as a basis for model construction. It means it tries to find a set of parameters and a prior probability distribution, such as the normal distribution, to construct the model representing the distribution over our data.

Along with the development of deep learning, there are many different loss functions. From single loss functions such as MSE, cross-entropy, L1 and L2 Regularization, etc., in classic neural networks such as MLP to composed loss functions built from basic functions such as adversarial loss, perceptual loss, or cycle-consistent loss seen in recent neural networks.

2.3.3 Back-propagation

As introduced in the previous part, back-propagation [174] is an algorithm that is used to train neural networks by adjusting weights and biases in the network so that they cause the predicted output to be closer to the expected output, thereby minimizing the cost function. The gradients of the cost function determine the level of adjustment with respect to those parameters.

To train a feed neural network with backpropagation, there are three main components:

- A dataset $\mathbf{X}$ consisting of N pairs of input and output $(\mathbf{x}_n, y_n)$ denoted as: $\mathbf{X} = \{(\mathbf{x}_0, y_0), ..., (\mathbf{x}_N, y_N)\}$ where y_n is desired output of the network on input $\mathbf{x}_n$
- A feedforward neural network consisting of m layers l with θ parameters. The parameters of primary interest are the weight w_{ij}^k of node j in layer l_k and node i in layer l_{k-1}; and bias b_i^k for node i in layer l_k
- The loss function, $E(\mathbf{X}, \theta)$, which defines the error between the desired output y and the calculated output $\hat{y}$ of the neural network on input $\mathbf{x}$ for set of pairs and value of the parameters θ.

Algorithm 1 Classic back-propagation

Require: α, a fixed learning rate
Require: random initialization of the parameters w_{ij}^k
a_i^k, product sum plus bias for node i in layer l_k
o_i^k, output for node i in layer l_k
δ_i^k, the error value for node i in layer l_k
r^k, number of nodes in layer l_k
for n=0 to N **do**
Forward phase
$(a_i^k, o_i^k) \leftarrow feedforwards(x_n, w_{ij}^k)$
Backward phase
1. Evaluate the error term for the final layer: $\delta_1^m \leftarrow a_1^m(\hat{y_n} - y_n)$
2. Backpropagate the error terms for the hidden layers at layer $k = m - 1$: $\delta_j^k \leftarrow a_j^k \sum_{l=1}^{r^{k+1}} w_{il}^{k+1} \delta_1^{k+1}$
3. Evaluate the partial derivatives of the individual error E_n with correspond to w_{ij}^k: $\frac{\partial E_n}{\partial w_{ij}^k} \leftarrow \delta_j^k o_i^{k-1}$

Combine gradients
Total gradient: $\frac{\partial E(X,\theta)}{\partial w_{ij}^k} = \frac{1}{N} \sum_{n=1}^{N} \frac{\partial E_n}{\partial w_{ij}^k}$
Update weights
According to the learning rate α and total gradient $\frac{\partial E(X,\theta)}{\partial w_{ij}^k}$:
$\Delta w_{ij}^k = -\alpha \frac{\partial E(X,\theta)}{\partial w_{ij}^k}$
end for

The back-propagation algorithm is formally described in algorithm subsection 2.3.3. When a feedforward neural network uses an input x and produces an output $\hat{y}$, information flows forward, then propagates up to the hidden units at each layer, and finally generates $\hat{y}$. This is denoted as the feedforward in algorithm subsection 2.3.3.

The back-propagation allows the information produced from forward flow to backward through the network to compute the gradient using a simple and inexpensive procedure. Computing derivatives by propagating information through a network in back-propagation can balance the trade-off between the difficulty of analytically finding the derivative for each neural network architecture and the computation cost of approximating the derivative. In general, back-propagation refers to the method that can compute the gradient of any function. Optimization algorithms in neural network training, such as stochastic gradient descent, are used to perform learning using this gradient.

2.3.4 Optimization

Deep learning approaches require optimization in several contexts, and neural network training is known as the most common and challenging problem. Network training aims to find set of parameters that significantly minimize the loss function, reflecting the difference between model output and desired output. When the loss functions are reduced to an acceptable value, models can learn the mapping function.

As previously mentioned, training neural networks is generally expensive in resources and computational time. For its resolution, a set of specific optimization techniques has been developed. There are several choices for optimizers, where each optimizer features tunable parameters like learning rate, momentum, and decay. The most commonly used optimizers are stochastic gradient descent (SGD), Adaptive Moments (Adam), and Root Mean Squared Propagation (RMSprop). Adam and RMSprop are variations of SGD with adaptive learning rates. Adam is used in the proposed classifier network since it has the highest test accuracy.

Stochastic Gradient Descent (SGD) ,and its variants are considered the most fundamental optimizer for deep learning. It is a faster version of the gradient descent (GD) in calculus. In gradient descent, tracing the curve of a function downhill finds the minimum value, much like walking downhill in a valley until the bottom is reached. By comparison, the actual gradient of the total cost function becomes small and then, when approached and reaches a minimum using the batch gradient called batch gradient descent (BGD).

Deep learning models crave data. The more the data, the more chances of a model being good. In BGD, all the examples are considered for every step of gradient descent. However, scaling BGD is ineffective when training data is huge due to a large amount of computation. If computing the loss function takes several floating-point operations, computing its gradient takes about three times to compute. To tackle this problem, SGD is a better solution to optimize training. In SGD, only one set is considered at a time to take a single step. A set is fed into the network first to calculate the gradient; then, the output is used to update weights. The process is repeated for all the examples in the training dataset.

Adaptive gradient (AdaGrad) [55] is a simple modification of SGD, which implicitly does momentum and learning rate decay by itself. Using AdaGrad often makes learning less sensitive to hyper-parameters. However, it often tends to be worse than precisely tuned SDG with momentum. AdaGrad adjusts the learning rate of all models individually by scaling them inversely proportional to the sum of all the gradient squared previous values. While parameters with small partial derivatives experience a relatively moderate fall in the learning rate, those with the most significant partial derivatives of the loss experience a similarly substantial decrease in learning rate. As a result, more significant progress is made in the parameter space in more gently sloping directions. AdaGrad is advantageous in convex optimization and has specific good theoretical characteristics. However, empirically, the growth of squared gradients from the start of training can lead to an early and disproportionate drop in the effective learning rate. AdaGrad operates well for some deep learning models, but not all in general.

Root mean square propagation (RMSProp) [79] is a modification of AdaGrad to perform better in the nonconvex setting by changing the gradient accumulation into an exponentially weighted moving average. AdaGrad is designed to converge rapidly when applied to a convex function. When applied to a nonconvex function to train a neural network, the learning trajectory may pass through many different structures and eventually arrive at a locally convex bowl region. Before reaching the convex structure, AdaGrad may have made the small learning rate by shrinking

it in accordance with the entire history of the squared gradient. On the other hand, RMSProp behaves as if it were an instance of the AdaGrad algorithm initiated within the identified convex bow by discarding history from the distant past using an exponentially decaying average.

Adaptive moment estimation (Adam) [102] is another adaptive learning rate optimization algorithm. It uses momentum [160] and adaptive learning rates to make training converge faster. When learning rates are reduced to a pre-defined schedule throughout the training phase, adaptive learning rates are thought of as modifications to the learning rate, while momentum is a method that accelerates SGD in the relevant direction. First, in Adam, momentum is incorporated directly as an estimate of the first-order moment of the gradient. Then, Adam includes bias corrections to the estimates of both the first-order and second-order moments to account for their initialization of the origin.

Currently, SGD with momentum, RMSProp, RMSProp with momentum, AdaDelta, and Adam are the most widely used optimization algorithms. It appears that the researchers expertise with the algorithm will significantly impact the algorithm to be used.

2.3.5 Learning rate

The learning rate is a hyperparameter that determines how much the model is controled in response to the estimated error whenever the model weights are updated. During the network training, optimizers provide the steps in the current direction of the slope, while the learning rate gives the length of each step that it take. The learning rate helps the network to abandon old beliefs for new ones.

Defining the learning rate is very important for network training because a too small value can lead to a long training process that could get stuck, while a too large value can lead to a suboptimal set of weights being learned too quickly or an unstable training process. Learning rate is a very critical hyperparameter in neural network configuration. Therefore, it is necessary to understand how to study the effects of the learning rate on model performance and to develop adaptive learning rate on model behavior.

2.3.6 Batch Normalization

Ensuring that the weights of the network stay within a reasonable range of values are a common issue when training a deep neural network. The network suffers from the exploding gradient problem if they become too large. As errors are propagated backwards through the network, the calculation of the gradient in the early stages can occasionally become exponentially large, leading to dramatic fluctuations in the weight values. When weights have grown large enough to cause an overflow error, it leads to vanishing gradients.

In general, it does not appear at the beginning of network training. However, it can happen in the middle of training or even after a few epochs when suddenly, the loss function fails to return a finite value, and the network has exploded. It can be incredibly irritating, especially if the network has appeared to be performing well for a while. It causes by the scaling of input data into a neural network which aims to guarantee a stable beginning to training over the initial few iterations. Since the network weights are initially randomized, unscaled input could potentially create huge activation values that immediately lead to exploding gradients. Because the input is scaled, it is natural to expect the activations from all future layers to be relatively well-scaled. It

may be true initially, but as the network trains and the weights deviate from their initial random values, this assumption can break down. This phenomenon is known as covariate shift [84].

To remain stable, each layer implicitly assumes that the distribution of its input from the layer beneath is approximately consistent across iterations when the network updates the weights. However, since nothing prevents any activation distributions from significantly shifting in one particular direction, this can occasionally result in runaway weight values and an overall collapse of the network.

Batch normalization [84] is an approach that significantly reduces this issue. The solution is relatively straightforward. The mean and standard deviation are calculated for each input channel across the batch by a batch normalization layer, which then normalizes by subtracting the mean and dividing by the standard deviation. The scale and shift are the following two learned parameters for each channel. The output is simply the normalized input that has been shifted by beta and scaled by gamma.

In practice, batch normalization is usually used as a layer after convolutional layers or fully connected layers to normalize the output. During training, a batch normalization layer calculates the moving average of each channel mean and standard deviation and stores this value as part of the layer. There are two trainable parameters within a batch normalization layer: the scale and shift. The moving average and standard deviation are nontrainable parameters, although they need to be calculated for each channel. However, they are derived from the data passing through the layer rather than trained through backpropagation.

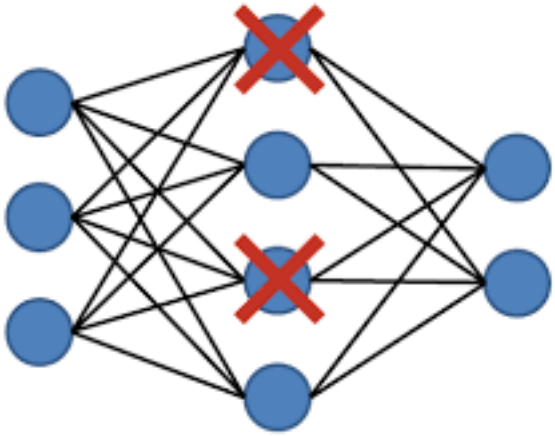


Figure 2.3: Dropout layer.

2.3.7 Dropout

Good training for a deep neural network is when it can generalize on unseen data rather than simply remembering the training dataset. It only performs well on the training dataset but not the test dataset, which leads to overfitting issues. To handle this problem, regularization techniques are usually applied to ensure that the model is penalized if it starts to overfit. Among many ways to regularize algorithms, dropout [81, 195] is one of the most popular methods. The idea is straightforward. During training, each dropout layer chooses a random set of units from the

preceding layer and sets their output to zero to remove it. Figure 2.3 illustrates dropout layer on hidden units.

This addition significantly reduces overfitting by ensuring that the network does not become overdependent on particular units or groups of units that, in effect, remember observations from the training set. It prevents the network from relying too much on a particular unit; therefore, knowledge is more evenly spread across the whole network. This makes the model much better at generalizing to unseen data because the network has been trained to produce accurate predictions even under unfamiliar conditions, such as those caused by dropping random units. There are no weights to learn within a dropout layer, as the units to drop are decided stochastically. In addition, the dropout layer does not drop any units during testing; hence the entire network is used to make predictions.

2.4 Convolutional Neural Network

Convolution is a mathematical operation fundamental to many well-known image processing operations. It is performed by multiplying pixel by pixel windows by a part of the image and summing the result. The result is more positive if the portion of the image exactly matches the filter and more negative if the portion of the image is the inverse of the filter. A filter, sometimes called a kernel, is a small matrix with a specific value. When the filter cuts through a group of pixels, it combines the values of the pixels to obtain a new matrix that picks out a particular feature of the input. Figure 2.4 indicates how convolutional operations work in the usual case.

Strides and padding are two additional parameters used in convolution in addition to the filter. The step size demonstrates the range that layers use to move the filters across the input is specified by the stride parameter. Therefore, increasing the stride causes the output to be smaller. For instance, when strides equal 2, the output height and width will be half as large as the input. It is useful for reducing the spatial size of the tensor as it passes through the network while increasing the number of channels. Padding refers to zero pads added to the input data so that the output size of the layer is identical to the input size. This allows the kernel to extend over the edge of the image so that it fits exactly into the convolution step.

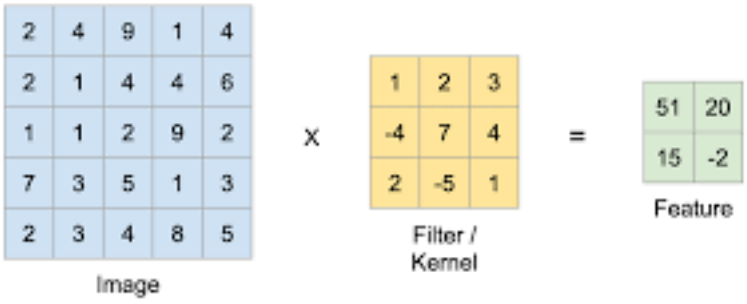


Figure 2.4: Convolutional operation.

Convolutional neural networks (CNN) [115] are specific neural networks that share weights and parameters across space. A typical CNN consists of multiple convolutional layers, followed by pooling layers and activation functions, and then a fully connected output layer. Convolutional operators are used to designing convolutional layers.

In convolutional network terminology, there are three primary components: the first argument to the convolution is often referred to as the input, the second argument is the kernel, and the output is the feature map. In machine learning applications, the input is referred to as a multi-dimensional array of data, and the kernel is typically an array of parameters that are adapted by the learning algorithm. The objective of convolution is to map and patch feature maps to the desired feature maps through the kernels. A stride is the number of elements in the multi-dimensional array that is shifted each time along with the kernel.

Unlike feedforward networks, where each unit contains an independent weight, the values of each element in the filters are the weights learned by the neural network through training. Therefore, it is considered that weights are shared between units. These values are initially random, but the filters gradually adapt weights to pick out interesting features such as edges or colour combinations. Parameter sharing refers to using the same parameter for more than one function in a network. In a conventional neural network, each component of the weight matrix is used precisely once when calculating the output of a layer. It is multiplied by one input element before being left alone. In a convolutional neural net, each kernel member is used at every position of the input. Since the convolution operation uses parameter sharing, the network only learns one set of parameters in general rather than multiple sets for each location.

Convolutional layers in a CNN are usually stacked together to form a pyramid-like structure. After each convolution layer, input is reduced in the spatial dimension while the feature map size is increased, roughly equivalent to its semantic complexity. Figure 2.5 illustrates a conventional convolutional neural network used for image classification.

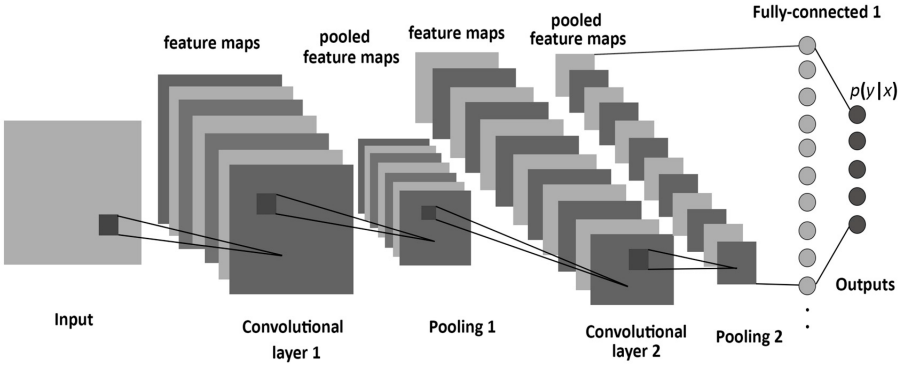


Figure 2.5: Convolutional neural netowrk architecture.

2.5 Representation learning

In machine learning, representation learning [19] presents a set of techniques enabling a system to exploit the representations required for features from raw data automatically. It replaces manual feature engineering and allows a machine to learn and use the features to perform a specific task. Representation learning can be either supervised or unsupervised. In supervised representation learning, features are learned using labelled input data, while features are learned with unlabeled input data in unsupervised representation learning.

For example, feedforward networks trained by supervised learning perform representation learning. Linear classifiers, such as a softmax regression classifier, are frequently used as the final layer, while the rest of the network work to learn a good representation of the classifier to

perform the final task. In unsupervised algorithms, it has the main training objective but also learns a representation as a side effect. No matter how a representation was created, it can be applied to different tasks.

Along with conventional deep learning methods, shared learning algorithms [224] also are becoming an exciting research topic in machine learning while they can share statistical strength across different tasks, including using information from unsupervised tasks to perform supervised tasks. Shared representations are useful when handling multiple modalities or transferring learned knowledge to tasks for which few or no examples are provided. Alternatively, multiple tasks can be learned together with some shared internal representation.

Representation learning is exciting because it provides one way to perform unsupervised and semi-supervised learning. For example, a dataset only includes a minimal amount of labelled data, while the number of unlabelled data is significant. Severe overfitting frequently occurs when supervised learning techniques are applied to the labelled subset. Semi-supervised learning provides a way to overcome the overfitting issue by incorporating learning from unlabeled data. The supervised learning task can be explicitly resolved by learning effective representations for the unlabeled data.

2.5.1 Pre-train

Training deep neural networks with many layers remains a challenge due to the complexity of the task and high computational cost. When the number of hidden layers increases, preserving information propagated back to earlier layers is significantly reduced. It results that weights in early hidden layers are rarely or never updated, whereas weights in final hidden layers are typically updated. This issue, also known as the vanishing gradient problem, typically prevented the training of very deep neural networks. It is a common problem in deep learning because as more layers are added to neural networks, the gradients of the loss function approach zero, making the network hard to train.

Greedy layer-wise pre-training, often known as pre-training, is a technique that initially allowed the development of deeper neural network models. Pre-training aims at adding a new hidden layer to a model and then retraining it. It allows the new model to learn to map on new inputs while maintaining the parameters of the existing hidden layers of the old model. Pretraining works based on the assumption that it is simpler to train a shallow network as opposed to a deep network, and it devises a layer-by-layer training approach that fits a shallow model.

There are two primary pre-training strategies: supervised and unsupervised greedy layer-wise pre-training. In general, the overall training scheme for both supervised and unsupervised pre-training is nearly the same in most cases. During supervised pre-training, hidden layers are successively added to a model trained on a supervised learning task. Unsupervised pre-training, on the other hand, relies on a single-layer representation learning algorithm that learns latent representations. Each layer is pre-trained by unsupervised learning, using the output of the previous layer to generate a new data representation. There are two important factors for unsupervised pre-training: the initialization of parameters and the learning distribution of input. The initialization for neural network parameters can significantly affect to model regularizing, whereas learning input distribution can extend the input-to-output mapping. When the number of unlabeled examples is significant and can be used to initialize a model before using smaller examples to fine-tune the model weights for a supervised task, unsupervised pre-training may be

appropriate.

Even though the weights in the last layers are held constant for both tasks, it is common to fine-tune all weights in the network after adding a new final layer. Pretraining is a significant milestone in developing deep learning, as it enables the creation of networks with more hidden layers than ever before.

2.5.2 Transfer Learning

Transfer learning refers to the topic in which a model learned to perform a task is exploited to improve generalization in a task [251]. In general, transfer learning allows new models to perform more than two different tasks, including tasks of the captured model. For example, a model is trained to classify visual categories, such as bikes and cars, and then it can be used in the second model to learn about different visual categories, such as planes and boats. Suppose the first model contains significantly more data than the second. In that case, this may make it easier to learn representations that can be quickly generalized using only a small number of examples from the second model.

In general, transfer learning can be achieved via representation learning when there exist features that are useful for different research purposes or tasks, corresponding to underlying factors that appear in more than one setting.

Domain adaption [165, 18] is a sub-category of transfer learning where the task remains the same between models, but the input distribution is slightly different. The objective is to use data from the first model to extract information that could be useful for learning or even for making predictions directly in the second model. The fundamental key of representation learning is that the single representation may be effective in various contexts. Representation learning can take advantage of the training data available for both tasks by using the same representation.

2.6 Conclusion

This section has provided an overview of machine learning with definitions and examples of supervised and unsupervised learning. The basic deep learning/deep neural network concepts are also presented. Deep neural networks are generally completely flexible by design, and there are no fixed rules for model architecture. In addition, the term convolutional layers and convolutional neural networks are introduced. The concept of convolutional layers is a significant part of deep neural networks, which has been successfully applied in many computer vision tasks. In the thesis, it has been used to build generative neural networks to solve different problems within the objective.

Chapter 3

Generative models

Deep learning algorithms hold great potential in computer vision, e.g., pattern recognition, classification, or regression. With this development, image synthesis using deep learning methods has become a potential research area due to its tremendous advantages. Generative models describe methods that produce outputs in a way that has no apparent relationship to probability distributions over possible input samples. The term generative indicates the primary purpose of the model: to generate new data.

Initially, generative methods were mostly based on statistical models, such as the Gaussian mixture model, the Hidden Markov model, etc. Later, with the development of neural networks, methods with autoencoders (AE) and later variational autoencoders (VAE) were widely applied.

When Ian Goodfellow et al. [69] first presented Generative Adversarial Networks (GAN), it was a breakthrough in image synthesis. GANs have achieved remarkable results long thought to be virtually impossible for artificial systems, such as transferring image styles or generating fake images with near-realistic quality without requiring huge amounts of tediously labelled training data. By proposing an architecture consisting of 2 separate neural networks, GANs have enabled computers to generate impressively realistic data.

GAN is a class of generative models judged primarily on comparing specific outputs to potential inputs. Before the invention of GANs, the best machines could produce a blurred countenance - and even that was celebrated as a breakthrough success using AEs. Later, advances in GANs enabled computers to synthesize false faces whose quality rivalled high-resolution portrait photos.

A generic GANs consists of two simultaneously trained models: the generator trained to generate fake data and the discriminator trained to discern the fake data from actual examples. As a generative method, GAN aims to produce new data with desired content from the input. On the other hand, the term "adversarial" points to the dynamic competition between the two models that constitute the generator and the discriminator.

The generator goal in GAN is to produce examples that capture the characteristics of the targets, so much so that the output it generates looks indistinguishable from real ones. The generator might be considered as an object recognition model but in a reverse direction. Object recognition algorithms learn the patterns of images to complete tasks. In contrast, the GAN generator learns to create the patterns essentially from scratch rather than recognizing them. The input to the generator is frequently just a vector of random numbers.

On the other hand, the discriminator goal is to determine whether a particular example comes

from the training dataset or is created by the generator. Therefore, the generator knows it did something right each time the discriminator is tricked into classifying a fake image as real. Conversely, each time the discriminator correctly rejects an instance as fake, the generator receives the feedback that it needs to improve.

The following section will introduce generative neural network architectures for image synthesis. Starting with the most fundamental algorithm, the autoencoder (AE), we will present generic AE several and its variants. Although AE has not been used during this PhD, it is an essential part of modern deep learning for image synthesis, which achieved state-of-the-art before. We will mention them briefly in terms of architecture. Then, we will introduce adversarial methods in detail. It will be the most important part of this thesis when we provide the generic architecture of GANs and their variants, which we have used extensively. The architecture GAN and its generator and discriminator are separated neural networks. Depending on the complexity of the GAN implementation, these can range from simple feed-forward neural networks to convolutional neural networks or even more complex variants.

3.1 Autoencoders

Autoencoders are neural networks that are trained to extract useful representations and reconstruct inputs in an unsupervised strategy while minimizing information loss [80]. AE models learn the approximation of the identity function; while this may appear straightforward, it extracts critical representations from dimensionality reduction and higher-level features by imposing various constraints on the network design and activations. Autoencoders have been utilized for decades and can be viewed as a non-linear alternative to PCA with the capacity to learn complex transformations of the input data. In addition, they can be trained with SGD, which has linear complexity with the number of samples.

3.1.1 Architecture

The general architecture of AE is represented in Figure 3.1. An AE contains three components: encoder, decoder and a latent space. The encoder maps the input from space $\mathcal{X}$ to a latent feature space $\mathcal{Z}$, and then the decoder reconstructs it back to the original data as closely as possible. Latent space is typically a representation of a smaller dimension and acts as an intermediate step. The generation part only happens in the latent space and the decoder. In other words, autoencoders can systematically and automatically uncover these information-efficient patterns, define them, and use them as shortcuts to increase the information throughput. The parametric mapping $p_\phi(Z|X)$ from X to Z are represented by a neural network with parameters ϕ. In general, only the Z is needed to be transmitted, which is typically much lower-dimensional, thereby saving the bandwidth.

Here, the encoder can be represented by a function $p_\phi : \mathcal{X} \to \mathcal{Z}$, and $q_\theta : \mathcal{Z} \to \mathcal{X}$, where ϕ and θ are the parameters of encoder and decoder respectively. The goal of training is to tune the parameters of the encoder and the decoder to find the best appropriate parameters for the two networks and get a sense of how the examples are represented in the latent space.

$$\max_{\phi,\theta} E_{(X,Z)\sim p_\phi}[\log_{q_\theta}(X|Z)] \tag{3.1}$$

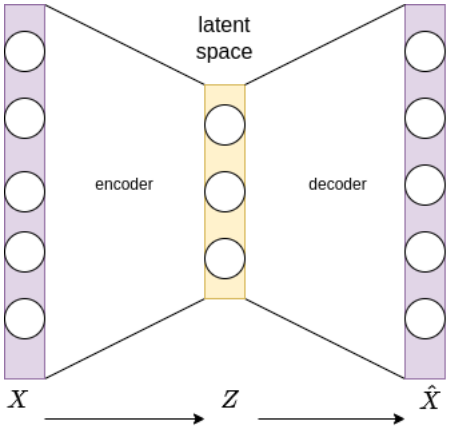


Figure 3.1: General autoencoder architecture.

One of the critical distinctions with autoencoders is that it only requires one loss function to train the whole network. First, the encoder is trained by passing the input X with the aim that the decoder is managed to reproduce $\hat{X}$ back as close to X as possible. That loss function is called the reconstruction loss.

In other words, the reconstruction $\hat{X}$ is presented for the mean of a distribution that has generated X. In the case of continuous variables, Gaussian distribution is the most popular approache which leads to the mean squared error (MSE) loss:

$$\mathcal{L}_{MSE}(X, \hat{X}) = ||X - \hat{X}||_2^2 \tag{3.2}$$

In case of binary variables, it can expressed to a probability which has range between [0,1], the common choice is Bernoulli distribution which lead to binary cross-entropy loss:

$$\mathcal{L}_{BCE}(X, \hat{X}) = -\sum_{i=1}^{N} X^i \, log \, \hat{X}^i + (1 - X^i) \, log(1 - \hat{X}^i)$$

3.1.2 Regularized AE

In general, AEs are constantly trained with some regularization to reduce the size of the hypothesis space to learn good representations of the data called regularized AE [3]. The most basic form of regularization is to use an intermediate feature space through a lower dimension, forcing the model to learn an efficient code with fewer parameters.

There are two kinds of regularized AE: under-complete, where the internal layer has a smaller number of units than the input AE and overcomplete, where the internal layer has more units than the input layer. Both map the input space to a lower-dimensional feature space, thus reducing non-linear dimensionality.

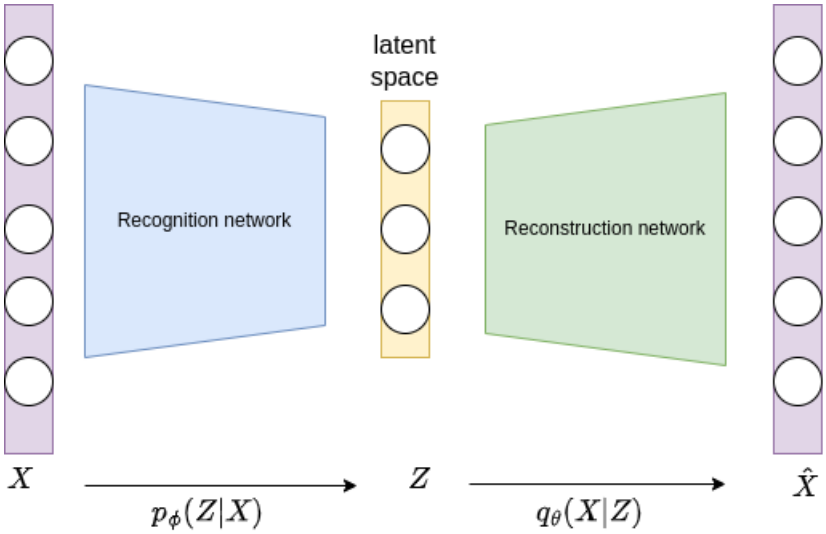


Figure 3.2: Variational autoencoders architecture.

Regularized AE are trained not only to optimize the reconstruction error but with additional constraints on the structure of the latent space. Several techniques have been proposed to prevent autoencoders from learning the identity function and also to improve the ability to capture important information and learn more meaningful representations. Among these, there are three main classes: sparse, denoising and contractive autoencoder.

Sparse autoencoders [147, 135, 9] presented data compression by applying a theoretically high number of units in each hidden layer, whereas only a set of units are active simultaneously. Sparsed output are obtained by the use of KL-divergence distance [147] or L1 loss [9]. In contrast, contractive autoencoders [169] utilize the loss function in the form of the squared Frobenius norm of the Jacobian matrix in the encoder. It becomes equivalent to L2 weight decay in the case of a linear AE.

Denoising autoencoders [217] learn to reconstruct a duplicated version of the input to minimize loss when the input vector is produced by the stochastic corruption process , e.g, random dropout or additive noise. Through the training to remove noise and recover the original data after noise removal, the network learns useful representations and extracts valuable features from the input distribution.

Later the stacked denoising autoencoders were proposed to enhance the performance of the generic model for several reconstruction tasks. The stacked model applied the advantage of pretrained (Section 2.5.1) to initialize each layer to enhance the representation of the feature. It consists in stacking several networks by using the output of each AE as the input of the next AE. First, the network is pretrained using greedy layer-wise training [21]. Then, all encoder layers followed by decoder layers are concatenated in reverse layer-wise training order, forming a deeper network. Next, the stacked denoising autoencoder is finetuned on reconstruction error.

3.1.3 Variational autoencoders

The standard under complete AE has limitations because latent spaces are unstructured objects and may not be continuous or overfit. Theoretically, even a single continuous latent variable can memorize the entire training set using one real number per sample.

Variational autoencoder (VAE) [103], a deep latent-variable probabilistic model, has solved the existing problems. The original VAE is built based on Bayesian machine learning. VAE aims to learn the distribution by finding the suitable parameters defining that function in latent space. Samples are produced from the latent distribution and then fed into the decoder to reconstruct the input. Figure 3.2 illustrates a general architecture of a VAE.

In VAE, the latent space is represented as a distribution composed of a learned mean and standard deviation instead of a set of numbers. The decoder act as a generative model that samples the data using the likelihood $q_\theta(X|Z)$, while the encoder is an inference or recognition model that applies the posterior distribution $p_\phi(Z|X)$. The goal of the encoder is to approximate the posterior, which is intractable.

As a machine learning problem, it is necessary to define a loss function to update the network weights through backpropagation. The objective is to jointly estimate the generative model parameters θ and the variational parameters ϕ to minimize the reconstruction error between input and output of the network, using maximum likelihood estimation (Equation 3.1). Similar to AE, the reconstruction loss of VAE are often mean squared error (Equation 3.2) and cross-entropy (Equation 3.1.1). Besides, to maximize the log-likelihood to improve the generated data quality and to minimize the distribution distances between the real posterior and the estimated one, VAE introduced the evidence of lower bound loss function (ELBO), which, based on Kullback–Leibler divergence [103]. In addition, to make the ELBO suitable for training purposes, the reparameterization trick is a sampling technique while maintaining different operations to allow end-to-end optimization by backpropagation [104].

Later, several studies were proposed with the aim of improving the performance of VAE. Popular studies can be highlighted, such as β-VAE [78] which uses tunable β hyperparameter weights to balance the ELBO loss, ladder VAE [192] which applies the batch normalization and deterministic warm-up to train the variational models with many stochastic layers. Other methods use multiple stochastic variables such as: importance weighted autoencoders [31], normalizing flows [168], inverse autoregressive flows [105], Variational Gaussian Processes [76]. Finally, recent studies has been demonstrated to improve the latent space structure and disentanglement such as : InfoVAE [246], β-TCVAE [36], FactorVAE [99], π-VAE[143].

3.2 Generative adversarial network

GANs are still considered part of the unsupervised learning strategy in deep learning, even through the automated labelling process. GANs are extremely potent because they can perform complex tasks instead of latent space interpolations of the autoencoder.

The main concept of GANs is straightforward. As a generative model, GAN contains two components: generator and discriminator, to learn the input distribution to produce samples. The primary function of the generator is to generate new examples that can fool the discriminator, while the discriminator is trained to classify that examples are generated or real data. Figure 3.3

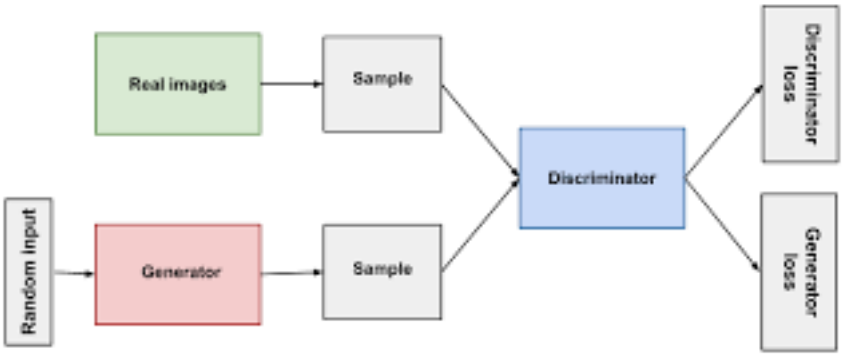


Figure 3.3: General GAN architecture.

illustrates the general architecture of a GAN model. In the ideal case, at the end of the training process, the discriminator will no longer be able to see the difference between the synthetically generated data and the real ones, and the generator can then be used to generate new, never-before-observed, realistic data.

The idea of training both the generator and discriminator simultaneously in order for both networks to learn the distribution of data is an exciting research topic. However, achieving stable training in the generator-discriminator network is not easy. When the discriminator computes the lost functions, network weights are rapidly updated. However, if the discriminator converges faster than the generator, it cannot receive sufficient gradient updates for parameters and consequently fails to converge. Moreover, GAN training is also affected by the partial or total modal collapse, where the generator produces nearly identical outputs for different latent encodings.

This section describes the concepts underlying GAN. Then, advanced concepts, including GAN and its variants, will be covered.

3.2.1 Adversarial architecture

Following Goodfellow et al. [69], GAN contains a generator G and a discriminator D. Formally, the generator and the discriminator are represented by differentiable neural networks, each with its cost function.

To regularize terms for the adversarial explanation, we use x to present real data, and z represents an arbitrary encoding or noise vector to synthesize new signals. Mathematically, the generator works to produce $G(z) = \hat{x}$ that is supposed to be as close to a real example as possible, $\hat{x} \approx x$. The discriminator takes either a real example x or a fake example $\hat{x}$ to classify. For the real examples, $D(x)$ seeks to be as close as possible to 1, while for fake examples, $D(\hat{x})$ strives to be as close as possible to 0, while the generator strives to produce fake examples $\hat{x} \approx x$ such that $D(\hat{x})$ is as close to 1 as possible.

The goal of adversarial training is to solve the adversarial min-max problem in Equation 3.4. The generator aims to minimize errors while the discriminator tries to maximize them. The final objective is to train the generator G to fool a differentiable discriminator D that is trained to distinguish generated SR images from A images. With this approach, the generator can learn to create highly similar solutions to real images, hence difficult to classify by D. The generator can learn to produce extremely similar solutions to real images, making them challenging for D to

classify. This promotes the existence of perceptually superior solutions in the subspace or manifold of real images.

$$\min_G \max_D E_{x \sim p_{data}(x)}[\log D(x)] + E_{z \sim p_z(z)}[\log(1 - D(G(z)))] \tag{3.4}$$

where $p_{data(x)}$ and $E_{x \sim p_{data(x)}}$ are respectively the empirical distribution and data distribution of real data x [67], while the $p_z(z)$ present for a prior on input variables.

Significantly, the training dataset determines the kind of examples the generator will learn to emulate. The input to the generator can be any kind of data, and the output is a synthesized one with desired content. Meanwhile, the discriminator input is the real or synthesized data. Genuine data comes from the true sampled data, while the generator produces the generated data.

The goal of the generator is to produce examples that capture the data distribution of the training dataset, while the discriminator learn distribution to distinguish real data from fake data. It acts as a classifier to produce the probability of a given sample. All valid data is labelled as a value in the range [0.0, 1.0]. Algorithms 2 summary the training process of a GAN model.

Algorithm 2 GAN training algorithm

for each training iteration **do**

- Train the discriminator:
  1. Take a random mini-batch of real examples x
  2. Take a mini-batch of random noise vectors z and generate a mini-batch of fake examples: $G(z) = \hat{x}$.
  3. Compute the classification losses for $D(x)$ and $D(\hat{x})$, and backpropagate the total error to update θ_D to minimize the classification loss.
- Train the generator:
  1. Take a mini-batch of random noise vectors z and generate a mini-batch of fake examples: $G(z) = \hat{x}$.
  2. Compute the classification loss for $\mathrm{D}(\hat{x})$, and backpropagate the loss to update θ_G to maximize the classification loss.

end for

Both generator and discriminator are trained using backpropagation. Theoretically, the generator and discriminator are trained simultaneously. However, the generator and the discriminator are trained alternatively in practice. During classifying, only the discriminator parameters will be updated. The discriminator is trained to classify the given input as the real or generated sample. At regular intervals, the generator defines its output as actual data and labels it 1.0. When that sample is presented to the discriminator, it will be classified as fake with a label close to 0.0 by default.

On the other hand, the generator parameters can only be updated only when it produces a realistic sample. During the generative phase, the discriminator parameters are typically temporarily frozen in practice. The gradients are then allowed to backpropagate from the final layer of the discriminator to the initial layer of the generator. The generator will finally utilize the gradients to update its parameters and enhance its ability to generate realistic samples.

Overall, the process is comparable to two networks competing against one another while also

cooperating. When models converge, they can synthesize data that appears genuine. The discriminator believes that the synthesized data is real or has a label near 1.0, allowing it to be discarded. The generator component will be useful for generating meaningful outputs from random noise inputs.

3.2.2 Loss function

Following the standard notation, let $\mathcal{L}_G$ denote the generator loss function and $\mathcal{L}_D$ for the discriminator loss function. Parameters of two networks are represented by θ_G for the generator and θ_D for the discriminator.

In contrast, GANs consist of two networks whose cost functions depend on both network parameters. Hence, the generator loss function is $\mathcal{L}_G(\theta_G, \theta_D)$, and the discriminator lost function is $\mathcal{L}_D(\theta_G, \theta_D)$. A neural network can generally tune all its parameters during the training process. However, each network can tune only its own parameters in a GAN. The generator can control only θ_G, while the discriminator can tune only θ_D during training. Accordingly, each network has control over only a part of what determines its loss. Because each loss function depends on the other network parameters, but each network cannot control the other network parameters, this scenario is most straightforward to describe as an optimization problem. Hence, it is not easy to balance the parameters for both networks to ensure efficient training.

The cost used for the discriminator is:

$$\mathcal{L}_D(\theta_G, \theta_D) = -E_{x \sim p_{data}}[\log D(x)] - E_{z \sim p_z}[\log(1 - D(G(z)))] \tag{3.5}$$

This loss function is based on the standard cross-entropy cost that is minimized when training a standard binary classifier with a sigmoid output. It is the negative sum of the expectation of correctly identifying real data $D(x)$, and the expectation of correctly identifying synthetic data, $1D(G(z))$. The log does not change the local minima.

The loss value of the generator is based on zero-sum, in which the sum of all loss functions for both networks is always zero. Hence, it is simply the negative of the discriminator loss function [67].

$$\mathcal{L}_G(\theta_G, \theta_D) = -\mathcal{L}_G(\theta_G, \theta_D) \tag{3.6}$$

Thus, based on Equation 3.6, from the perspective of the generator, its loss functions should be minimized, but from the point of view of the discriminator, its loss function should be maximized. Maximizing with respect to θ_D, the optimizer updates gradient on discriminator in order to pretend synthesized sample to be real. Simultaneously, by minimizing with respect to θ_G, the optimizer updates parameters of generator to fool the discriminator.

3.2.3 Training GAN difficulty

Training a GAN can be complicated since it has to balance the training of two separated neural networks simultaneously. Several studies have figured out shortcomings of GAN training in different scenarios [87, 60]. These drawbacks of GAN training can be categorized into five main tasks:

- Oscillating loss: the loss of the discriminator and the generator may start to oscillate wildly instead of showing long-term stability. Typically, the loss fluctuates slightly between batches. However, it is expected to stabilize or gradually increase/decrease rather than fluctuate erratically to ensure that your GAN converges and improves over time.

- Mode collapse: it occurs when the generator exploits a small number of examples that trick the discriminator and, as a result, cannot generate any other examples. The generator, for instance, is trained over multiple batches without updating the discriminator in between. The generator would tend to identify a single mode that consistently fools the discriminator, at which point it would assign every point in the latent input space to that observation. This indicates that gradients of loss function can reach zero. Even if the discriminator is retrained to avoid being fooled by this generator, the generator will simply find another mode that fools the discriminator, as it has become accustomed to its input and therefore has no incentive to diversify its output.

- Uninformative loss: since deep neural networks are compiled to minimize the loss function, it is reasonable to assume that the smaller the loss function of the generator, the higher the image quality. The loss functions cannot be compared to evaluate them at different stages of the training process because the generator is only evaluated against the current discriminator and the discriminator is constantly improving. Because there is no correlation between generator loss and quality of data, it can be challenging to monitor GAN training.

- Overgeneralization: when the model produces outputs with more content than expected or vice versa. This happens when GAN overgeneralizes and learns things that should not exist based on the real data.

- Hyperparameters: there are several hyperparameters that must be configured. In addition to the discriminator and generator overall architecture, there are stack normalization parameters, dropout, learning rate, activation layers, convolutional filters, kernel size, striding, batch size, and latent space size to take into account. GANs are extremely sensitive to small changes in each of these parameters, and finding a working set of parameters is often a matter of trial and error as opposed to following a set of predetermined guidelines.

Recently, several solutions have been proposed to solve the difficulty of training GAN and also to improve the training process, such as increasing network depth [67], feature matching [178], mini-batch discrimination [84], normalization techniques, etc.

Besides, as mentioned previously, training GAN is based on the min-max problem. The discriminator minimizes the loss function, but the generator maximizes the same loss function. This is unfortunate for the generator because when the discriminator successfully rejects generator samples with high confidence that lead to the gradient vanishes for the generator. As a result, the generator fails to converge. In practice, the discriminator is confident in its prediction in classifying the synthetic data as fake and will not update the GAN parameters. Furthermore, the gradient updates are small and have diminished significantly as they propagate to the generator layers. To solve this problem, the non-saturating GAN techniques are used to solve this problem. In order to keep minimizing the loss function on the generator while maximizing it on the discriminator, the content of the loss function for the generator is flipped to construct the new loss function. The cost for the generator then becomes:

$$\mathcal{L}_G(\theta_G, \theta_D) = -E_{z \sim p_{data}} \log D(G(z)) \tag{3.7}$$

By training the generator, the loss function simply maximizes the likelihood of the discriminator believing that the synthetic data is real. The new formulation is purely heuristic and no longer zero-sum. The generator parameters are updated only after the entire adversarial network has been trained. Because gradients are transmitted from the discriminator to the generator. During adversarial training, the discriminator weights are frozen only temporarily.

3.2.4 Deep Convolutional GAN

In general, the architecture of the generator and discriminator can be any model. It can range from a simple feed-forward neural network with a single hidden layer to a deep neural network. If the data is an image, both the generator and discriminator networks will use a CNN. That GAN architecture using deep CNN is known as Deep Convolutional GAN (DCGAN).

The first DCGAN was introduced in [163] in 2016, marking one of the most important early innovations in GANs. It was not the first attemp to use CNNs in GANs, but it was the first time group of researchers were able to successfully integrate CNNs into a full-scale GAN model. In fact, the use of CNNs increase the difficulty for GAN training due to the instability and gradient saturation. Existing challenges require researchers to propose alternative approaches, such as the LAPGAN [50]. This method applied a cascade architecture within a Laplacian pyramid, whereas the model is trained at each level using the GAN framework.

The DCGAN introduced a new scheme and optimizations to scale up CNN architecture to the full GAN framework. It did not require to reduce general GAN architecture or modify the underlying GAN. By using batch normalization [84] during feature extraction, the network can stabilize the training process by normalizing inputs at each layer.

Normalization is the technique to scale data to reduce the size of the computation. Normalization has several advantages. Perhaps most important, it makes comparisons between features with vastly different scales easier and, by extension, makes the training process less sensitive to the scale of the features. The idea behind is to normalize the inputs to each layer for each training minibatch as it flows through the network. And as the parameters get tuned by back-propagation, the distribution of each layer input is prone to change in subsequent training iterations, which destabilizes the learning process covariate shift [84]. Batch normalization solves it by scaling values in each minibatch by the mean and variance of that mini-batch.

3.2.4.1 Generator

CNNs have traditionally been used for classification tasks, in which the network takes in multi-dimensional data as input and—through a series of convolutional layers—outputs a single vector of class scores which relate to the probability of input for a domain.

In order to generate an image by using the CNN architecture, instead of taking an image and processing it into a probability vector, hidden layers are added to transform feature maps into images. The key to this process is the upsampling method.

The generator starts with an input which is in the form of a multi-dimensional matrix. Though fully connected layers, features are extracted and reshaped into a smaller size and larger depth. Then, at some points, the input is progressively reshaped such that its base grows while its depth decreases until we reach desired size and contents using sampling operators. The types of upsampling methods can be ranged from static interpolation to convolutional upsampling called

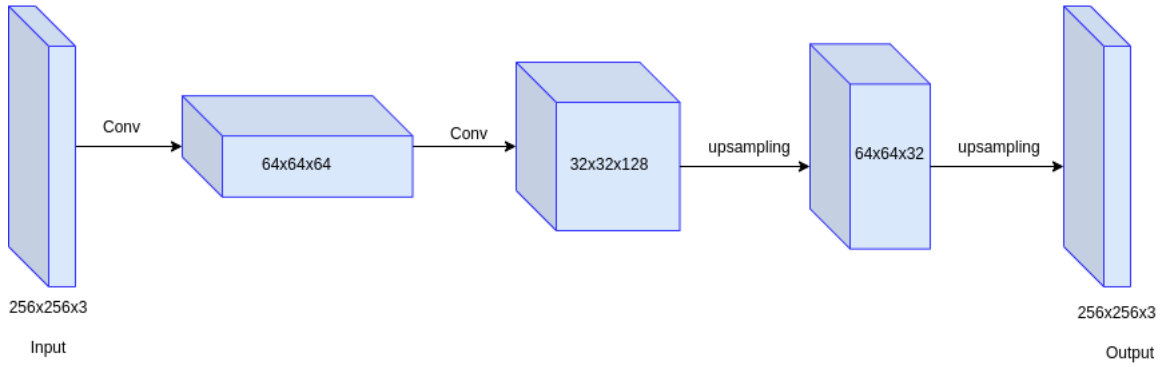


Figure 3.4: Illustration of DCGAN generator architecture.

transposed convolution (or deconvolution). Figure 3.4 and Figure 3.5 illustrate data flow in a simple DCGAN generator and discriminator with several layers.

3.2.4.2 Discriminator

The discriminator is a CNN that is similar to a classification network. The input is a multi-dimensional matrix, and the output is a prediction vector: in this case, a binary classification indicating whether the input image was classified to be real rather than fake.

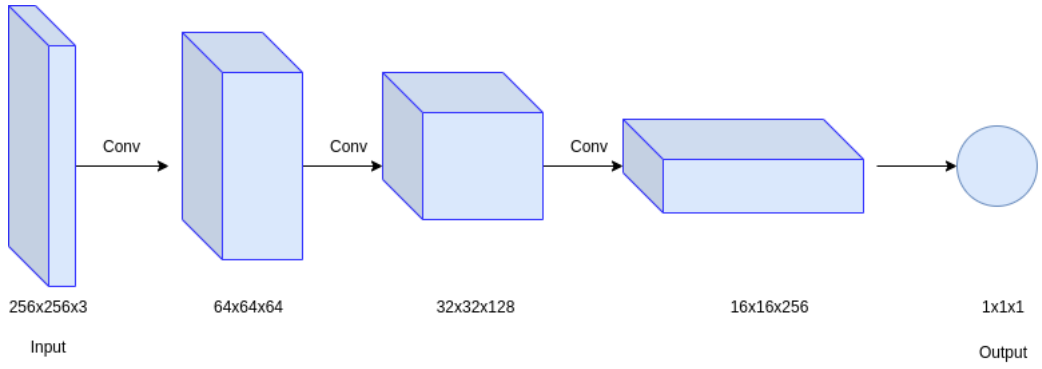


Figure 3.5: Illustration of DCGAN discriminator architecture.

3.3 GAN variations

Semi-supervised learning is one of the most promising areas of the practical application of GANs. Unlike supervised learning, in which we need a label for every example in our dataset, and unsupervised learning, in which no labels are used, semi-supervised learning has a class label for only a small subset of the training dataset. By internalizing hidden structures in the data, semi-supervised learning strives to generalize from the small subset of labelled data points to classify new, previously unseen examples effectively. Importantly, for semi-supervised learning to work, the labelled and unlabeled data must come from the same underlying distribution.

Since Goodfellow et al. [69] have proposed the generative adversarial networks (GAN) model, many follow-up studies of GAN and its variations have been applied in several computer vision tasks. GAN has proved its efficiency in achieving state-of-the-art performance in the image synthesis field. In this section, we will present variants of GAN networks, which has been used in this thesis.

3.3.1 Wasserstein GAN

As mentioned in previous sections, GANs are notoriously hard to train. The opposing objectives of the two networks, the discriminator and the generator, can easily cause training instability. The discriminator attempts to classify the fake data from the real data correctly. Meanwhile, the generator tries its best to trick the discriminator.

If the discriminator learns faster than the generator, the generator parameters will fail to optimize. On the other hand, if the discriminator learns more slowly, the gradients may vanish before reaching the generator. In the worst case, if the discriminator is unable to converge, the generator will not be able to get any useful feedback.

Wasserstein GAN (WGAN) [8] has been introduced by proposing a novel method to improve GAN training. WGAN argues that the stability in training a GAN depends on the loss of functions. In this study, they introduced the earth mover distance (EMD) as a loss function that clearly correlates with the visual quality of the samples generated. Figure 3.6 and Figure 3.7 indicate the training scheme of WGAN for both generator and discriminator.

In EMD, the loss function of the discriminator are defined as:

$$\mathcal{L}_D = -E_{x \sim p_{data}} \log D_w(x) + E_z D_w(G(z)) \tag{3.8}$$

This equation is quite similar to Equation 3.5, with some important differences.

Here, the D_w represent the discriminator. It aims to estimate the earth mover's distance and looks for the maximum difference between the real examples and the generated examples distribution under different valid parametrizations of the D_w function.

The discriminator is trying to make the generator harder to generate samples by looking at different projections using D_w into shared space in order to maximize the amount of probability mass it has to move.

In the case of a generator, the loss function is defined as:

$$\mathcal{L}_G = -E_z D_w(G(z)) \tag{3.9}$$

The objective is trying to minimize the distance between the expectation of the real distribution and the expectation of the generated distribution. WGAN contains more understandable loss with more tunable training.

Similar to GANs, WGAN alternately trains the discriminator and generator. However, in WGAN, the discriminator trains for some critic iterations before training the generator for one iteration. It contrasts with GANs with an equal number of training iterations for the discriminator and generator. Training the discriminator means learning the parameters of the discriminator. It requires two mini-batches from real and fake samples to compute the gradient of discriminator

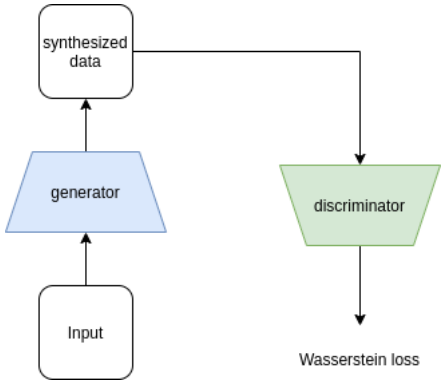


Figure 3.6: WGAN generator training. The generator produce images during this training. Synthesized data is pretended to be real with label=1. The discriminator weights are frozen but gradients propagate back to the generator

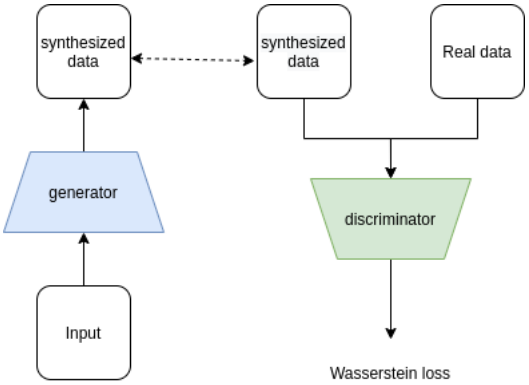


Figure 3.7: WGAN discriminator training.

parameters after feeding the sampled data to the discriminator network. Then, EMD optimization is imposed by clipping the discriminator parameters. After several critical iterations of discriminator training, the discriminator parameters are frozen. The generator training starts by sampling a batch of fake data. The sampled data is labelled as real (1.0), endeavouring to fool the discriminator network. Both generator and discriminator parameters are optimized using RMSProp optimizers. After training the generator, the discriminator parameters are unfrozen, and another iteration of discriminator training starts. These processes repeat until the model convergence.

Similar to GANs, the discriminator can be trained as a separate network. However, training the generator always requires the participation of the discriminator through the adversarial network since the loss is computed from the output of the generator network. The most practical implication of WGAN is that it allows the training of discriminators can be stopped by measuring the Wasserstein distance. Besides, the correlation between the discriminator loss and the perceptual quality also helps inform when to stop.

3.3.2 Conditional GAN

GANs are capable of producing various examples with any desired content. Although the synthesis domain can be controlled by learning to emulate the training dataset, GAN cannot specifically generate any of the characteristics of the data samples. The ability to decide what kind of data will

be generated opens the door to a vast array of applications. The conditional GAN (CGAN) [142] is one of the first GAN innovations that made targeted data generation possible, and arguably the most influential one.

Conditional GAN is a generative adversarial network whose generator and discriminator are conditioned during training by using some additional information. CGAN is generally similar to DCGAN except for the additional one-hot vector input. The one-hot label is concatenated with the latent vector before the dense layer for the generator. For the discriminator, a new fully-connected layer is added. The new layer is used to process the one-hot vector and reshape it so that it is suitable for concatenation to the other input of the next CNN layer.

During CGAN training, the generator learns to produce realistic examples from given labels, and the discriminator learns to distinguish fake example-label pairs from real example-label pairs. Moreover, the discriminator in a CGAN does not learn to identify which class is which. It learns only to accept real, matching pairs while rejecting mismatched pairs and pairs in which the example is fake.

The generator uses examples with their label to synthesize a fake example with the condition of label value. This fake example aims to look as close as possible to a real example for the given label.

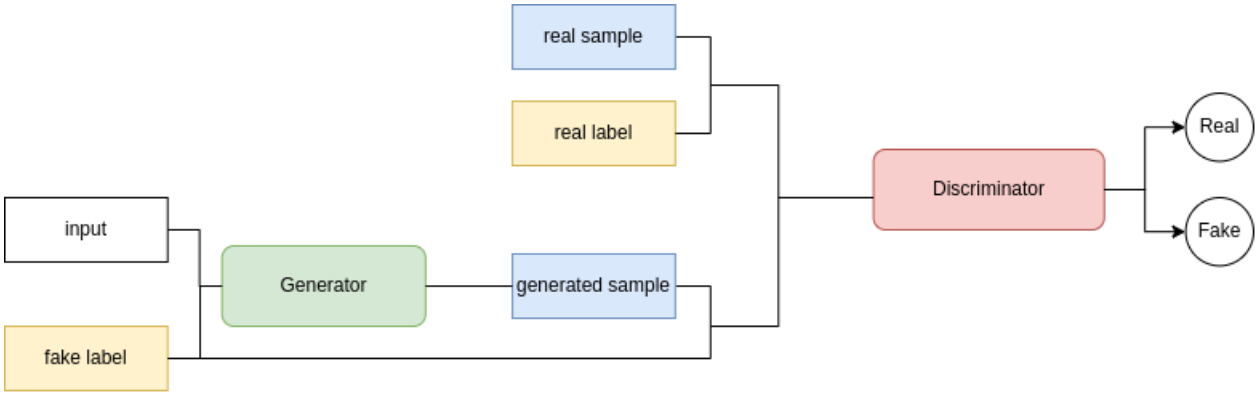


Figure 3.8: Illustration of CGAN workflow.

The discriminator receives real/fake examples with labels. On the real example-label pairs, the discriminator learns how to recognize real data and how to recognize matching pairs. On the synthesized examples, it learns to recognize fake image-label pairs, thereby learning to tell them apart from the real ones. The discriminator outputs a single probability indicating its conviction that the input is a real, matching pair. The discriminator's goal is to learn to reject all fake examples and all examples that fail to match their label while accepting all real example-label pairs. Figure 3.8 demonstrates the data flow of CGAN with fake/real image-label pairs.

The loss function of the discriminator and the generator of CGAN are defined as:

$$\mathcal{L}_D(\theta_G, \theta_D) = -E_{x \sim p_{data}}[\log D(x|y)] - E_{z \sim p_z}[\log(1 - D(G(z|\hat{y})))] \tag{3.10}$$

$$\mathcal{L}_G(\theta_G, \theta_D) = -E_z \log(1 - D(G(z|\hat{y}))) \tag{3.11}$$

where y and $\hat{y}$ are labels of real and synthesized data, respectively.

The new loss function of the discriminator aims to compute the error of lost function betweeen real images coming from the dataset and fake images coming from the generator, given their one-hot labels. In terms of the generator, the loss function of the generator maximizes the correct prediction of the discriminator on fake images conditioned on the specified one-hot labels. The generator learns how to generate the specific examples given its one-hot vector, which can fool the discriminator. Algorithm 3 summarize the training of CGAN for both generator and discriminator.

Algorithm 3 CGAN training algorithm

for each training iteration **do**

- Train the discriminator:
  1. Take a random mini-batch of real examples x with corressponding label y
  2. Compute $D((x, y))$ for the mini-batch and backpropagate the binary classification loss to update θ_D to minimize the loss.
  3. Take a mini-batch of input z and a class label y to generate a mini-batch of fake examples: $G(z, y) = \hat{x}|y$.
  4. Compute $D(\hat{x}|y, y)$ for the mini-batch and backpropagate the binary classification loss to update θ_D to minimize the loss.
- Train the generator:
  1. Take a mini-batch of input z and class labels (z, y) to generate a mini-batch of fake examples: $G(z, y) = \hat{x}|y$.
  2. Compute $D(\hat{x}|y, y)$ for the given mini-batch and backpropagate the binary classification loss to update θ_G to maximize the loss.

end for

Along with the DCGAN, CGAN is one of the most influential early GAN variants that has inspired countless new research directions. It might be the most impactful and promising adversarial network as a general-purpose solution to image-to-image translation problems. One of the most successful early implementations based on the Conditional GAN paradigm is pix2pix, which uses pairs of images to learn to translate from one domain into another. In theory and practice, the conditioning information used to train a CGAN can be much more than just labels to provide for more complex use cases and scenarios.

3.4 CycleGAN

Image-to-image translation can be considered as a special case of image synthesis. The pix2pix [85] algorithm is an example of a cross-domain algorithm. It was developed based on conditional GAN [142]. The condition here is related to a complete image rather than a class, typically of the same dimensionality as the output image that is then provided to the network as a kind of a label.

The idea is powerful and versatile; however, the main disadvantage of neural networks similar to pix2pix is that the training input and output images must be in pairs to start the training phase. In practice, aligned image pairs are not available or expensive to generate from the source images, or we have no idea how to generate the target image from the given source image.

Cycle-consistent adversarial network (CycleGAN) [250] proposed by Zhu et al. is an image-to-image translation model for learning to translate an image from a source domain to a target domain in the absence of paired examples. No alignment is needed. CycleGAN learns the source and target distributions and how to translate from source to target distribution from given sample data. CycleGAN has become a very effective method in the image synthesis and image translation domain.

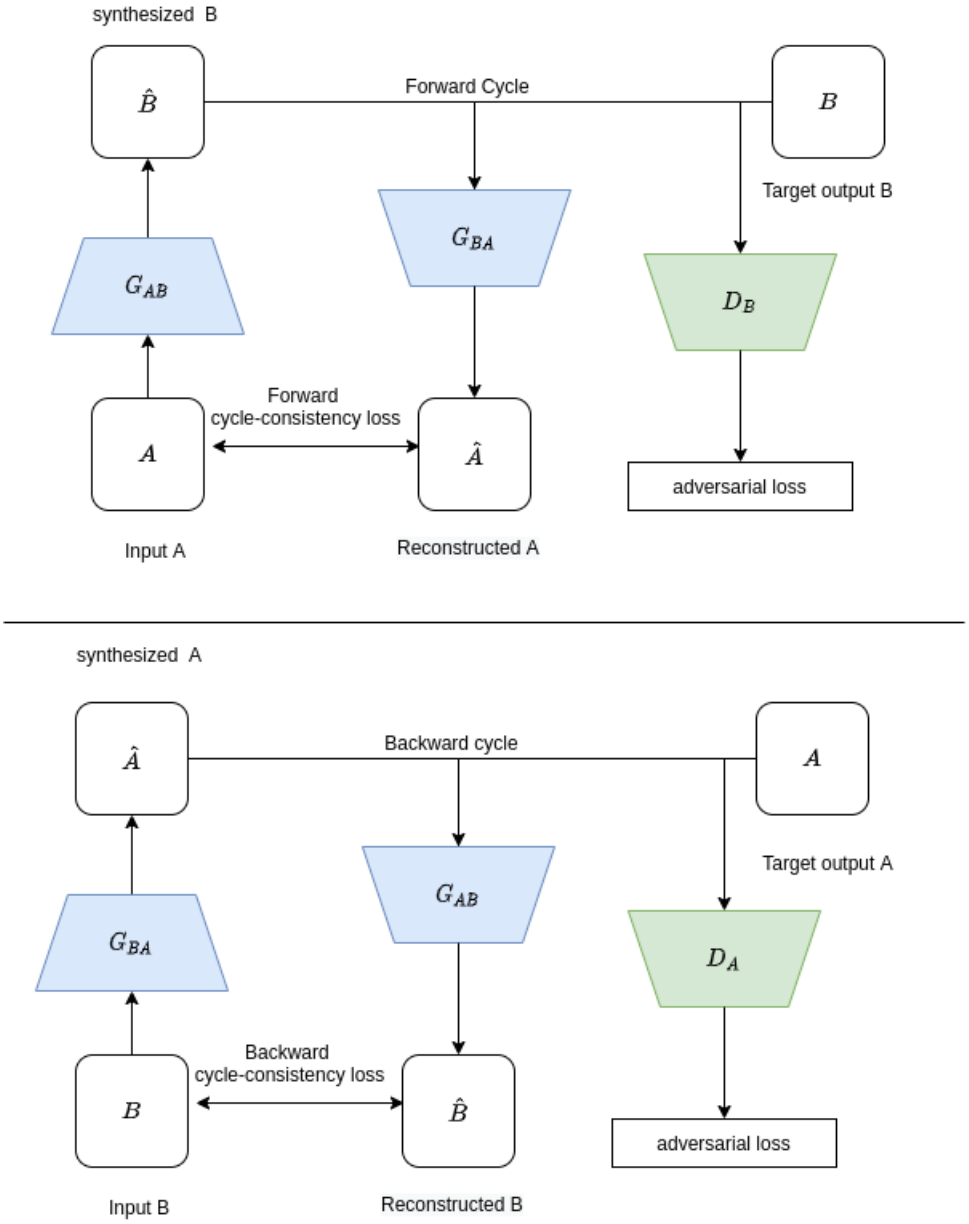


Figure 3.9: CycleGAN generator workflow.

3.4.1 General architecture

The overall CycleGAN architecture can be viewed as training GANs in reversed directions. Following the fundamental CycleGAN [250] model, the final purpose aims at synthesizing an image of the domain A into another image with desired content belonging to a different domain B. The

model contains two generators to produce images between two domains and two discriminators to predict whether the output is real or generated. The first generator G_{AB} translating the image from domain A to B, has a corresponding discriminator D_B; and G_{BA} has the discriminator D_A .

Generators and discriminators work in pairs. They aim to learn to create solutions that are highly similar to real images and difficult to classify. This encourages perceptually superior solutions residing in the subspace, or manifold, of real images. For both tasks, we used the same architecture of discriminator, except for the difference between functions. In the end, the discriminators can evaluate generated outputs by values ranging from 0 to 1. The overall architecture of CycleGAN is shown in Figure 3.9.

3.4.2 Loss function

CycleGAN is used for cross-domain translation. To regularize symbols for CycleGAN architecture explanation, we use A for the first domain B for the second domain. From the original CycleGAN model [250], the loss function of model based on the adversarial loss $\mathcal{L}_{adv}$ from original GAN [69] and the cycle consistency loss $\mathcal{L}_{cycle}$ which is used to prevent the generators can produce unrelated output during unpaired training, thereby improving training efficiency. The final loss $\mathcal{L}$ function of CycleGAN can be written as:

$$\mathcal{L} = \mathcal{L}_{adv}(G_{BA}, D_B, A, B) + \mathcal{L}_{adv}(G_{AB}, D_A, B, A) + \lambda \mathcal{L}_{cycle}(G_{BA}, G_{AB}) \tag{3.12}$$

In Equation 3.12, the first generator G_{BA} implements the mapping function $A \rightarrow B$ with its discriminator D_B, while the second generator G_{AB} implements the mapping function $B \rightarrow A$ with discriminator D_A.

3.4.2.1 Adversarial loss

In CycleGAN, since there are two generators and two discriminators, the adversarial loss is applied for both generators. Training with adversarial loss solves Equation 3.4. From Equation 3.4 and Equation 3.12, the adversarial loss of the first generator from A to B can be presented as:

$$\begin{aligned} \mathcal{L}_{adv}(G_{BA}, D_A, A, B) = E_{B \sim p_{data}(B)}[\log D_B(B)] + \\ E_{A \sim p_{data}(A)}[\log(1 - D_B(G_{BA}(A)))] \end{aligned} \tag{3.13}$$

where $B \sim p_{data(B)}$ and $A \sim p_{data(A)}$ are data distribution of B and A based on [67].

G_{BA} generates images $G_{BA}(A)$ that are expected to look as similar as images B, while D_B aims at distinguishing between generated samples $G_{BA}(A)$ and real samples B. G_{BA} work to minimize this objective against an adversary D_B that tries to maximize it. On the other hand, the second generator G_{AB} which downsamples images from B to A, we solve a similar adversarial loss function $\mathcal{L}_{adv}(G_{AB}, D_B, B, A)$:

$$\begin{aligned} \mathcal{L}_{adv}(G_{AB}, D_B, B, A) = E_{A \sim p_{data}(A)}[\log D_A(A)] + \\ E_{B \sim p_{data}(B)}[\log(1 - D_A(G_{AB}(B)))] \end{aligned} \tag{3.14}$$

3.4.2.2 Cycle consistency loss

Theoretically, the adversarial loss can control the learning mapping of generators to produce outputs identically distributed as target domains [67]. However, Zhu et al. [250] show that using only the adversarial loss cannot guarantee that the learned function will map an individual input to output in a large capacity network.

CycleGAN introduced the cycle consistency loss intending to reduce the space of possible mapping functions. For each generated images $G_{BA}(A)$ from A, the generating cycle must be able to reconstruct it back to the original image such that $G_{AB}(G_{BA}(A)) \approx A$. Thus, in the CycleGAN model, for each generating process between classes, there exists two consistent cycles to secure the generating process between domains:

$$\begin{aligned}\mathcal{L}_{cycle}(G_{BA}, G_{AB}) = E_{A\sim p_{data}(A)}[\|G_{AB}(G_{BA}(A)) - A\|_1] \\ + E_{B\sim p_{data}(B)}[\|G_{BA}(G_{AB}(B)) - B\|_1]\end{aligned} \tag{3.15}$$

Cycle consistency is the key for a training that does not require pairs of input and ground-truth data. It prevents generators from producing images that do not relate to the output in an unpaired training, while the comparison of adversarial functions ensures the procedure with unpaired data.

3.4.3 Training procedure

As can be seen from Figure 3.9, the workflow of CycleGAN contains two cycles: the forward cycle that produces images from domain A to B and the backward cycle that produce images from domain B to domain A. CycleGAN is symmetric in which each flow is similar but in reversed direction. In the forward cycle, examples of domain A are firstly fed to the generator G_{AB} to produce a sample that supposes to belong to domain B, then it is evaluated by the discriminator $D(B)$ to see if it looks real in domain B, and finally translated back to domain A using G_{BA} to measure the cyclic loss. The cycle consistency check implies of transformation between real samples of each domains and synthesized ones which should remain intact and be recoverable. Details of CycleGAN training algorithms are shown in 4.

3.4.4 Generator architecture

Current architectures of generators for GAN-based methods in general, or CycleGAN in particular, are mostly based on convolutional neural networks which contain convolutional layers rather than fully connected layers as a feed-forward network. Among different types of CNNs, U-net and Residual network (ResNet) are the base architectures that are currently used to solve image translation tasks. The following paragraphs will present the architecture of these methods as the generator of CycleGAN.

3.4.4.1 U-net

U-net [170], based on the fully convolutional network, was introduced for biomedical image segmentation. Similar to VAE, U-net itself has parts referred to as encoder and decoder. A general

Algorithm 4 CycleGAN training algorithm

for each training iteration **do**

- Train the discriminator:
  1. Take a mini-batch of random images from each domain A and B.
  2. Use the generator G_{AB} to translate images A to domain B and vice versa with G_{BA}.
  3. Compute $D_A(A)$ and $D_A(G_{BA}(B))$ to get the losses for real images in A and translated images from B, respectively. Then add these two losses together.
  4. Compute $D_B(B)$ and $D_B(G_{AB}(A)$ to get the losses for real images in B and translated images from A, respectively. Then add these two losses together.
  5. Compute total loss for discriminator based on $D_A(A)$ and $D_B(B)$.
- Train the generator:
  1. Input the images from domain A and B.
  2. Compute the validity of A $D_A(G_{BA}(B))$ and the validity of B $D_B(G_{AB}(A))$
  3. Compute the reconstructed of A $G_{BA}(G_{AB}(A))$ and reconstructed of B $G_{AB}(G_{BA}(B))$
  4. Compute the identity mapping of A $G_{BA}(A)$, and identity mapping of B $G_{AB}(B)$
  5. Update the parameters of both generators inline with the cycle-consistency loss, identity loss, and adversarial loss.

end for

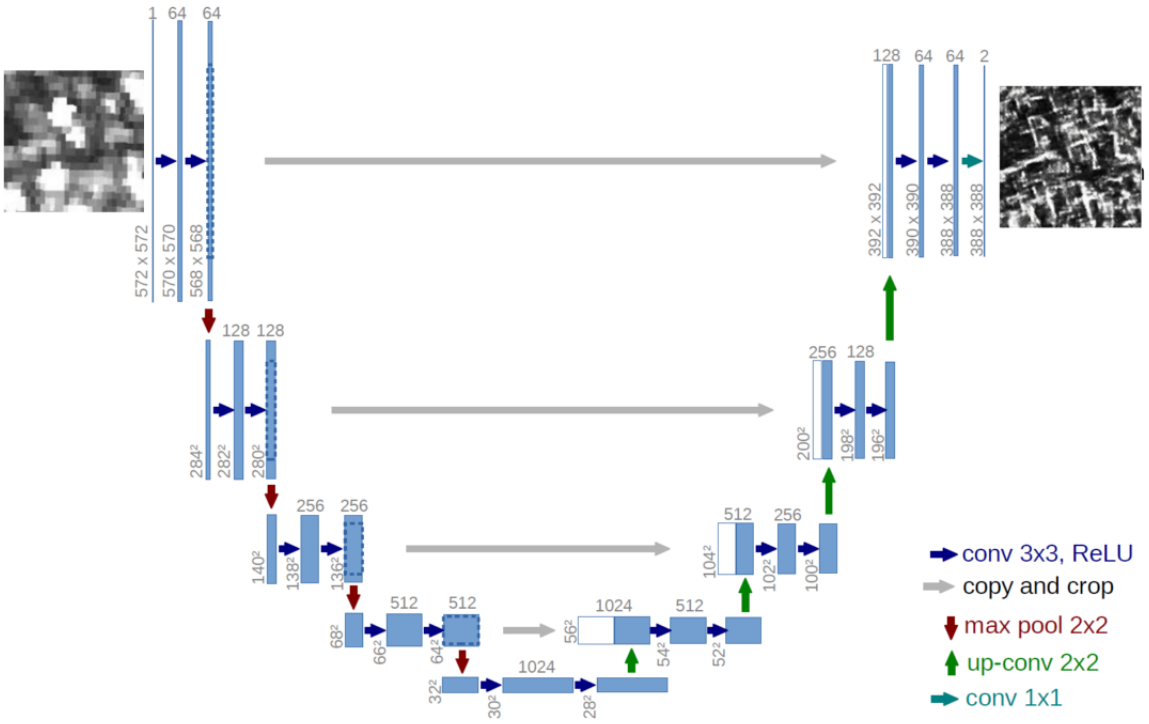


Figure 3.10: The original U-net architecture [170]. It contains a contraction path and expanding path The contraction and expanding paths are sometimes referred to as encoder and decoder, respectively.

network consists of two halves: the downsampling half, where input is compressed spatially but increased depth, and an upsampling half, where representations are expanded spatially while the

depth is reduced. The overall architecture of U-net has been successfully applied in several tasks of computer vision with better performance instead of only segmentation tasks. The key idea is to focus on classification and understanding large regions, including higher resolution skip connections, to preserve the detail that can then be accurately segmented, although it is compressed.

Unlike the linear structure of VAE, in which data flows through the network from input to output, one layer after another, U-net uses skip connections that allow information to shortcut parts of the network and flow through to later layers. For classification models, a final fully connected layer is enough to output the probability of a particular class being present in the image. However, it is critical to upsample feature maps back to the input size for image segmentation and synthesis tasks without losing information. Through the downscaling and the subsequent upscaling, U-net compresses the image to capture the most meaningful representation but, at the same time, can add back all the detail. Feature maps learn contextual understanding of what input is while reducing information about where it is located.

Figure 3.10 shows the original architecture of U-net generator. U-net contains convolutional layers to encode information for data and transpose convolutional (deconvolutional) layers to upsampling feature maps. Similar to CNN-based methods, these layers are followed by activation functions such as ReLU and normalization functions.

The concatenated layers join a set of layers together along a particular axis. In the U-net, concatenated layers connect upsampling layers to the equivalently sized layer in the downsampling parts. The layers are joined together along the depth dimension, while the number of feature map sizes remains the same. There are no weights to be learned in a concatenated layer. Besides, CycleGAN uses instance normalization layers [209] rather than batch normalization layers [84], which in style transfer problems can lead to more satisfying results. The instance normalization normalizes every single observation individually rather than as a batch. Unlike batch normalization, it does not require mean and stand deviation parameters to be calculated as a running average during training since at test time, the layer can normalize per instance in the same way as it does at train time. The means and standard deviations used to normalize each layer are calculated per channel and observation. Also, the instance normalization layers do not need to learn weights since there are no scaling or shifting parameters like batch normalization.

3.4.4.2 ResNet

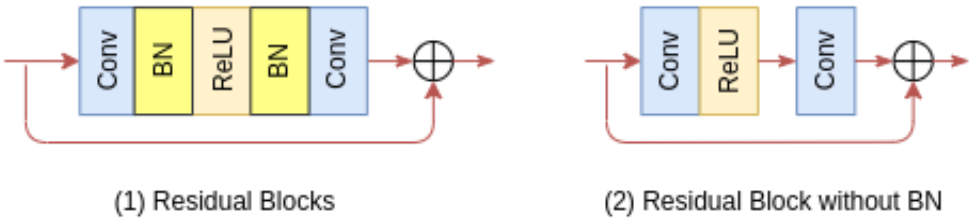


Figure 3.11: A single residual block. (1) generic residual block (2) residual block with removal of normalization

Along with U-net, residual network (ResNet) [75] is a popular generative architecture used for the generative model. The ResNet architecture allows information from previous layers in the network to skip ahead of one or more layers. However, instead of using a U-shape design to connect layers from the downsampling of the network to corresponding upsampling, ResNet uses

a series of stacked residual blocks where each block contains a skip connection that sums the input and output of the block before passing this on to the next layer. ResNet for the generator also contains downsampling and upsampling layers. The overall architecture of the ResNet is shown in Figure 3.12

Since AlexNet [109], the state-of-the-art CNN architecture is going deeper and deeper. However, training a deep neural network with many layers is difficult due to the vanishing gradient problem [189], where the repeated multiplication may make the gradient infinitely small during back-propagation. When the network goes deeper, its performance gets saturated or degrades rapidly. ResNet has solved this issue by introducing an identity shortcut connection that skips one or more layers. When the error gradients can backpropagate freely through the network through the skip connections that are part of the residual blocks [75]. Adding additional layers never degrades the network performance because the skip connections ensure that it is always possible to pass through the identity mapping from the previous layer if no additional informative features can be extracted. The shortcut connection reduce not only the number of network parameter, but also the computational complexity of algorithms. Figure 3.11 presents the original residual blocks and the residual blocks with removal of normalization layers.

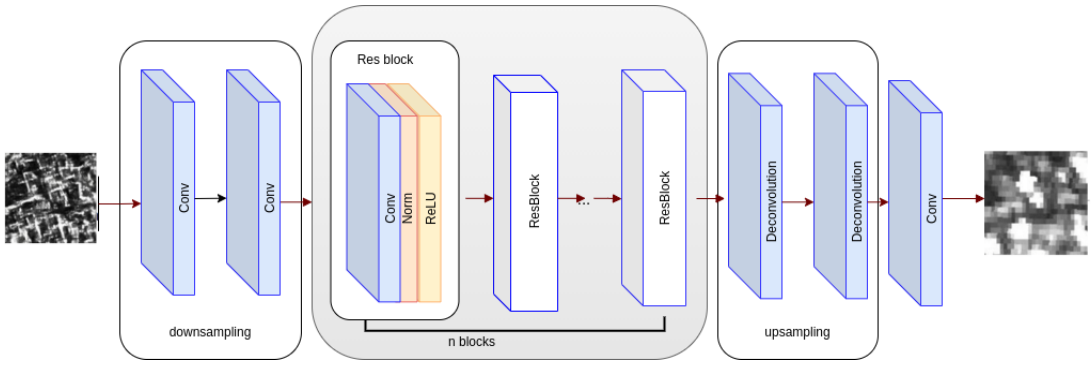


Figure 3.12: U-net generator architecture.It contains a contraction path and expanding path The contraction and expanding paths are sometimes referred to as encoder and decoder, respectively.

3.4.5 Discriminator architecture

The goal of discriminators is to produce the probability of a given input to classify whether it is real or not. The CycleGAN discriminator is based on the PatchGAN [85] architecture, where the discriminator divides the input into square overlapping "patches" and classifies each patch as real or fake, rather than predicting the whole input. Therefore the output of the discriminator is a tensor that contains the predicted probability for each patch, rather than just a single number. Patches are predicted simultaneously when input flow through the network. Input is not divided up the image manually and passes each patch through the network. The division of the image into patches arises naturally as a result of the convolutional discriminator architecture. It allows the design of the CycleGAN to be fully convolutional, meaning that it can scale relatively easily to higher resolutions. Other than that, the discriminator of CycleGAN is still a relatively straightforward implementation.

Using a PatchGAN discriminator is that the loss function can then measure how good the discriminator is at distinguishing images based on their style rather than their content. Since each element of the discriminator prediction is based only on a small square of the image, it must use

the style of the patch rather than its content to make its decision.

3.5 Conclusion

This section has provided fundamental knowledge of the generative models in autoencoder and generative adversarial neural networks. Autoencoders are fundamental generative models to map high-dimensional input into a low-dimensional latent space so that high-level features can be extracted from raw input. However, there are still drawbacks to using plain autoencoders as a generative model sampling. Variational autoencoders solve these problems by introducing randomness into the model and constraining how points in the latent space are distributed.

In terms of generative adversarial networks, all GANs are characterized by a generator versus discriminator architecture, with the discriminator trying to classify between real and fake images and the generator aiming to fool the discriminator. By balancing the adversarial training of two networks, the final model can gradually learn to produce similar observations to those in the training set. Overall, the GAN framework is highly flexible and able to be adapted to many exciting problem domains.

The last part is the CycleGAN - a generative model used for image translation and style transfer. The CycleGAN methodology allows training a model to translate images from one domain to another. Hence, the potential of CycleGAN is enormous, which allows us to apply it to different topics related to medical images. Crucially, CycleGAN does not require paired images from each domain to implement, making it a powerful and flexible technique.

Chapter 4

MRI super-resolution

The work presented in this chapter lead to the following publications:

[1] Do, H., Bourdon, P., Helbert, D., Naudin, M., Guillevin, R. (2021). 7T MRI super-resolution with Generative Adversarial Network. Electronic Imaging, 2021(18), 106-1.

[2] Do, H., Helbert, D., Bourdon, P., Naudin, M., Guillevin, C., Guillevin, R. (2021, October). MRI super-resolution using 3D cycle-consistent generative adversarial network. In 2021 Sixth International Conference on Advances in Biomedical Engineering (ICABME) (pp. 85-88). IEEE.

4.1 Introduction

Magnetic resonance imaging (MRI) is widely used in medical imaging because it provides a noninvasive assessment of the anatomy and physiology of the body in health and disease while offering the best contrast resolution for soft tissue. High-quality MRI is preferred in clinical centers and research settings because it can provide important structural details with a smaller voxel size, enabling accurate image analysis. At the same time, low-resolution MRI (LR) is plagued with noise and a lack of structural information. Therefore, the demand for image quality with sufficient detail in medical imaging is rapidly increasing. However, MRI images are usually acquired with limited resolution and low spatial coverage, which is limited by signal-to-noise ratio (SNR) or longscan time [158]. For example, a 3 Tesla (3T) MRI scanner may take 2 to 48 hours to produce a high-resolution (HR) result, depending on the clinicopathologic question and the size of the scanned area. Therefore, it is not easy to achieve the desired resolution with high-resolution MRI. Recently, the improvement of medical image quality has become an important issue of great value for both research and practice. Figure 4.1 illustrates an examples of LR and HR MRI. In the full view, the differences between low and high-resolution MRI are not clear but can be seen clearly in zoom-in view.

Single-image super-resolution (SR), an image quality enhancement technique, has become a potential post-processing method to increase the spatial resolution of medical scans after acquisition virtually. It produces high-resolution images from single- or multi-frame low-resolution images, using either explicit geometric regularity constraints or self-learned rules. In recent years, SR has been getting much attention from the community because of its benefits in medical image analysis [125]. Moreover, several studies has been proposed to enhance spartial resolution MRI. The variety of methods stretches from statistical method such as interpolation [121, 202], dictionary mapping

[126, 11], self-learning [136] to automatically techniques using neural networks [40, 239] or hybrid methods [162]. Specially, with advanced network architectures and training process, CNN-based super-resolution has achieved significant success on both objective (peak signal-to-noise ratio - PSNR) [127, 117, 100] and subjective (human visual quality assessment tests) [117, 177] criteria.

Existing methods in SR require paired datasets to implement. Generally, a training dataset for super-resolution contains pairs of low and high-resolution MRI to update loss value during training. Hence, these are usually not large enough due to paired data retrieval tedious and time-consuming task. However, as mentioned above, the scanning to produce HR MRI can take a very long time. Thus, it is not easy to obtain a paired medical dataset in the field of super-resolution. Currently, there does not exist a public dataset with paired MRI for the SR task.

Within the scope of the thesis, the work focuses on implementing a method that enhances the spatial resolution on routine 3T MRI to improve diagnosis and assessment. To address data availability problems, we propose a method that can implement either on paired/unpaired datasets. The proposed method is based on a cycle-consistent generative adversarial network (CycleGAN) [250] to benefit the unpaired training. As presented in Chapter 3, CycleGAN is a well-known method for image-to-image translation. However, different from the literature methods, we propose a novel architecture of generators to improve the performance of medical data.

With the support of MRI infrastructure at CHU Poitiers, we are able to explore the performance of the proposed method on practical data. Hence, for the super-resolution task, we aim to examine the performance of methods on both research and practical dataset, focusing on 3T and also 7T MRI.

Our main contributions are:

- We propose a unified framework based on CycleGAN for super-resolution. With a different generator from the generator of CycleGAN, the proposed model can perform the super-resolution on MRI through unpaired training. Besides, the method is implemented on both 2D and 3D MRI data.
- We do experiments on both research and practical MRI dataset to ensure the performance of the dataset. With training medical dataset has high-quality enough, the proposed methods can work on any available MRI dataset.
- Finally, we make a comparison of the proposed method with traditional super-resolution and other well-known GAN-based methods to highlight the performance.

In the next section 4.2, we present concepts and state-of-the-art image super-resolution with different CNN-based methods for medical images. Then we discuss the details of network architectures used in our experiments in Section 4.3 and 4.4. After that are the experimental results and comparison between methods; finally will be the discussion.

4.2 Related work

4.2.1 Image super-resolution

Super-resolution is a process that produces HR images from single- or multi-frame LR images. In the case of medical analysis, it becomes a potential solution as a post-processing technique

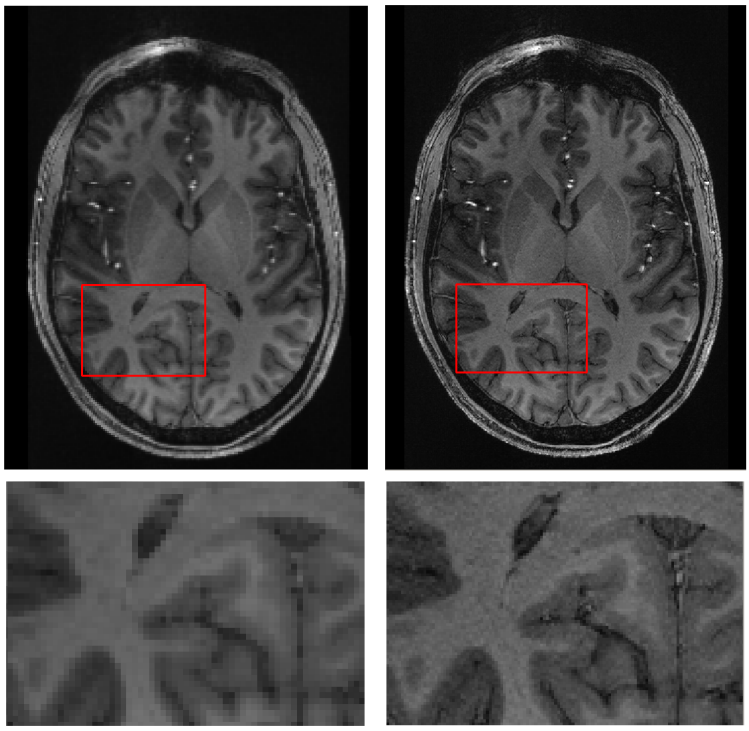


Figure 4.1: Examples of low and high-resolution MRI. From left to right: low-resolution MRI and high-resolution images, from top to bottom: full view and zoom-in view.

to improve the spatial resolution of MR images [213]. Figure 4.1 illustrates an examples of low and high-resolution MRI. In the beginning, SR is similar to an optimization problem to minimize the cost between observed LR image and regularization terms. However, statistical methods are limited by concepts of data representation. It results to non-robust methods, and the performance is not stable on images with great structural details.

Before deep learning-based approaches achieved state-of-the-art performance, super-resolution methods mostly relied on interpolation [190], edge-preservation [198], and dictionary learning [228]. regularization [95].

In general, interpolation-based methods, such as linear or nearest-neighbor interpolation, are simple and easy to implement. However, the effectiveness of basic interpolation is not high when the result usually contains artifacts, blurred sharp edges, or recovering fine details in complex textures [186]. Later, upsampling methods via patch-based non-local reconstruction have been proposed [137] to improve MRI image quality. On the other hand, regularization approaches such as manifold regularization [133], non-local similarity regularization [240], total variation regularization [208] are usually used for SR MRI. Both interpolation- and regularization-based SR methods mainly perform SR operations directly on the testing image using information only from the testing image itself [186].

Along with the development of neural networks in computer vision, Dong et al. [52] firsty proposed the super-resolution convolutional neural network (SRCNN). It contains three convolution layers for feature extraction, feature space building, and image reconstruction together in end-to-end training. After that, many follow-up approaches have been inspired with improvement on network structures [100, 206, 200]. Deep-learning-based studies have been proposed in

medical analysis to apply to MRI images. Dense network [40] full use of hierarchical features for reconstruction. Residual network [150] reconstructs 3D HR cardiac volume from multiple 2D LR slices.

4.2.2 Learning-based methods for super-resolution

Up to now, many deep learning-based methods have presented excellent performances in the field of super-resolution. The number of learning-based methods is fast increasing with different algorithms to improve model performance. The size of the complexity of models is significantly increasing, wherein the depth of models becomes a practical problem. Kim et al. [100] have proved that expanding the network architecture is the key to obtaining the high-quality super-resolution outputs. However, a deep network using advanced techniques to increase network depth could ease the difficulty of training. In addition, when the network is too deep, gradient disappearance and gradient explosion issues are declared [189]. Although data regulation and batch normalization can solve these gradient issues, they can lead to model performance degradation.

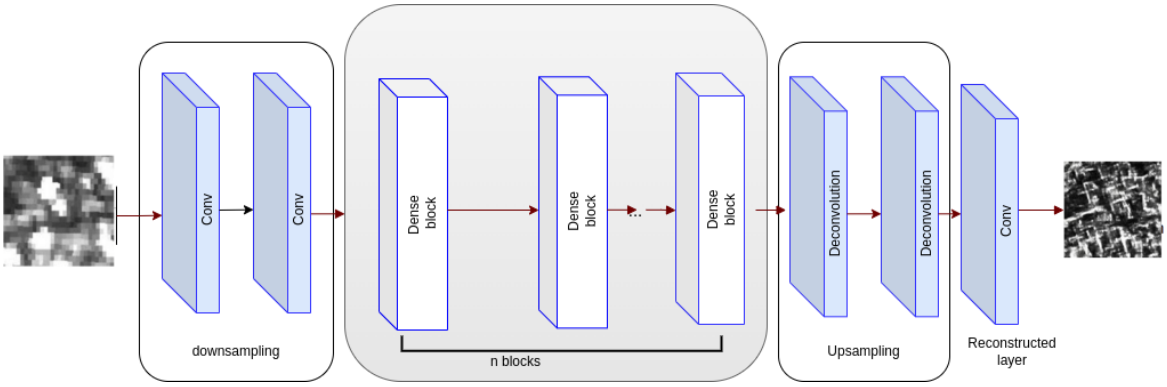


Figure 4.2: SRDenseNet architecture.

ResNet, proposed by He et al. [75], has addressed this problem by introducing the residual learning. The output of the previous convolutional layer is connected to the next for smoother information flows through a shortcut. It has been proved to neither increases the number of network parameters nor the computational complexity of algorithms. Kim et al. [100] presented a very deep convolutional network (VDSR) that uses an architecture composed of 20 layers. Advantage of ResNet, this method introduced residual learning in SR tasks and initialized a higher learning rate to accelerate the training process.

Besides, Wang et al. [221] introduced a method that combines deep learning terms and conventional sparse coding for super-resolution. It outperformed in SR fields, while the sparse coding algorithm contributes to training speed and model compactness. Shi et al. [188] proposed the efficient sub-pixel convolutional neural network (ESPCN). The network architecture consists of two convolution layers to extract feature information; then, the sub-pixel convolution layer aggregates the feature information and rearranges elements from low-resolution space to the output in high-resolution space. Later, the sub-pixel layer achieved superior super-resolution performance and became a popular and efficient upsampling method in many other studies. Laplacian Pyramid Super-Resolution Network (LapSRN) [113] is a method based on gradually reconstructing the sub-band residuals of high-resolution images. Feature maps of LR images are taken as the input of the next layer at each level of the pyramid, and then it predicts the high-frequency residuals.

DenseNet [82] proposed a connectivity pattern to improve the flow of feature information by concatenating previous layer information. The network is more efficient and presented an appreciable performance improvement compared to ResNet with fewer parameters. Later, recent studies in image super-resolution show that removing unnecessary batch normalization (BN) layers in residual blocks [220] and dense blocks [127] can reduce the computational cost, memory usage and boost model performance. The flow of the gradient is unobstructed due to the direct link between layers.

Inspired from DenseNet, Tong et al. [206] have proposed the densely connected network (SRDenseNet) that uses dense blocks with skip connections to solve the SR problems. The architecture of original SRDenseNet are shown in Figure 4.2. This architecture allows subsequent layers can effectively use extracted feature information from each convolutional layer; hence the information convention is preserved between different levels from different convolution layers in the network. Features of ground truth images are retained to a greater extent; therefore, the dense connection can effectively improve the quality of image reconstruction. SRDenseNet has significantly improved performance over the model using multi-level features, indicating that level fusion benefits SR problems. Figure 4.3 shows the architecture of residual and dense block, and later is blocks with removal of normalization layers.

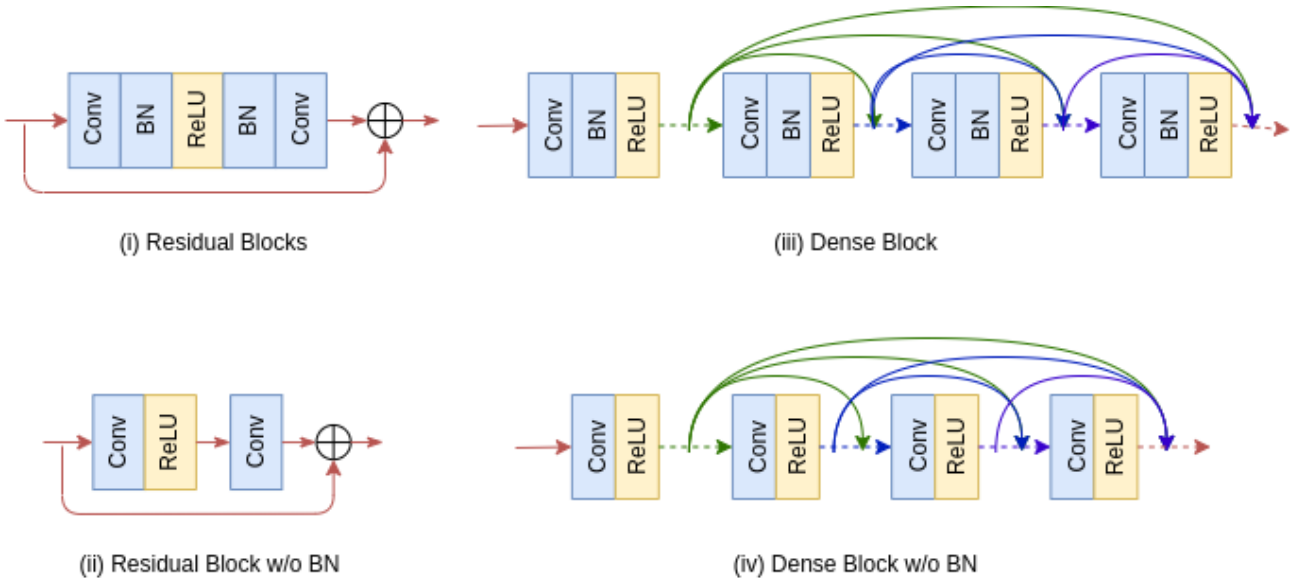


Figure 4.3: Residual blocks and dense blocks with skip connection.

Residual Dense Network (RDN) introduced by Zhang et al. [244] - a combination of DenseNet and ResNet to solve the image super-resolution. This network presented the residual dense blocks (RDB) that use the densely connected structures effectively utilize local features from convolution layers and the combination of dense skip connection and residual learning. Besides, the idea of fusing local and global features with residual learning makes full use of the features from the proceeding layers.

4.2.3 Generative adversarial networks for super-resolution

Since Goodfellow et al. [69] proposed the generative adversarial network model (GAN), many follow-up studies of GAN and its variations have been used in various computer vision tasks. In

general, GAN has demonstrated its efficiency in achieving top performance in image synthesis. Recently, GAN has also been used in the field of super-resolution. Super-resolution GAN (SRGAN) [117] and Enhanced super-resolution GAN (ESRGAN) [220] have been successfully applied to solve SR problems for color images. In addition, CycleGAN [250], proposed by Zhu et al. is an image-to-image translation model for learning how to translate an image from a source domain to a target domain in the absence of paired examples. In general, CycleGAN has been used in image synthesis and image translation domains. By taking advantage of CNN-based methods, we can also use CycleGAN to solve super-resolution in medical data. The advantage of this method is that it does not require paired images for efficient training, which is a challenge for high-resolution MRI.

Most previous super-resolution methods aim to optimize HR image reconstruction by minimizing the pixel-wise difference between original and generated images. However, one drawback is that reconstructing small, critical details is extremely difficult if one is only concerned with local, pixel-wise differences. In contrast, if global perceptual constraints can be taken into account, the SR model will be guided by both local intensity information and patchwise perceptual information, likely resulting in a better and sharper SR reconstruction [117].

With benefits from the GAN framework of Goodfellow et al. [69] for its unsupervised-learning potential of capturing perceptually essential image features, Ledig et al. [117] proposed the SR-GAN to handle the super-resolution issue. The adversarial loss is defined in the GAN model [69], and it also extends with perceptual loss in SRGAN. In perceptual loss concepts, this model used a pre-trained model for feature extraction and compared it with features of the generator to minimize loss of information. Besides, SRGAN defines the activation layers of a pre-trained deep network, where the distance between two activated features is minimized. Several techniques have been implemented to provide different building unit architectures to transform LR into HR images and reduce computation costs during the training phase. Residual blocks from ResNet and Dense blocks from DenseNet are the most popular architecture used in GAN-based methods for SR tasks because these units can be combined or modified to speed up the training process and improve model performance.

4.2.4 MRI super-resolution

Super-resolution on medical images is getting much attention from the community because of its benefits in practice [125]. Methods are applied on different types of medical images MRI, CT, PET, ultrasound [179, 39, 239]

In terms of MRI, many deep learning-based methods have been proposed to improve the spartial resolution. In general, digital MRIs are stored as types of 2D (slices) or 3D (volumes). Thus, deep learning methods applied to MRI can be both in 2D or 3D space. The following paragraphs will survey applying deep learning methods on MRI at both levels.

In terms of 2D MRI, there are many studies. Zeng et al. [238] proposed a method for simultaneously estimating single-contrast and multi-contrast MRI images. Single-contrast sub-network solves the super-resolution problem of low-resolution T2 images; the multi-contrast sub-network estimates multi-contrast T2 images based on the reference T1 images and T2 super-resolution images. HR images are MRI brain images, and LR images are simulation data. Shi et al. [187] used local residual block and global residual network to extend SRCNN to solve a 2D MRI SR problem. Zhao et al. [247] extended an SRCNN architecture for 2D MRI brain images. The

network consists of three main sub-networks: feature extraction sub-network, non-linear mapping sub-network, and reconstruction. The non-linear mapping sub-network comprises a set of cascaded channel splitting blocks where each block follows a merge-and-run strategy; it splits features into two branches, precisely, one for DenseNet, and the other for residual learning. Liu et al. proposed a multi-scale fusion convolution network (MFCN) [128]. This network has several multi-scale fusion units (MFU) in which each unit corresponds to an estimation obtained with filters at a specified scale. Oktay et al. introduced the T-L network [151] based on U-net architecture, which was used for image segmentation and super-resolution for 2D cardiac images.

In terms of 3D MRI super-resolution, Pham et al. in [157] attempted to perform SR on 3D MRI brain images with the SRCNN framework. The comparison between the networks trained on natural and MRI images was also given. Oktay et al. [150] proposed a 3D CNN based on residual learning for SR of cardiac images. The proposed network is modified from VDSR with the replacement of the interpolation operation by a deconvolution layer at the top of the network. Zhao et al. [245] presented an extended EDSR for MRI brain super-resolution. The EDSR-based network is trained with the paired low and high-resolution images in the axial direction. Then the low-resolution image in sagittal and coronal directions is reconstructed from the trained model. Chen et al. in [40] proposed a densely connected super-resolution network (DCSRN) for brain MRI images, similar to the DenseNet. The DCSRN is densely connected, while the output of each block will be reused in the latter blocks.

In general, a 3D model is preferable to fully solve the ill-posed super-resolution because it can directly extract 3D structural information. Recent studies [157, 40] demonstrated that a 3D CNN outperforms its 2D counterpart since it fully exploits the 3D volume information. The network architectures use 3D convolutions. With the additional dimension introduced by a 3D CNN, the number of parameters of the network overgrows. The performance of a deep network generally improves with more layers and weights, but with 3D, the model becomes computationally expensive. A densely connected network has efficient memory usage and is practical for 3D images.

4.3 Methodology

The core problem of super-resolution for medical images in the real world is the lack of paired data. The research objective is to reconstruct LR MRI volumes obtained under a down-sampled protocol to HR MRI volumes following a scaling factor s; through a training process that does not require paired data for analysis. In the scope of the work, we expect the final model to perform super-resolution on a low-resolution volume to a high-resolution volume following a scaling factor.

The proposed method is a generative model based on the cycle-consistent design to perform the super-resolution through paired/unpaired training. The hybrid model contains two different generators with changes from the original models to perform the up and down resolution. All the modification of resolution is only executed during the training process.

CycleGAN is a generative model used to perform efficient training with unpaired data, while the new generator responds for the super-resolution part.

4.3.1 Network architecture

4.3.1.1 Adversarial network architecture

As presented in the previous chapter, the general GAN contains a generator G and a discriminator D. Generators and discriminators work in pairs to solve the adversarial min-max problem. The generator aims at minimizing errors against a discriminator that tries to maximize them. The goal is to train the generator G to fool a differentiable discriminator D that is trained to distinguish generated high-resolution images from low-resolution images. However, existing GAN-based methods for super-resolution often requires paired LR and HR images to let the discriminator compare real and generated images. On the other hand, the core problem of super-resolution for medical images in the real world is the lack of paired data. Moreover, MRIs contain different or more complex spatial variations, correlations, and statistical properties than natural images, thus limiting the SR imaging performance of most traditional methods.

When Zhu et al. introduced CycleGAN [250] - an image-to-image translation framework using unpaired data, it has inspired many studies on different computer vision tasks and the potential for SR. In this case, the CycleGAN aims to translate input LR MRI volume into an HR MRI volume without requiring paired images during training. Following the fundamental CycleGAN, the network contains two generators to produce images between low and high-resolution MRI and two discriminators to predict real or generated data. Along with the adversarial loss from the GAN model, to prevent the generators from producing synthetic images that are irrelevant to the inputs, the cycle-consistency loss is used for two generators to force the synthesized images to be identical to their inputs for each class. This loss function prevents CycleGAN from requiring paired data for training. The details of loss functions and general architecture have been presented in Section 3.4.

4.3.1.2 Building blocks

The proposed generator uses Residual Dense Blocks (RDB) [244] as building units to extract feature information. Figure 4.4 illustrates the design of RDB in the super-resolution model by removing normalization layers. These blocks have been successfully applied in the fields of super-resolution [73, 127, 1]. A RDB contains several convolutional layers followed by a ReLU activation function [65] in continuous connection. RDB can reduce computational time and memory usage and speed up the training process by removing batch normalization from general residual and dense blocks.

RDB introduced the contiguous memory mechanism that allows access to information between block layers bypassing the state of the preceding block to layers of the current block [244]. The size of feature maps is increased through each convolutional layer with a specific growth rate to synthesize information from raw input. The high growth rate can further improve the performance of the network. In the end, feature output has information of all subsequent layers, which not only extracts dense local features but also preserves the feed-forward flow.

Moreover, by concatenating components of blocks, local features are synthesized from the states of preceding RDBs and whole layers in the current RDB. Then, a convolutional layer adaptively fuses the output information [244]. Local feature learning improves the preservation of information inside the network, combining the hierarchical features. Information of all layers is fully used with RDBs to obtain dense features.

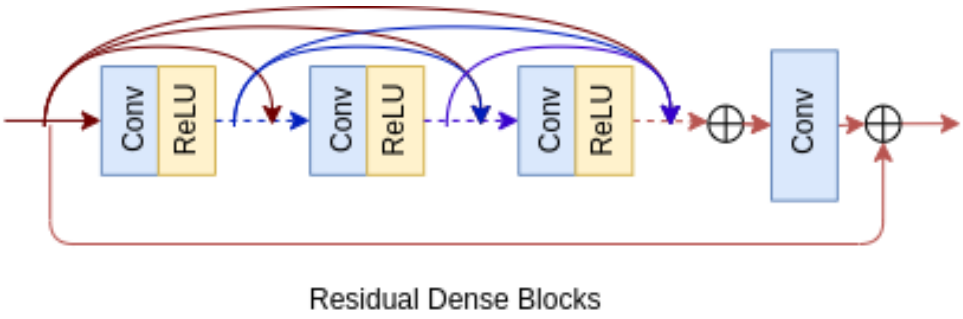


Figure 4.4: Residual dense block architecture.

4.3.1.3 Generative network architecture

Different from the generic CycleGAN [250], generators of the proposed model for super-resolution is modified with the addition of several RDBs for feature extraction. In the beginning, shallow features are produced from input through two single convolutional layers in order to reduce input size for the following computation. Then, local features from RDBs are extracted and concatenated to form global features. The number of features is increased through each layer with a fixed growth rate to synthesize information. Then, the information of all layers is preserved based on local and global features fusion mechanism. Next, global features from RDBs are stacked to fully use features from all the preceding layers.

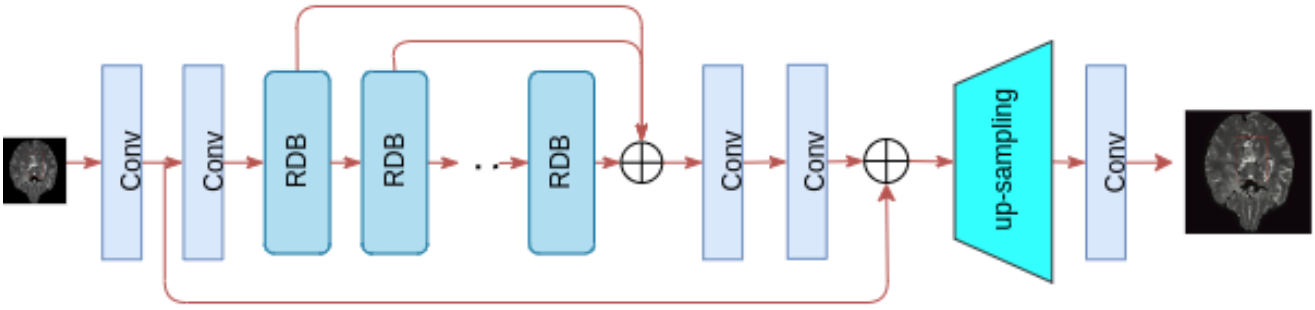


Figure 4.5: Generator architecture.

After the feature extraction process, a bottleneck convolutional layer is applied to adaptively fuse hierarchical features, followed by a convolutional layer to extract additional features for global residual learning. Then, fused features are concatenated to obtain the dense feature. The final process is to upsample this dense feature to form the desired output.

The detail of super-resolution generator is shown in Figure 4.5. Within the scope of the work, we implement the proposed methods on both 2D and 3D space in order to compare the model performance. In general, the architecture of the 2D and 3D models are similar for feature extraction. However, the main difference lies in the upsampling part, which is the most critical process. A model working well on 2D volumes cannot be ensured to work well in 3D due to the data structure and uniformity.

For the 2D SR model, the upsampling operator uses the sub-pixel function from ESPCNN.

Introduced by Shi et al. [188], it is an efficient way to upscale resolution for 2D images. It has proved its potential and has been used in many state-of-the-art in SR field [117, 127]. The sub-pixel function consists of a convolutional layer followed by a shuffler function to arrange pixels from input into an output of with the desired size. The sub-pixel can be represented as a standard convolution in a low-resolution space followed by a periodic shuffling operation. However, during our experiments, we have found that the sub-pixel function does not work stable on our 3D MRI.

For the 3D model, we implement two types of upscaling. The first approach uses a linear interpolation layer in 3D space to dense upscale features before the final convolutional layer forms the HR output. For the second solution, we use deconvolutional layers to perform the upsample operators for dense features. By striding a set of kernels over s steps, the element of a feature-map is rearranged to form consecutive pixels in the HR space. The advantage of this approach over static upsampling is that it contains learnable parameters with the same computational complexity, improving model performance. However, since consecutive pixels depend on different feature maps that are independently randomly initialized, it might lead to new artifacts in the HR output.

Although generating an image from high-resolution to low-resolution has no value for research, it is a part of the network for unpaired training. The architecture of the generator for downgraded is similar to the first one. To down-sample features, we use convolutional/ pooling layers as downscale operators to reconstruct LR from HR MRI for the downsample generator.

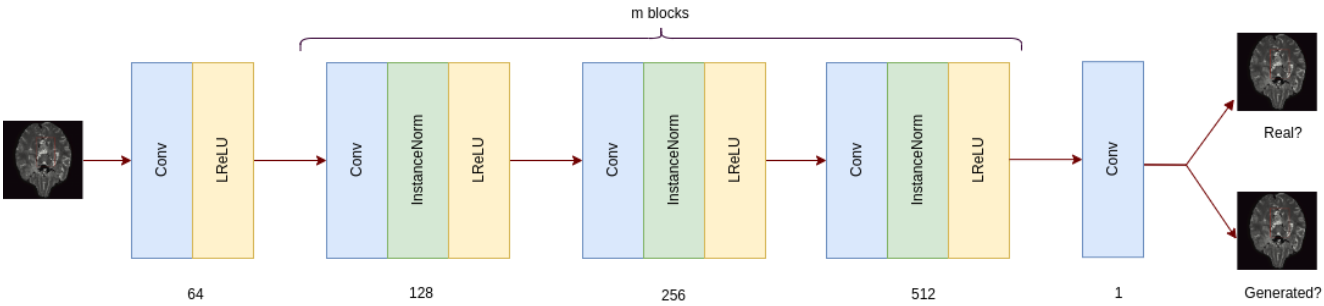


Figure 4.6: Architecture of discriminator. It contains several convolutional layers followed by Instance Normalization and Leaky ReLU.

The structure of discriminators is shown in Figure 4.6. Because the main work of discriminators is to classify the input and generated output, the architecture of the two discriminators are similar.

It is a CNN with multiple convolutional layers followed by instance normalization layer, and Leaky ReLU (LReLU) activation to synthesize information from given input. The network depth is modifiable based on the number of building blocks within the network. The last convolutional layer with a single output channel is used to generate values ranging from 0 (the generated image) to 1 (real MRI). In addition, we provide the option to predict using the sigmoid function with the convolutional layer.

4.4 Experiments

4.4.1 Dataset

4.4.1.1 Imaging dataset

As presented above, we want to examine the performance of the proposed method on both research and practical data. For this task, we firstly focus on super-resolution on T1 and T2-weighted MRI - the most common MRI sequences and then extend to different types of sequences. Thus, only T1 and T2-weighted MRIs are processed as training data.

We use the MICCAI BraTS 2018 [140, 13, 12], a dataset containing 3T MRI volumes with different types of sequences as research data. The variety within BraTS is ensured by samples acquired with various clinical protocols and scanners from multiple institutions. It contains 285 subjects in the training set and 59 subjects in the validation set, including T1-weighted, T2-weighted, post-contrast T1-weighted and T2-FLAIR. MRI volumes are pre-processed by co-registration to the same anatomical template, interpolation to the same spacing at 1mm^3 and skull-stripping. The field of view on each volumes is $155 \times 240 \times 240$.

With the support of Siemens Healthineers MRI devices at Poitiers University Hospital, we have built a practical dataset (appearing later as the *CHU dataset*), including both 3T and 7T MRI. A total of 46 3D MP-RAGE brain MRI subjects, T1-weighted for 3T MRI at different slice spacing, are collected. 3T MRI volumes were acquired from a Siemens Magnetom Skyra scanner, including 18 samples at 0.9mm^3 voxel spacing with field of view $240 \times 288 \times 192$ and 28 samples at 0.6mm^3 with field of view $336 \times 416 \times 448$.

In addition, we also want to find out how the super-resolution model works on 7T MRI along with 3T volumes and take advantage of the Siemens Magnetom Terra scanner at CHU Poitiers. These 7T volumes were recently extracted in late 2021 when the COVID-19 situation became more stable, and the 7T machine was put into operation. 7T MRIs in CHU dataset consists in 61 samples at 0.5mm^3 voxel spacing with field of view is $448 \times 448 \times 320$ and 20 samples at 0.75mm^3 with size $340 \times 340 \times 240$.

4.4.1.2 Data pre-processing

Super-resolution requires both low and high-resolution MRI for training. Based on studies of Rueda et al. [173], with a MRI dataset with high enough quality, low-resolution MRI can be obtained with minimal loss and reduced appearance of new artifacts through the degradation process [173]. For both the BraTS and CHU datasets, as in [185], LR volumes are downgraded from MRI volumes at original resolution through down-sampling operators at isotropic scaling factors with Gaussian filters to avoid aliasing artifacts.

4.4.2 Training setup

Two generators are built with slightly different components to perform the up-sample and down-sample operators from our design. The upsampling operators include two different methods: linear upsampler (SRCycleGAN) and deconvolutional layer (SRCycleGAN+DC). While the first operator involves static re-sampling, up-sampling using DC is more flexible with learnable parameters

that will be optimized during the training phase. We consider both model configurations deserve performance assessment. Each generator contains three RDBs, where each block includes three dense blocks in residual connection.

The complexity of GAN models remains a considerable problem with millions of parameters and increase along with model depth or the input size. To reduce the computational cost of the model, the model is trained on patches. The ratio between high- and low-resolution patches is defined as scaling factor s. This work trains separate single-scale networks with $s = 2$ and $s = 4$ scaling factors. To ensure the trade-off between data size and model performance, for each batch, patches are randomly extracted into a maximum size of $64 \times 64 \times 64$ patches on HR volumes corresponding to a size of $64/s \times 64/s \times 64/s$ on LR volumes. In our experiments, we choose the maximum number of patches on each sample as 15 to secure the diversity of data and avoid patch duplication in the training phase. The size and number of patches are modifiable arguments.

The batch size is set to 2. The learning rate is initialized to $1e^{-4}$, and decay starts after every 20 epochs. The ADAM optimizer [102] is used to update network weights based on training data. With this configuration, the training of SRCycleGAN takes an average learning time of 10 hours with a GPU NVIDIA A100 40GB for 200 epochs.

4.4.3 Evaluation metrics

To evaluate the image quality between ground-truth and output MRI volumes in both tasks, we use peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) metrics. The PSNR evaluates the ratio between the maximum power of an input image and the power of features that distort the image. On the other hand, SSIM evaluates perceptual image quality instead of calculating pixel-wise variations. Two images can be considered similar if they show the highest PSNR and SSIM.

In MRI super-resolution, we compare the performance of two states of the proposed 3D model (SRCycleGAN and SRCycleGAN+DC), the 2D SRCycleGAN to tricubic interpolation, Enhanced super-resolution GAN (ESRGAN) [220] for measurement purposes.

ESRGAN

ESRGAN, proposed by Wang et al. [220], is an improved version of SRGAN to increase the resolution of images. ESRGAN introduced residual-in-residual dense blocks (RRDB) without batch normalization (BN) layers as buidling units instead of residual blocks of SRGAN. Details of the generators with the additions of RRDB are shown in Figure 4.7

RRDB is a combination of residual blocks with multiple levels and dense block connections [127]. Recent research has demonstrated that removing BN layers can reduce computational cost and memory usage, and improve model performance. Despite the fact that BN layers use mean and variance computations to normalize the features during training and testing, BN layers have a tendency to produce undesirable artifacts and limit generalization ability when the difference between the training and testing sets is substantial [218].

Along with adding layers and connections to improve model performance, RRDB exploits deeper and more complex connections than the residual blocks of SRGAN. Meanwhile, the general high architecture of the ESRGAN model is kept as SRGAN. The discriminator of ESRGAN is also improved from SRGAN, based on the relativistic GAN [94]. Instead of estimating the probability

of an input being real or natural, the relativistic discriminator calculates the probability of a real image to be relatively more realistic than a fake image [218].

The ESRGAN is built only for 2D data. Besides, ESRGAN uses features before the activation layers, which helps overcome the drawbacks of the original design. ESRGAN also used the perceptual loss from a VGG pre-trained model during training to improve model performance. The perceptual loss in SRGAN can cause inconsistent reconstructed brightness compared with the ground-truth image or the sparsely activate of features when the deep network is intense. A recent study [218] based on ESRGAN for MRI data has proved the potential of this method on super-resolution images. In this work, we re-implement the ESRGAN from the original as a reference to compare the performance of GAN-based model.

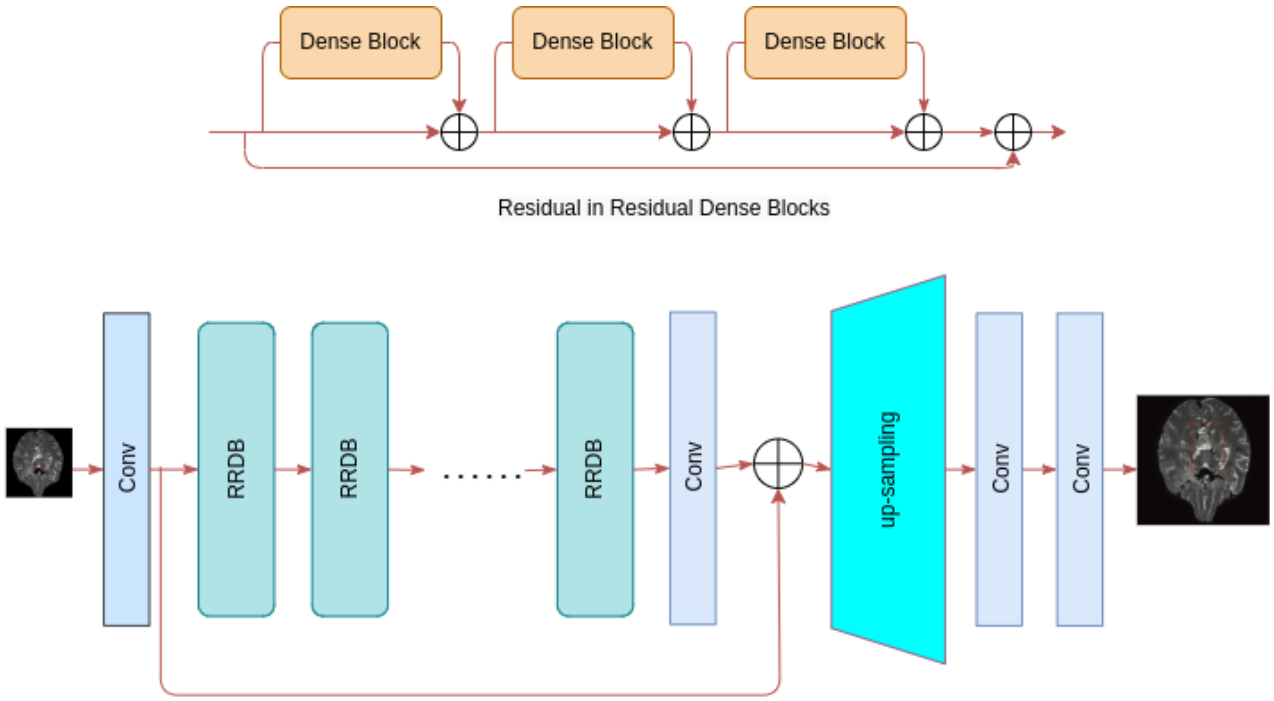


Figure 4.7: ESRGAN architecture.

Overall, we evaluate the super-resolution on routine clinical data (CHU dataset) and research data (BraTS test set) to examine model performance. Super-resolution enhances the structural information of the low-resolution 3T MRI input. For each ground-truth in the test set, there are corresponding reconstructed MRIs at the exact resolution, which are generated from 2D and 3D SR-CycleGAN, ESRGAN, and interpolation methods. Besides, there is also a test of model performance on 7T MRI.

		Tricubic		ESRGAN		2D SRCycleGAN		SRCycleGAN		SRCycleGAN+DC	
	Scale	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
BraTS	2x	26.24	0.7254	40.12	0.8409	44.57	0.8327	48.75	0.8782	**51.22**	**0.8825**
	4x	22.26	0.4782	35.85	0.6820	36.76	0.6441	34.78	0.7168	**37.82**	**0.7283**
CHU 3T	2x	26.07	0.7397	43.89	0.8220	42.31	0.8254	48.56	0.8606	**48.86**	**0.8749**
0.9mm	4x	22.20	0.4916	30.60	0.6533	30.48	0.6594	33.09	0.7022	**34.24**	**0.7103**
CHU 3T	2x	26.65	0.7300	44.17	0.8464	44.26	0.8492	49.76	0.8803	**49.86**	**0.8902**
0.6mm	4x	22.75	0.4953	30.27	0.6679	31.59	0.6806	35.07	0.7227	**35.18**	**0.7327**
CHU 7T	2x	25.67	0.7347	42.75	0.8443	44.68	0.8399	49.27	0.8758	**49.75**	**0.8815**
0.75 mm	4x	22.67	0.5050	31.33	0.6586	30.85	0.6742	34.21	**0.7212**	**34.67**	0.7187
CHU 7T	2x	26.32	0.7467	44.30	0.8577	46.01	0.8637	48.84	0.8980	**51.48**	**0.9028**
0.5mm	4x	23.03	0.5116	31.09	0.6935	31.48	0.6976	35.84	0.7309	**36.42**	**0.7455**

PSNR, peak signal-to-noise ratio; SSIM, structural similarity index.

Table 4.1: Average value of PSNR (dB) and SSIM for scale factors $\times 2$ and $\times 4$ on BraTS and CHU MRI dataset. Resolution-enhancement methods (tricubic interpolation, ESRGAN, 2D SRCycleGAN, 3D SRCycleGAN and 3D SRCycleGAN+DC) were compared with ground-truth images for quantitative evaluation.

4.5 Results

Table 4.1, Figure 4.8, Figure 4.9 show the quantitative results obtained by all tested methods on the different types of MRI of BraTS and CHU datasets. Image quality measurements in terms of PSNR and SSIM show that our 3D SR-CycleGAN method is able to achieve the lowest distortion for $s = 2$ and $s = 4$ scale factors. The SSIM values on reconstructed images are higher than other GAN-based and interpolation methods, indicating that our method maintains optimal perceptual quality compared with ground-truth images. PSNR values also indicate that SR-CycleGAN outperforms the other methods for objective quality measurement.

Both versions of our 3D CycleGAN provide an outperformed super-resolution result with detailed textures compared to 2D models and tricubic interpolation. For example, SSIM scores exceed 0.88 on 7T MRI at 0.75 mm at $2\times$ scale, ensuring a better fidelity of image structures such as contours or fine details compared to the original volume shown in visual quality.

Figure 4.10 illustrates model performance on T1-weighted 3T and 7T MRI from the CHU dataset by displaying a selection of data slices. Both versions of our SRCycleGAN model maintain fine visual structures compared to ground truth data. Zoomed-in views corresponding to the red boxes on whole-brain images also highlight the superiority of the 3D SRCycleGAN model configured with deconvolutional layers compared to the one using static up-sampling, which contains more blur and brightness/contrast fidelity issues. Object in higher-zoom (yellow circles) is finely reconstructed from the low-resolution MRI compared to the ground truth, without crashing of

voxels.

4.6 Discussion

As discussed in the previous section, 3D CycleGAN demonstrates higher performance compared to other methods of 3D MRI super-resolution. This demonstration in terms of SSIM values is particularly important, as SSIM computation approaches subjective visual quality metrics derived from the human visual system.

It is critical to analyze network performance. In interpolation, high-resolution features to be resolved are usually treated as low-order representations of existing low-resolution features. Although this is an acceptable assumption for natural images, it may not be valid for medical images due to the complex morphology, textures, and low SNR of tissue. In addition, such interpolation methods consistently perform the interpolation over a certain length scale (usually a few pixels around the pixel to be interpolated). Therefore, they are also not ideal for medical images, especially for MRI of the brain, since its structure and components are very complex. As a result, there is no way to distinguish whether a particular pixel truly represents the anatomy or whether that pixel is affected by noise or artifacts such as motion or aliasing.

In terms of ESRGAN, it is a well-known GAN-based super-resolution method for natural images. One advantage of ESRGAN is the outbreak architecture, which uses a new architecture of building units to improve model performance while reducing training costs. However, recent studies [218] have presented its drawbacks for medical images and proposed improvements for dealing with MRI. The efficiency of ESRGAN also comes from the improvement of the loss function. The original version proposed a perceptual loss based on using features extracted from the pre-trained model (VGG). In the case of medical data, there are no standard pre-trained models for this task, while using models such as VGG does not optimize.

We also observe that the performance of the model is limited in terms of blur for small objects, although the reconstructed images are quite detailed. However, we consider the CHU dataset a practical problem, and the reconstructed images are relatively impressive. Besides, we have to focus on the quality of input. In general, when the resolution of input is not good enough, the overall degradation can lead to the lack of information on objects that later happen to be only partially reconstructed in the output [125].

Although our proposed method can achieve compelling results in many cases, some limitations currently exist in the output. In the case of CHU data, we realize that the brightness of reconstructed images is slightly higher than the ground truth. It is automatically changed during the super-resolution process without any interference. Unlike images that are 8-bit color, medical images are usually 16-bit color. Because of the larger tone color, the value of voxels in medical data on the reconstructed image might be slightly different from the ground truth, which leads to this issue.

To optimize the performance and complexity of the model, training is performed on patches to reduce the computational cost of the model and to ensure the diversity of the data. The adversarial and cycle consistency losses are retained for the primary model. The generators are modified by adding several dense residual blocks. The number of blocks increases as the model depth or input size increases.

MRI volumes contain different or more complex spatial variations, correlations, and statistical

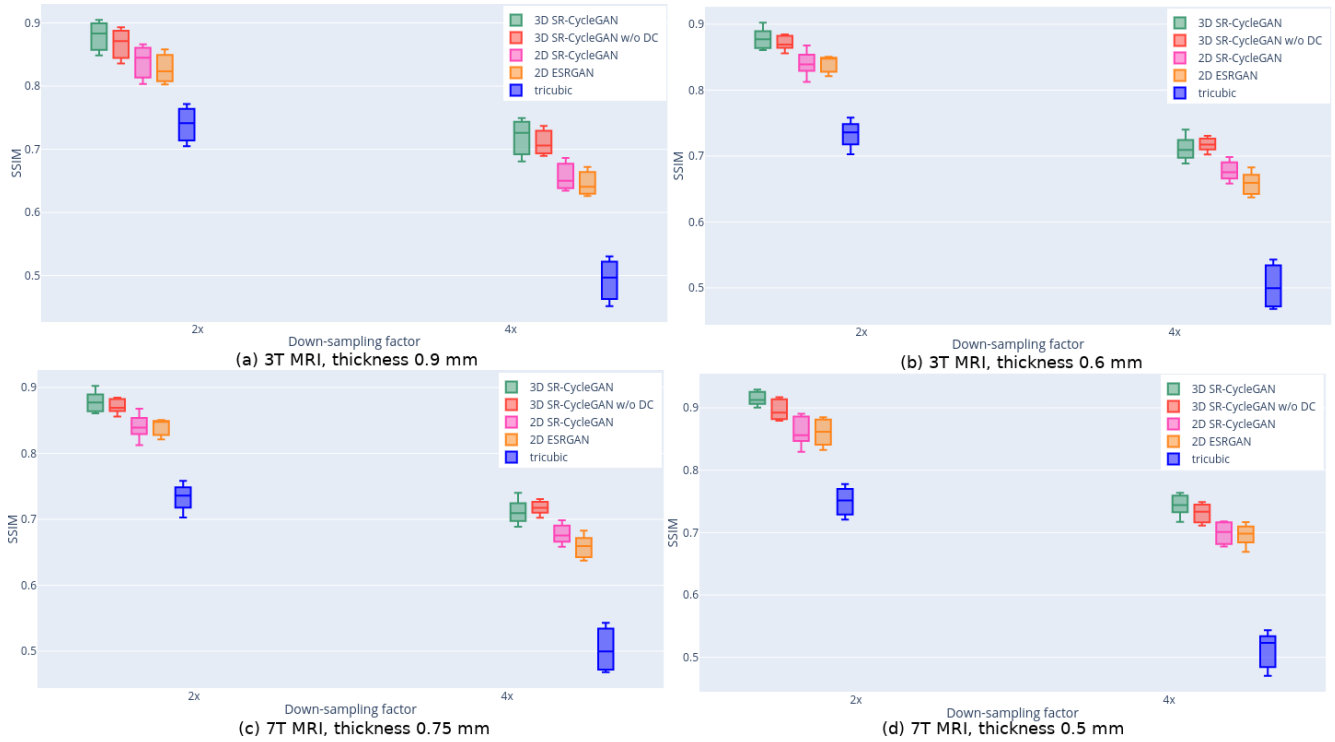


Figure 4.8: Comparison of model performance on the different ground-truth MRI for quantitative image similarity metrics using SSIM: (a) 3T MRI with 0.9 mm slice thickness, (b) 3T MRI with 0.6 mm slice thickness, (c) 7T MRI with 0.75 mm slice thickness, (d) 7T MRI with 0.5 mm slice thickness

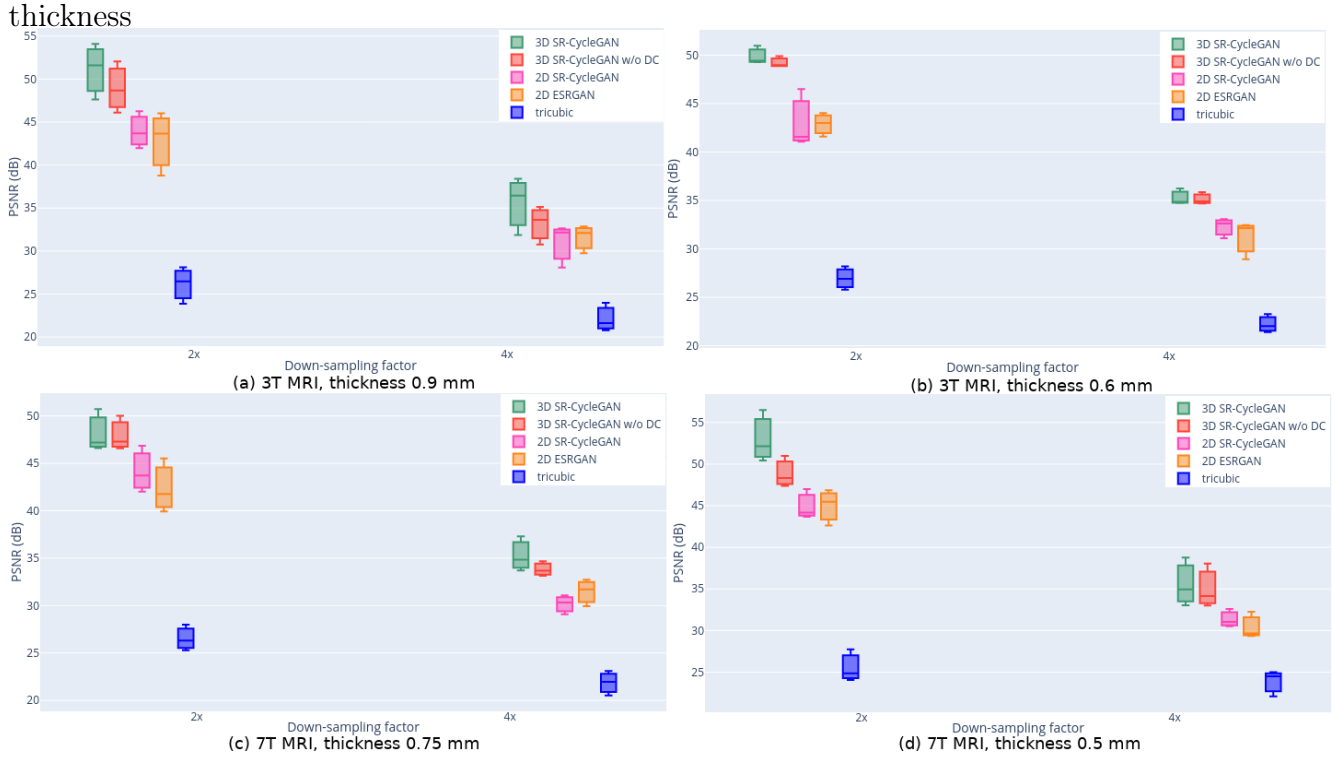


Figure 4.9: Comparison of model performance on the different ground-truth MRI for quantitative image similarity metrics using PSNR: (a) 3T MRI with 0.9 mm slice thickness, (b) 3T MRI with 0.6 mm slice thickness, (c) 7T MRI with 0.75 mm slice thickness, (d) 7T MRI with 0.5 mm slice thickness

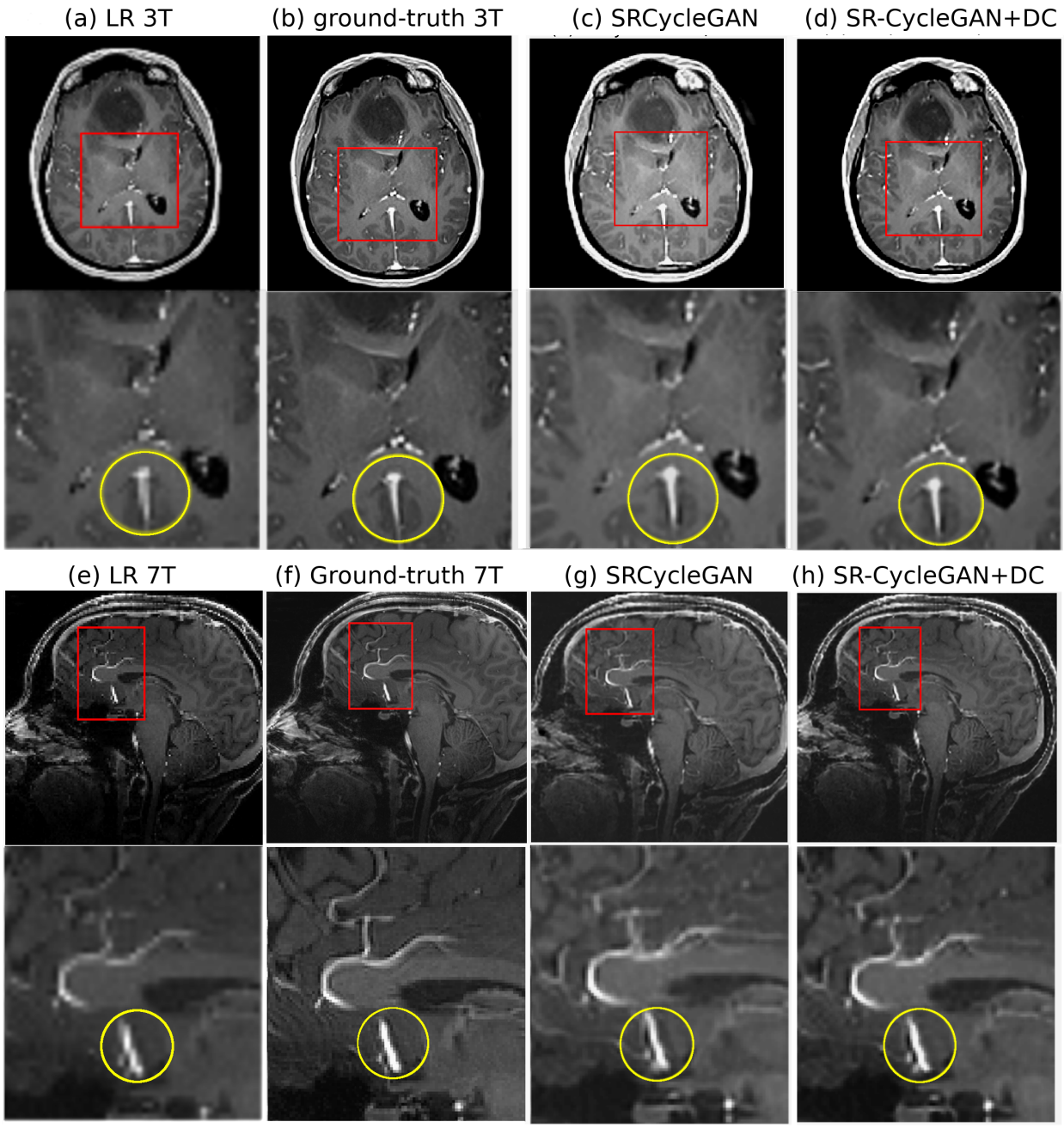


Figure 4.10: Visualization of 3D model performance on CHU 3T and 7T MRI: (a),(e) low-resolution MRI, (b),(f) ground-truth MRI, (c),(g) SRCycleGAN, (d),(h) SRCycleGAN with deconvolutional layers output. On randomly selected sample, zoom-ins are shown in the red box. The vessel in the yellow circle is blurred out LR MRI (a),(e) and partially recovered in (d),(h) and preserves more details in (c),(g)

properties than natural images, which limits the SR imaging performance of most traditional methods. Because the sampling and degradation operations are coupled and poorly posed, the tasks of SR can only be performed to a limited extent by traditional methods. These methods cannot effectively recover some fine features and run the risk of blurring new artifacts. In unpaired training, the generated output cannot be compared with the target output to improve performance because of the different voxel values.

A 2D model can directly apply to 3D volumes slice-by-slice or combine coronal, axial, and sagittal views. However, structural information in 3D volumes is more natural with details rather than combining slices. The 3D model reconstructs the whole volume; therefore, the volume uniformity is more secure than a 2D model. In 3D volume, a blood vessel may cast its edge to neighbor slices when the 2D SR model processes by the single slice. It also can reduce the appearance of noise/artifacts or even the gap between the object in reconstructed results.

4.7 Conclusion

Current approaches in MRI super-resolution require paired low- and high-resolution data for training, which are difficult to obtain due to limited resources and computational time. In the first work of this dissertation, we proposed the SRCycleGAN to solve super-resolution on MRI data. The advantage of self-learning between two classes can be used to perform the unpaired training. The proposed methods can work stably on different types of MRI. The evaluation of the reconstructed images on both 3T and 7T MRI shows exploitable results with low distortion and detailed texture.

We also compared SRCycleGAN with different methods in the same areas to have an objective perspective on model performance. Quantitative evaluation shows that SRCycleGAN is better and more measurable than other methods at different scaling factors.

We want to improve the quality of MRI further. Since super-resolution aims to reconstruct the input with the highest quality possible, a gap exists between high-resolution 3T and standard 7T MRI remains. In the next section, we will implement the synthesis task to produce UHF-MRI to compare the performance with the SR task.

Chapter 5

Ultra-high field MRI synthesis

The work presented in this chapter is under submission as:

[1] Do, H., Bourdon, P., Helbert, D., Naudin, M., Guillevin, R. (2022). Realistic ultra-high field MRI rendering using cycle-consistent generative adversarial networks.

Submitted to SPIE Journal of Medical Imaging

5.1 Introduction

In recent years, ultra-high field (UHF) MRI *e.g.* 7-Tesla (7T) or higher devices were introduced, which provide better signal-to-noise ratio (SNR) sensitivity and higher spatial resolution compared with 3-tesla (3T) or 1.5T MRI [149]. However, at this moment, 7T MRI machines are significantly more expensive and are still in the early stages of large-scale deployment; thus, they are less common in hospitals and clinical centers. Therefore, enhancing medical image quality through artificial means such as machine learning-based synthesis holds potential for clinical and research interests.

Image synthesis techniques are based on a paradigm shift in which a transform model can learn to regenerate images from a given input domain to another desired domain. Image synthesis in medical imaging is an active research topic with numerous applications in radiology. The idea behind medical image synthesis is to accelerate the usual procedure by replacing all or part of the acquisition, which is usually limited by time, labor or cost constraints. Applications of image synthesis in medical imaging range from cross-modality translation within individual types (*i.e.* MRI T1 $\leftrightarrow$ T2) or between different types of medical images (*i.e.* CT $\leftrightarrow$ MRI) to field- strength conversion (*i.e.* 3T $\leftrightarrow$7T) [219].

Within the scope of the thesis, we aim to support the medical diagnosis by improving the quality of routine MRIs. In the previous chapter, we presented quality enhancement on routine MRI by performing super-resolution. In this work, we aim at achieving UHF (7T) quality out of routine (3T) MRI. While both super-resolution and image synthesis enhance MRI quality, their requirements and output are different.

Super-resolution produces high-resolution out of low-resolution through reverse transform rules trained on a set of images emulating HR content, with their downsampled versions serving as LR content. While achieving good performance in aesthetic properties, ones could argue that pixel

or voxel downsampling in the digital image domain is unrealistic and inconsistent with actual resolution changes in MRI field maps. Synthesis, on the other hand, holds more ambitious promises by relying on standard MRI data pairs (*e.g.* 3T and 7T) to emulate conversion on a physical level. A substantial drawback for MRI synthesis is the difficulty in having reliable training data *i.e.* both 3T and 7T MRI content for the same subject acquired simultaneously. For example, to obtain paired 3T and 7T MRI, patients must take the test on the same day. Later, the post-process has to be prepared manually with an exact alignment process. With the CHU dataset that has been built from MRI systems at CHU Poitiers, we keep using 3T and 7T MRI to implement the MRI synthesis for the 3T-to-7T conversion task.

Our main contributions are:

- We take advantage of the proposed CycleGAN model with appropriate modification to do the synthesis on entire 3D MRI volumes. The architecture of the generator is modified to work on the synthesis task. Although the proposed model can perform the task through unpaired training, this task is implemented on weakly-aligned data pairs to maximize the performance;
- We do experiments on practical MRI data to evaluate the performance of the dataset. Real 3T and 7T MRI are processed and used for experiments;
- Results show the efficiency of the proposed CycleGAN, overcoming limitations in training data. Finally, we compare the proposed method with traditional methods to highlight the performance;

In the next Section 5.2, we present related work for synthesis tasks with the most common architectures for medical images. Then we discuss the details of network architectures used in our experiments in Section 5.3 and 5.4. After that are the experimental results and comparison between methods; finally will be the discussion.

5.2 Related work

5.2.1 Common network architectures

Recently, image synthesis has gained much interest in many exciting clinical applications for different types of medical images such as MRI, CT, PET, etc. Statistical methods are usually implemented with explicitly defined rules for converting data between two domains and require a specific case-by-case parameter to optimize performance. Hence, these specific methods, which usually depend on the characteristics of the involved imaging modalities, lead to application-specific complex methodologies. Besides, it is also a challenge to build these models on the two imaging modalities, including precise information, such as anatomical and functional imaging.

Following the rapid development in deep learning, neural networks and their variants have been proposed and become popular methods for medical image synthesis. Through a network learning process to map between the input domain and desired domain, these methods can perform a prediction stage to synthesize the target from an input. In contrast with statistical methods, learning-based methods are more generalizable due to the regularization of architecture different

modalities with minimal adjustment. Hence, these methods allow the robust transformation of various clinical modality imaging.

Deep learning methods in medical image synthesis are mainly categorized into three types: auto-encoder (AE), U-net, and generative adversarial network (GAN). The details of these three networks are present in chapter 3; hence, in this section, we quickly summarise their architecture with a brief introduction to how it is applied for medical image synthesis. These methods are not different from each other, but their complexity is stepwise increases.

5.2.1.1 Autoencoder

An auto-encoder (AE) and variation autoencoder (VAE) is a class of generative model trained to learn to reconstruct their inputs by extracting useful intermediate representations. A basic AE consists of input, output, and functions that encode feature maps into latent space and then decode them to form desired output. In computer vision, the architecture of encoder and decoder network in AE is usually forms of CNNs. It is composed of several convolutional layers with trainable parameters. As a CNNs-based network, normalization, activation function, dropout or pooling layers are also applied to improve the performance of model. Activation function and normalization are the most common components due to their benefits in reducing internal covariate shift for faster convergence. Besides, dropout and pooling layers are usually used to avoid overfitting and save memory.

However, the AEs are limited on non-regularized latent space; where it is challenging to apply to current medical synthesis. Instead, VAEs are used to perform different complex tasks in medical image synthesis. For example, studies in [144, 56] used ResNet in AE architecture due to its shortcut connections that skip one or more layers, easing the training of the deep network without adding extra parameters or computational complexity. It allows feature maps from the initial layers that usually contain fine details to be easily propagated to the deeper layers.

In addition, the study in [72] used AE to synthesize CT from MR images. The encoder architecture uses several convolution layers, ReLU, batch normalization, and pooling layers to extract hierarchical features. The decoder architecture is the same, except the pooling layers are replaced by deconvolution layers to reconstruct the CT images from low to high resolution. The encoder and decoder are connected by shortcuts on multiple layers to enable high-resolution features from the encoder to be used as extra inputs for the decoder. The model was trained on pairs of MR-CT slices in 2D space.

5.2.1.2 U-net

The architecture of U-net is close to an AE variant, where it consists of an encoder and a decoder. The encoding part extracts hierarchical features from the input using a CNN-based architecture to reduce the input and increase depth. In contrast, the decoder uses deconvolution layers to reconstruct features into output with desired content. Two parts of the U-net are linked and concatenated from top to bottom layers. These layers can also learn simple features captured in different levels.

Most studies using U-net followed the general architecture, with modification and improvement in training procedures to perform different tasks. Studies in [130] used the same architecture. However, the model is trained on discretized maps from CTs to produce CT synthesis for segmentation

problems instead of MR-based CT synthesis.

Components in U-net are also modifiable. For example, batch normalization layers, can be replaced by other function such as instanace normalization. Study in [146] used generalized parametric ReLU instead of the usual ReLU layer to adaptively adjust the activation function. [207] added a dropout layer before the first transposed convolution in the decoder to avoid overfitting.

In [88], they proposed an U-net-based uses fully connected conditional random field to provide complementary information between neighbouring voxels and the base classifier to attenuationaly correct PET/MR imaging. Based on conditional random field [112], their method built pairwise potentials between all pairs of voxels from original volumes and the output of models in 3D space. Dong et al. [53] proposed U-net architecture to synthesize CT for attenuation correction of PET/MRI. The network uses a landmark advance and a self-attention design to use the feature maps from coarse-scale in the encoder to identify relevant features and eliminate noise prior by assigning attention scores. In addition, Gupta et al. [71] introduced a U-net based model to generate synthetic CT images from MRI for treatment planning.

In addition, the architecture of U-net now is more modified with building blocks in the encoding and decoding part. Instead of only convolutional layers, residual blocks from ResNet are used to produce feature maps with residual shortcuts and to save computational memory.

5.2.1.3 GAN

At this moment, GANs are widely applied to medical imaging synthesis [234]. GANs and their many variations have been applied to enhance the quality of medical images through super-resolution [241, 38, 218], cross-modality synthesis tasks such as CT to MRI [123, 90, 225], or 3T to 7T MRI [148, 161]. Many GAN-based methods have been proposed. For example, a study in [148] used GAN with binary cross-entropy loss function and patch-based training to generate different modality images.

In addition, varying structures composed of building blocks have proven useful for different applications. Several studies have demonstrated the efficiency of residual blocks in GAN architecture for medical image synthesis tasks, where differences between input and output are not too large such as CT $\leftrightarrow$ CBCT or low-counting PET $\leftrightarrow$ full-counting PET [74]. Study in [56] integrated residual blocks in a CGAN architecture to MR$\rightarrow$CT. Two deconvolution layers replace fully connected layers. Kim et al. [101] proposed a GAN that used the U-net architecture with the residual training scheme for the generator. Olberg et al. [152] introduced a pyramid convolutional network within U-net generator to exploit the characteristic of single pixels.

The difficulty of training GAN is also mentioned due to vanishing gradients or mode collapse when both generator and discriminator are trained to be optimal. To address this problem, Yang et al. [231] used Wasserstein loss in WGAN for the discriminator as an alternative with even smoother gradient flow and faster convergence. Besides, a study in [156] indicated that a feature-matching approach by using a new objective function, in which the generator produces the synthesized images to be close to the expected value of the discriminator instead of directly maximizing the final output of the discriminator.

Finally, since CycleGAN [250] was published, many following studies in medical synthesis have been inspired. The CycleGAN introduced a cycle-consistent mapping workflow within two separate directions that do not require paired training sets. The potential of CycleGAN for medical image synthesis is appealing due to the complexity of exactly matching pairs of data or misalignment

errors.

5.2.2 MRI synthesis

Research in medical image synthesis focusing on MRI can be categorized into two groups based on the objectives: inter-modality and intra-modality synthesis.

The inter-modality synthetic techniques include studies of image synthesis from medical imaging to another, such as from MR $\leftrightarrow$ CT, MR $\leftarrow$ PET, etc. Synthesizing MR from CT (and vice versa) using neural networks is a fundamental and common topic in medical image analysis. The primary purpose of CT-based MR synthesis is to use CT acquisition to replace MRI when MRI was not too popular. In the beginning, the image quality and visualization of synthesized MRI in these studies are considerably different from actual MRI, leading to the limit of direct usage for diagnostic. However, it can be used for non-diagnostic purposes, such as treatment planning for radiation therapy. Using both imaging modalities leads to not only time expense and cost for patients but also the issues of alignment during the fusion process [98] thus, the requirement for the automatic process has increased, which opens the opportunity for deep learning methods.

In general, both MR and CT imaging are usually utilized for treatment planning, such as simulation in the current radiation therapy workflow. MRI demonstrated an excellent contrast on soft tissue, which is useful for delineating organ status or locating tumours. On the other hand, CT gives reference images for pre-treatment positioning and electron density maps for dose calculation. The lack of relationship between voxel intensity of MR and CT leads to the gap in visualization and contrast of images, which can cause the failure of intensity-based calibration approaches. Statistic methods usually segment MR images into material classes to assign to CT. Hence, it usually relies on the segmentation and registration process, which contains a significant error due to the ambiguous boundary between classes. Regarding deep neural generative networks, popular studies using VAEs to synthesize MR-based from CT can be mentioned, such as [227], [51]. In terms of U-net, the number of studies is more diverse. [37, 5] applied U-net based methods to produce brain CT-based from T2-weighted MRI, while [146, 131] used T1-weighted MRI with similar architecture.

The group of intra-modality investigations consists of studies that translate data between two different protocols from an imaging modality. Various applications in this domain for MRI have been proposed, including translation between sequence types, high strength-field MRI rendering, or restoring undersampled acquisitions.

In the scope of this chapter, only studies related to ultra-high field synthesis will be mentioned. Related work for cross-modality translation will be demonstrated in the next chapter. In general, for all tasks in medical image synthesis, preserving contrast and resolution is the most critical factor that decides the effectiveness of a method. Rendering ultra-high field MRI from low magnetic field MRI allows acquisition on broadly available low-magnetic-field equipment while providing greater spatial resolution and improved contrast, similar to what might be obtained from a cutting-edge devices. In contrast, translation between sequences and restoration of undersampled acquisitions can both shorten acquisition times. Although these applications are motivated by distinct clinical goals, they pose the same technical challenges during the task of image synthesis: preserving contrast, resolution, and biological details that are relevant for diagnosis

In the beginning, conventional methods have been applied to address these problems. Compressed sensing (CS) approaches perform on data that have a sparse representation in the trans-

formation domain [233]. For example, in the case of multi-contrast MRI translation, the contrast level of a patch is expressed as a sparse linear combination of patches in an atlas, and the combination is then applied to image patches in the target contrast [122]. However, these methods require time and resource-intensive optimization as an iterative algorithm, while optimization is usually implemented as an iterative algorithm, which is time and resource-intensive.

Deep learning methods, in contrast, encourage the integration of neural networks into these strategies for their superior mapping capability of nonlinear relationships and significant savings in compute time. It is also common to apply neural networks in synthesis domains. Moreover, learning-based methods provide comparable or better performance on quantitative image quality metrics with less computational time. Zhang et al. [243] proposed a cascaded design using dual-domain and interactive multi-layer network streams in the spatial and frequency domains. Compared with a single spatial domain, the dual-domain method presented better visual results. Qu et al. [162] designed a wavelet-based affine transformation layer to modulate feature maps from the spatial and wavelet domains in the encoder, followed by an image reconstruction in the decoder that synthesizes 7T images from wavelet-modulated spatial information. In addition, also synthesizing 7T MRI from 3T MRI, Nie et al. [148] proposed a GAN-based method that solves the task at the image level.

GAN for image synthesis serves as a form of data augmentation and also as an anonymization tool. Other research topics also in the fields of MRI synthesis can be mentioned, such as brain tumour segmentation using coarse-to-fine GANs[145], generating synthetic medical images to address retinal fundus images [70], synthesizing realistic prostate lesions in T2-weighted and apparent diffusion coefficient resembling [106]; synthesis of patient-specific transmission image for PET attenuation correction in PET/MR imaging of the brain using a CNN [194]; image synthesis with GANs for tissue recognition[242]; synthetic data augmentation using a GAN for improved liver lesion classification [62].

5.3 Network architecture

In the scope of the work, our objective is to produce an ultra-high field MRI volume from a routine 3T MRI. Similar to the previous work on super-resolution, the general architecture is based on cycle-consistent GAN. Unlike the generator in super-resolution model, the generator for MRI synthesis is modified to fit the problem and reduce the complexity and cost. CycleGAN work to perform the adversarial training while the new generator responds for the synthesis part.

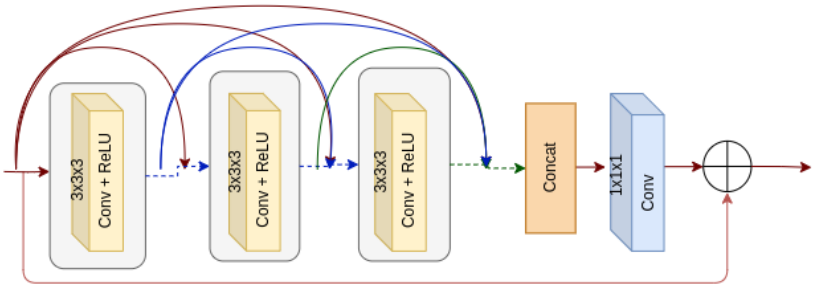
5.3.1 Adversarial network architecture

As presented in Chapter 3 and 4, CycleGAN contains two generators and two discriminators. Two generators work to produce output between 3T and 7T MRI while two discriminators predict real or generated data. The training for this task is paired, but the cycle-consistency loss is still kept to force the synthesized images to be related to their inputs for each domain and ensure model performance. Details of pre-processing for pairing are presented in Section 5.4.2

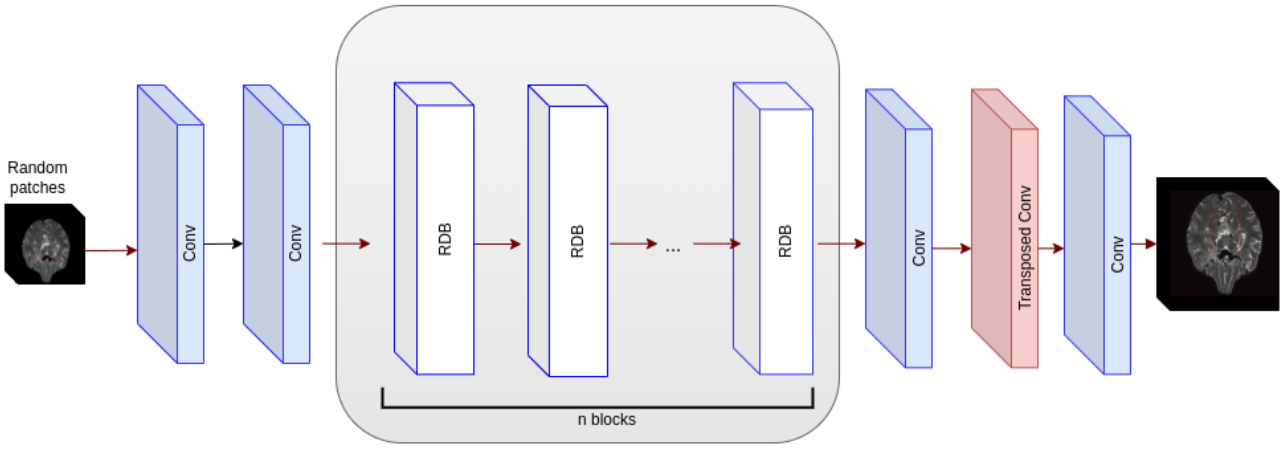
Based on the result of MRI super-resolution in both 2D and 3D space, we found that the performance of the 3D model is better than the 2D model in all aspects. Thus, in this work, we only focus on the 3D model.

5.3.2 Generative network architecture

The generator of CycleGAN for UHF MRI synthesis also use residual dense blocks (RDB) [244] - a combination of residual blocks and dense blocks as building units to extract feature information. In terms of residual blocks, when the input bypasses these hidden layers via the residual connection, the hidden layers enforce the minimization of a residual image between the source and ground truth target images, thereby minimizing noise and artifacts. In contrast, dense blocks concatenate outputs from previous layers rather than feed-forward summation as in a standard AE block, capturing hierarchy information to better represent the mapping from the source to the target image modality. RDB has proved its potential in the MRI super-resolution task; hence we would like to apply it also in the synthesis task. Figure 5.1.a shows the 3D implementation of the RDB block. In this implementation, a RDB also contains 3D convolutional layers followed by a ReLU activation function [65] in continuous connection.



(a) Residual Dense Block



(b) Generator

Figure 5.1: Architecture of generator. Here, we use 3 RDBs for feature extraction. The number of blocks can change the size and complexity of whole model.

Although using RDBs as building units, the generators for synthesis work are quite different from the super-resolution model. Shallow features are generated from raw input through two single convolutional layers to reduce input size for the following computation. Then, RDBs work

as building units to synthesize information from feature maps. However, only local features are used to store information. The number of features is also increased through each block with a fixed growth rate to synthesize information.

In this model, we remove the fusion operator after the final blocks to reduce the complexity and computation time for training. The bottleneck convolutional layers are kept to represent the features with reduced dimensionality. Moreover, we also remove the concatenation between shallow and fused features before the resampling part. Finally, a transposed convolutional layer is used to reshape encoded features to form the desired output. The details of 3D generator implementation are shown in Figure 5.1.

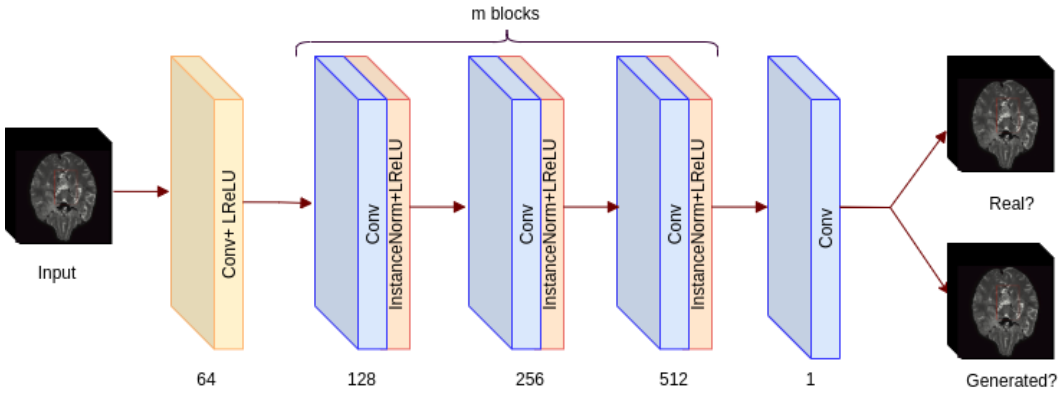


Figure 5.2: Architecture of discriminator. It contains several convolutional layers followed by Instance Normalization and Leaky ReLU for 3D data.

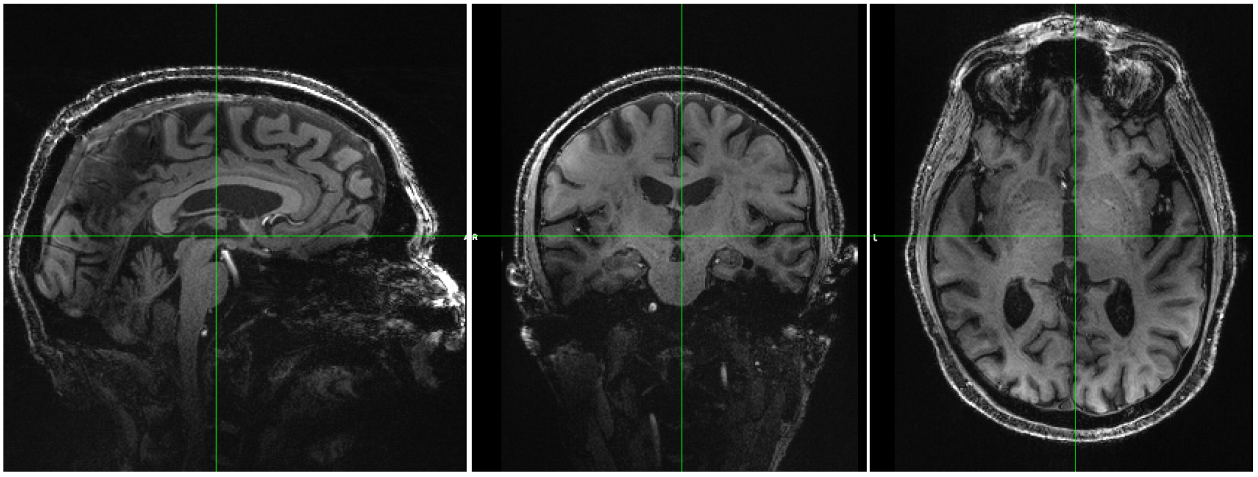
Unlike super-resolution, the generators in the synthesis task do not consider the difference in size between input and output. Hence, the architecture of the two generators is similar and the same for the two discriminators. The structure of 3D discriminators is shown in Figure 5.2. It is a CNN that contains several convolutional layers mixed with an instance normalization layer followed by Leaky ReLU (LReLU) activation to extract information from 3D volume and label whether it is an actual or generated image. The depth of the network is customizable depending on the number of mixed blocks in the network. In the end, a convolutional layer with a single output channel is used to produce values ranging from 0 (generated image) to 1 (real MRI). Besides, we also make an option to use the sigmoid function with the convolutional layer to predict.

5.4 Experiments

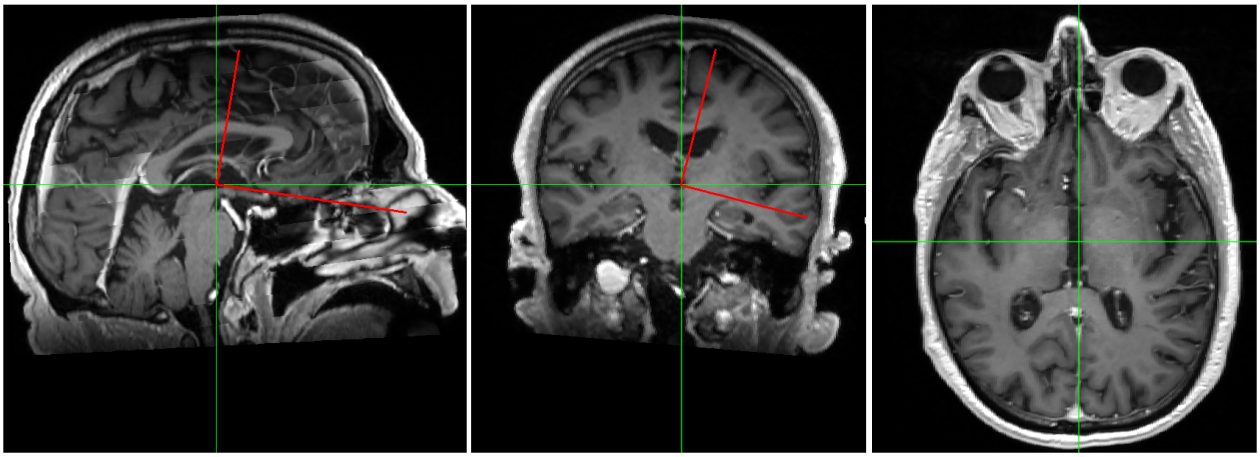
5.4.1 Dataset

As presented in previous work, we have built the CHU dataset with the support of Siemens Healthineers MRI devices at Poitiers University Hospital, which contains both 3T and 7T MRI with high-resolution. 127 3D MP-RAGE brain MRI subjects are T1-weighted for both 3T and 7T at different slice spacing. 3T MRI volumes were acquired from a Siemens Magnetom Skyra scanner, while a Siemens Magnetom Terra scanner produced 7T MRI volumes.

Retrieving standard pairs of 3T and 7T data for MRI synthesis is a tedious task with unsatisfiable constraints such as acquiring both scans simultaneously on the same patient or having an exact slice thickness match between both resolutions. In the CHU dataset, only ten pairs



(a) 7T MRI

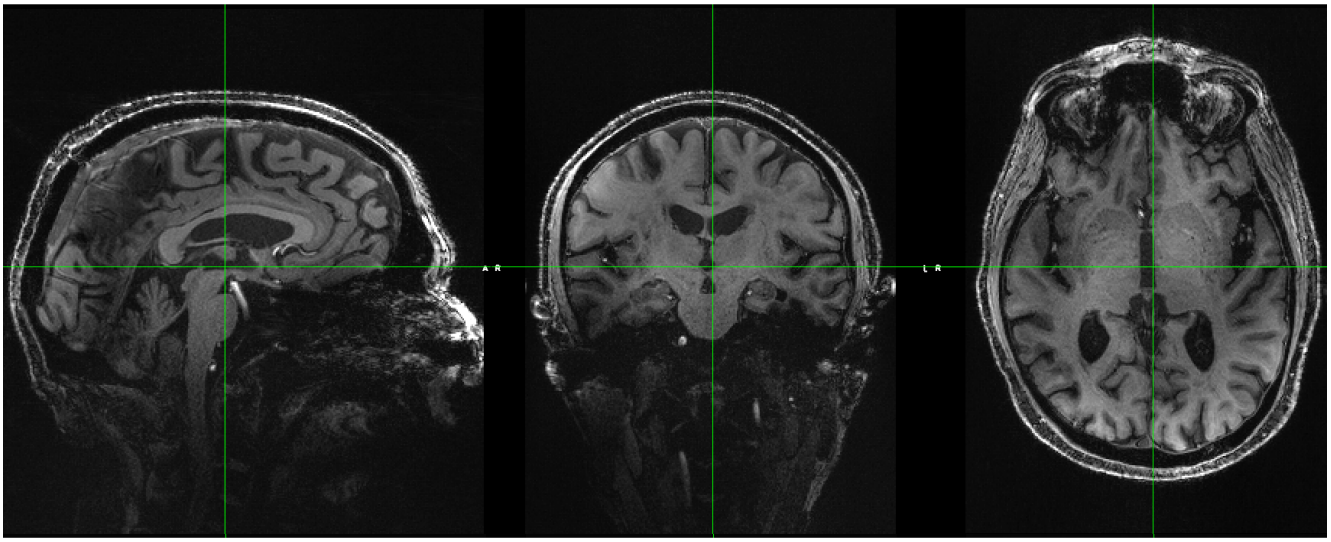


(b) 3T MRI

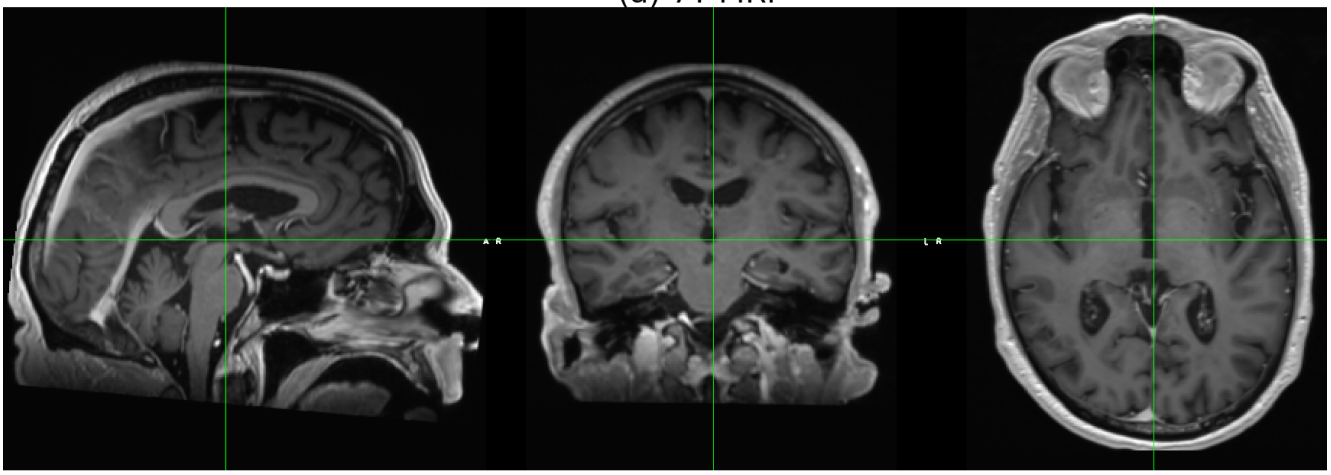
Figure 5.3: Illustration pair 3T and 7T MRI before preprocessiing. 3T samples are inclined with 7T samples, lead to a mismatch of voxel position.

of 3T MRI at 0.9mm and 7T MRI at 0.75mm qualified for the synthesis task. Pairs of 3T and 7T were acquired within the closest time to avoid the appearance of artefacts between subjects. Moreover, since the 3T 0.9mm with voxel size $240 \times 288 \times 192$ and 7T 0.75 mm with voxel size $340 \times 340 \times 240$ were used, we can minimize the mismatch in position and have the most accurate volume alignment because of comparable slice thickness. Due to the difference between volumes properties, the mismatch in voxel slice and position between volumes are quite clear observed, even though we have selected samples with the closest spacing. Figure 5.3 illustrates raw pairs of 3T and 7T MRI before alignment.

Although the design of CycleGAN can perform the unpaired training, due to the limit of weakly-registered pairs and the size of the dataset, the synthesis model is trained with paired data to optimize the model performance. Details of experiments are presented in subsection 6.4.3. Here, the objective of the synthesis aims to produce the UHF MRI from a given routine MRI in the CHU dataset. In this work, we separated pairs into a train/test set with ratio 8:2, respectively. Cross-validation is also applied during training phase due to the limited size of dataset. To evaluate the performance of the proposed model, the synthesized 7T MRIs are compared to ground-truth 7T MRIs.



(a) 7T MRI



(a) 3T MRI aligned by 7T

Figure 5.4: Illustration pair 3T and 7T MRI after alignement. Voxel position of 3T and 7T samples are similar

5.4.2 Pre-processing

The synthesis for MRI in this work mainly requires the alignment between 3T and 7T MRI subjects for the paired training. We used the FLIRT linear registration tool from FSL [89] - a comprehensive library of analysis tools for MRI brain imaging data, to register pairs of 3T and 7T MRI to standard space using and remove pose differences. Each 3T MRI is rigidly aligned to its corresponding 7T image. Figure 5.4 shows 7T and 3T aligned by 7T MRI. After the alignment process, volumes are considered to have same voxel size and spacing.

To reduce adverse impact of signal variation across different scanners and sites, training samples of each domain are standardized with other samples in the same domain using histogram standardization to make voxel values in the same intensity range. Then, all aligned samples are normalized using z-norm to reduce the computation cost and avoid training frag. The z-normalization uses terms of mean and standardization to re-scale transform voxel value within range [0, 1].

Finally, augmentation techniques such as random rotation and flipping are applied to extend

the size of the dataset, increase variability, and avoid over-fitting before training.

5.4.3 Training setup

In this work, we evaluate the performance of our method is the context of MRI synthesis, which is its most challenging task. We perform experiments to produce 7T-like MRI from 3T and vice versa on a subset extracted from CHU data in Section 4.4.1. Training images are paired by registration to preserve quantitative pixel values and reduce baseline geometric mismatch, allowing the network to focus on mapping details and accelerate training.

Two generators of the network are built in a fashion similar to significant modifications from the previous model. Fused connections of building units are removed to reduce the complexity and enhance the performance. Besides, the link between shallow feature and extracted features are also cut to reduce the dependence of output on input. The bottleneck convolutional layer and transposed convolutional layer are kept to obtain a representation of the features with reduced dimensionality and reshape the feature to the desired size.

In terms of complexity, it will increase along with the increase of model depth or the input size. However, since we have reduced several unnecessary fused connections and the concatenation of the shallow feature, the model configuration can be enhanced to improve the performance. In this work, each generator contains six RDBs, where each block includes three dense blocks in residual connection.

Model is also trained on patches to ensure diversity within samples. After alignment, 3T and 7T MRI volumes are fairly registered, then augmenation techniques such as flip, ratation, are applied to expand number of samples. For each batch, patches are randomly extracted into a maximum size of $64 \times 64 \times 64$ patches on 3T volumes and corresponding to the size on 7T volumes. The ADAM optimizer is also adapted for optimization.

The batch size is set to 4. The learning rate is initialized to $1e^{-4}$, and decay starts after every 20 epochs. The training CycleGAN for MRI synthesis takes an average of 25 hours with a GPU NVIDIA A100 40GB for 200 epochs. Results of the synthesis model are presented in the next section.

5.4.4 Evaluation metrics

Similar to the previous task, to evaluate the image quality between ground-truth and generated MRI volumes in both tasks, we use peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) metrics.

For practical reasons detailed in Section 5.4.1, actual MRI synthesis performance assessment can only be performed with the *CHU dataset* at the present time. For this experiment our CycleGAN model is now trained on actual 3T and actual 7T pairs as required. Because there are only a very limited number of studies dedicated to generating synthetic 7T out of 3T MRI (or *pseudo-7T*), with prediction algorithms published in this field [148, 243, 11] being trained on private datasets with limited size, we decided to run the WATNet [162] source code on the *CHU dataset* for bench-marking.

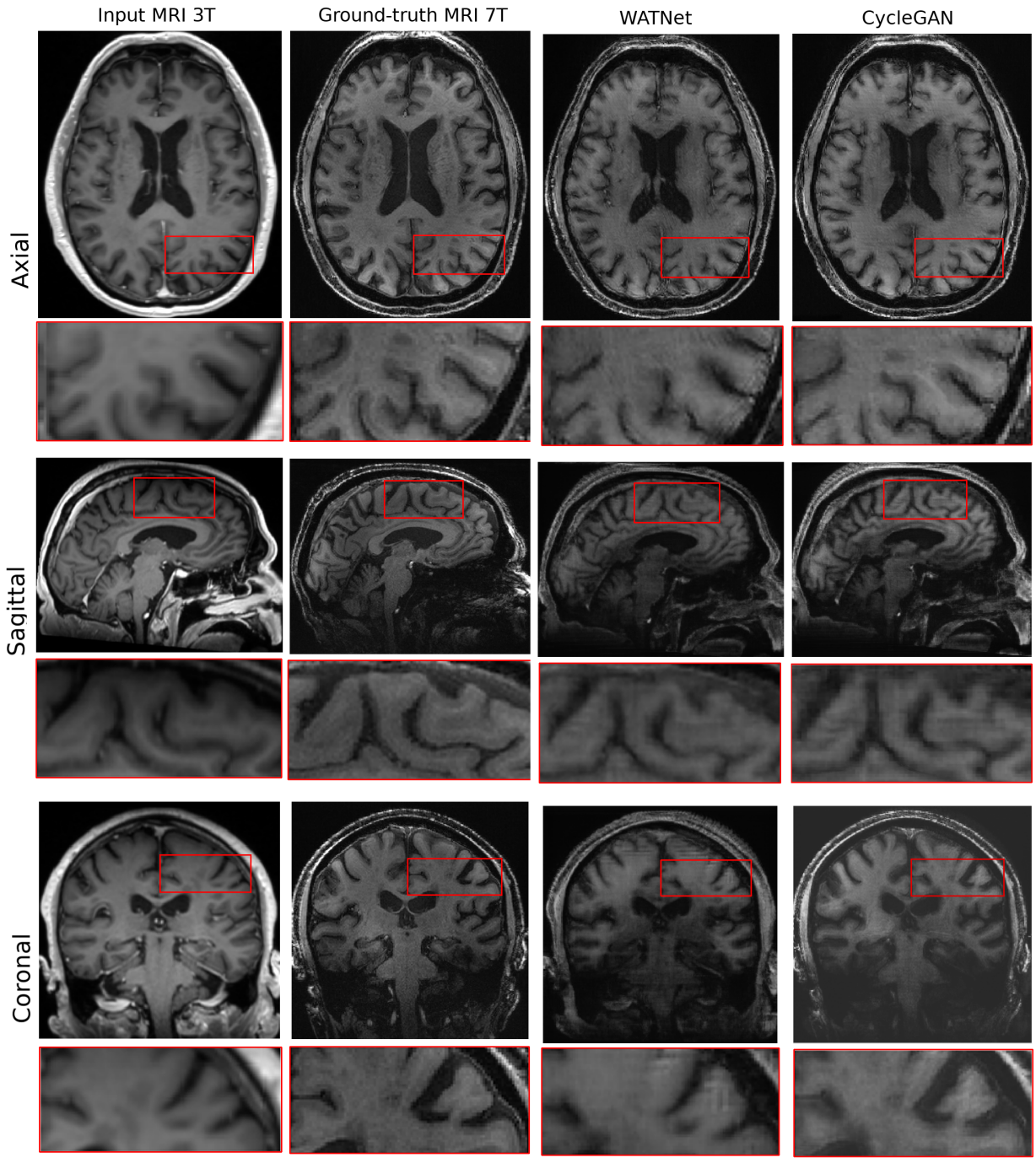


Figure 5.5: Axial, sagittal and coronal views of synthetic 7T MRI. From left to right: 3T MRI (input), ground-truth 7T (expected output), synthetic 7T generated by WATNet, synthetic 7T generated by our model. Zoomed-in areas appear as red rectangles.

5.5 Results

Table 5.1 illustrates the model performance for MRI synthesis tasks in terms of PSNR and SSIM assessment. CycleGAN outperforms the baseline methods. In case 7T MRI synthesizing, the SSIM/PSNR values reach 0.83/48.50. In addition, Figure 5.5 visualizes the generated outputs from the model in both directions (3T to 7T). Axial zoomed-in views in the red rectangle indicate that our method clearly improves the quality of grey/white matters from 3T MRI compared to the 7T ground-truth. Moreover, the details of the sulcus and gyrus are precisely reconstructed with high accuracy, close to 7T MRI quality.

Table 5.1: MRI synthesis quality assessment in terms of PSNR (dB) and SSIM values. Synthetic 7T data generated out of 3T data with our method and WATNet is compared to its corresponding ground-truth samples for quantitative evaluation.

Metric	WATNet	CycleGAN
SSIM	0.81025	**0.8309**
PSNR (dB)	43.29	**48.53**

With PSNR values over 40dB, the results presented in Table 5.1 demonstrates that while CycleGAN was originally designed for weakly paired or unpaired data fitting tasks such as image style translation, we can notice that combining its cycle consistency constraint with input samples that are at least rigidly aligned can help to overcome the limitations in data size and weak 3T/7T pairing.

In addition, due to the design of CycleGAN, we also can perform the 3T synthesized from 7T. Table 5.1 and Figure 5.6 also presented the result on model for the backward generation. Although the applicative value of 3T-synthesized from 7T MRI is of low importance, it has proved that the proposed model can perform a good translation between two domains. With PSNR and SSIM value even higher than the forward cycle, the proposed CycleGAN ensure the performance of the model on the CHU dataset without overfitting problems.

5.6 Discussion

5.6.1 Overview

The CycleGAN model has presented a potential result in the field of MRI synthesis. However, we also observed limitations in the generated output in visual assessment and measurement values.

This task is completed on a dataset with a limited size. We have ten pairs of 3T and 7T MRIs among different samples in our dataset. In general, samples are not genuinely standard for medical image synthesis. First, pairs of 3T and 7T MRIs come from the same patient but were not acquired simultaneously, with the same thickness of slices. Thus, pairs of data for MRI synthesis are mostly similar. In practice, 3T and 7T MRI volumes were collected with the closest thickness to maximize the similarity between subjects at precise locations. Nevertheless, since 3T and 7T MRI are not precisely equal in size, applying the alignment techniques to original data for the image translation task is mandatory. However, the alignment techniques do not fully overcome

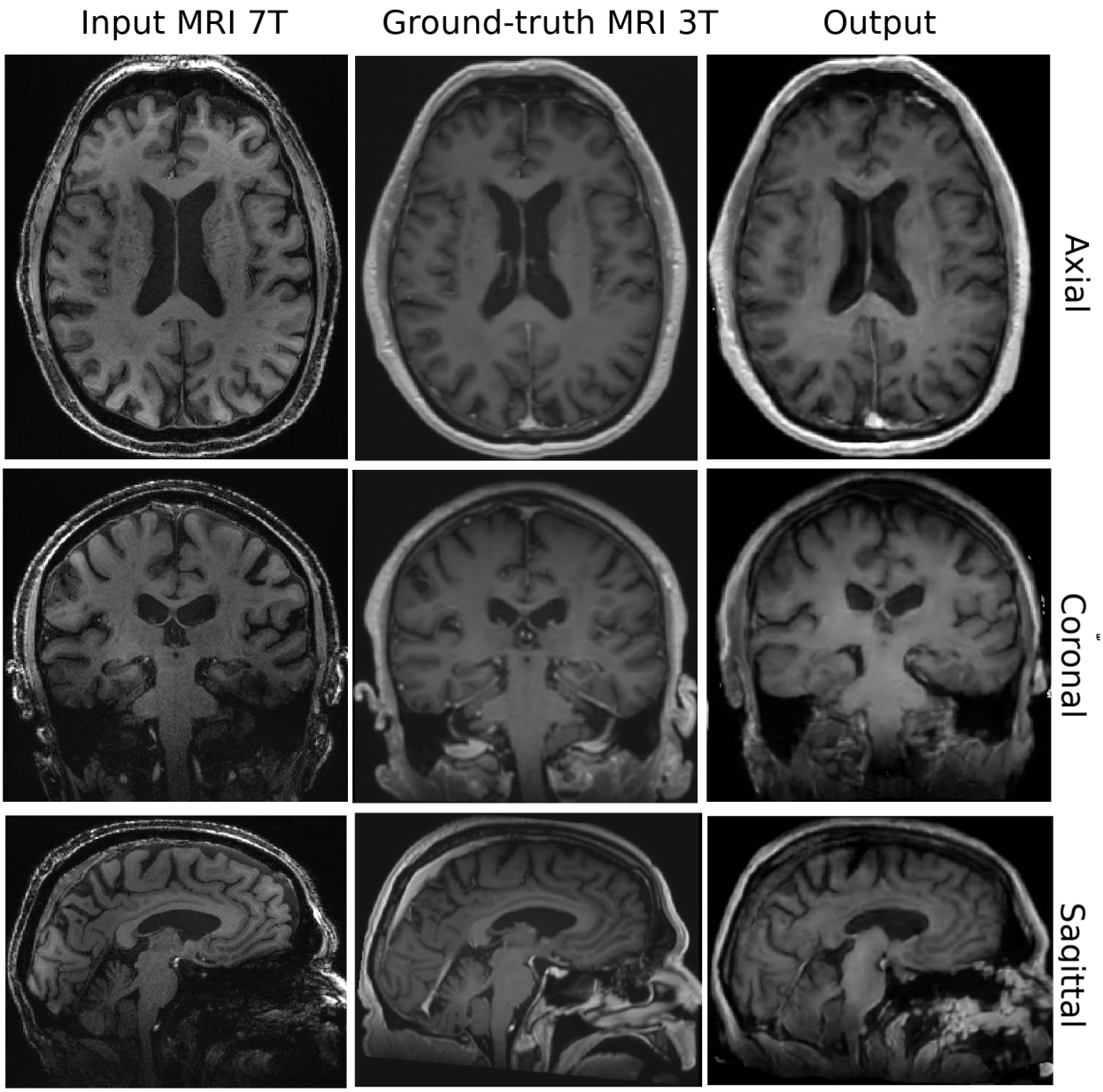
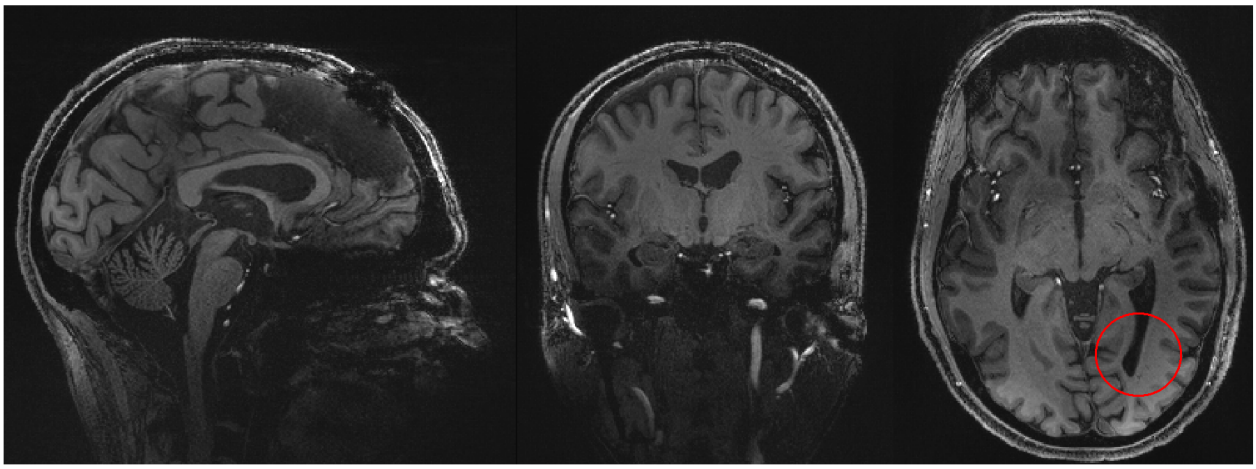


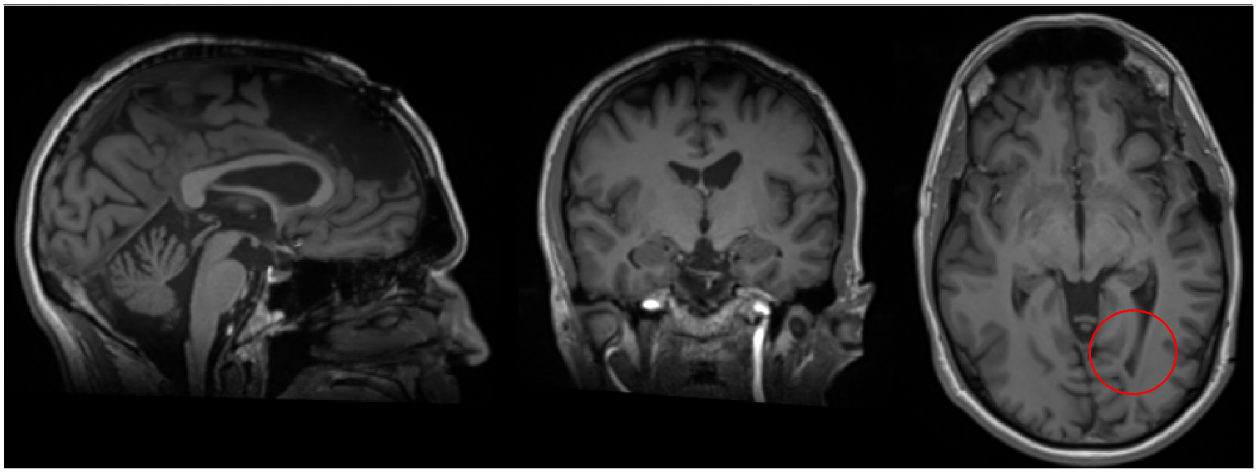
Figure 5.6: Visualization in axial, sagittal and coronal views of 3T synthesized MRI from 7T. From left to right: 7T MRI input , real 3T MRI, and CycleGAN output.

the voxel spacing difference from data acquisition.

The mismatch position can be observed visually from Figure 5.7. Because pairs are weakly paired through automatic alignment, the appearance of artefacts is inevitable. We tried to minimize errors during pre-processing. The proposed method is trained to handle the difference based on semi-supervised learning. We consider that the performance can be improved by adding or standardizing data. For another example, in a close-up coronal view in Figure 5.5, the difference between 3T input and 7T output is clear, but our model can reconstruct 7T MRI from 3T MRI with detailed texture and close to 7T ground-truth. In addition, since the amount of training and testing sets is limited while the diversity of training is only ensured by the augmentation process



(a) 7T MRI



(b) 3T MRI aligned by 7T

Figure 5.7: Illustration mismatch between pair 3T and 7T.

and patch training, CycleGAN provides an exploitable result in measured values on the test set.

Based on [219], we also realize the advantage of using 3D models compared to 2D models for medical images. A 2D model can directly apply to 3D volumes slice-by-slice or combine coronal, axial, and sagittal views. However, structural information in 3D volumes is more natural with details than combining slices. The 3D model reconstructs the whole volume; therefore, the volume uniformity is more secure than a 2D model. In 3D volume, a blood vessel may cast its edge to neighbour slices when the 2D SR model processes by the single slice. It also can reduce the appearance of noise/artifacts or even the gap between the object in reconstructed results.

In addition, we conclude that MRI synthesis has provided not only higher but also more coherent performance to fit with the research objective than super-resolution to produce a realistic UHF MRI. Although both models are trained to enhance the quality of the input, the SR model generally aims at enhancing the quality of MRI while maintaining the basic input properties, which are not necessarily those of UHF MRI. It means that the SR model is not trained to change the properties of the original input, e.g. the contrast made from the magnetic strength field; hence, the SR model can only produce better 3T with more texture details instead of generating 7T-like MRI from 3T MRI. On the other hand, the MRI synthesis is directly trained to transform input from a domain to the desired output; thus, it is not just performance but also consistency.

Overall, although reconstructed 7T MRIs are still noisy compared to ground-truth 7T MRIs, the detail of white-grey matter has improved differentiation over the 3T example without global morphological alterations. The spatial resolution is strongly increased from 3T MRI in all directions. Besides, the proposed provides a better definition of the basilar trunk and anterior cerebral circulation. In addition, the straight sinus in reconstructed MRI has no deviation of the centerline without morphological alteration compared to the ground truth.

5.6.2 Network characterization

Since pairs of 3T and 7T images are similar in terms of structure but their appearance are quantitatively different, in this work, we take advantage of the residual dense blocks to extract features from input. As mentioned in previous sections, residual connection enforces minimization between input and target images, thereby minimizing noise and artifacts while dense blocks can capture important frequency information to better represent the mapping from the source image to the target image modality. The residual dense blocks have been successfully applied in the SR task; hence we examine the performance of these building units for the synthesis task.

Besides, the complexity of GAN-based models is very considerable, with millions of parameters. The main model keeps the content of adversarial and cycle consistency losses. The generators are modified with severe changes from the previous model to enhance performance.

In fact, in the beginning, we have done several experiments using the SRCycleGAN from previous work to examine its performance on the synthesis task. In these experiments, upsampling operators have been replaced to regularize the input and output size. However, the obtained results are not too impressive. Through the research, we found that the fused operator after feature extraction and the link of shallow feature has limited the model performance. In super-resolution, pairs of LR and HR images are similar in both structure and appearance, while the main difference is the spatial resolution. Information of all layers is preserved based on local and global features fusion mechanisms. Global features from RDBs are stacked to use features from all the preceding layers fully. These have proved efficient in memorizing the critical input information and avoiding artifact generation during the unpaired training, which is the main objective of the research.

However, in the case of MRI synthesis, the contrast and the appearance of 3T and 7T MRI are clearly different, while the training is executed on pairs which have been aligned to have the same structural information. This fusion and link operators seem to prevent the significant transformation in contrast to images. By removing unnecessary operators, the results obtained from experiments are fastly improved.

Besides, it also helps to reduce the complexity of the model. As mentioned above, it increases along with the increase of model depth or the size of the features. By cutting unnecessary connections, we have significantly reduced the size of feature maps while improving the performance. With the same configuration as the super-resolution model as mentioned in Section 4.4.2, the training model is faster than 30%. Hence, we can expand the training setup with the deeper configuration of network components. Table 5.2 and Figure 5.8 show the quantitative and visualization of CycleGAN using SR architecture and the CycleGAN with reducing connections.

	CycleGAN using SR architecture		CycleGAN for synthesis	
	SSIM	PSNR	SSIM	PSNR
3T to 7T	0.6827	38.87	**0.8309**	**48.53**

PSNR, peak signal-to-noise ratio; SSIM, structural similarity index.

Table 5.2: Average value of PSNR (dB) and SSIM between SRCycleGAN and the current CycleGAN model for MRI synthesis. Synthesized 7T volumes by two CycleGAN models are compared with corresponding ground-truth MRI for quantitative evaluation.

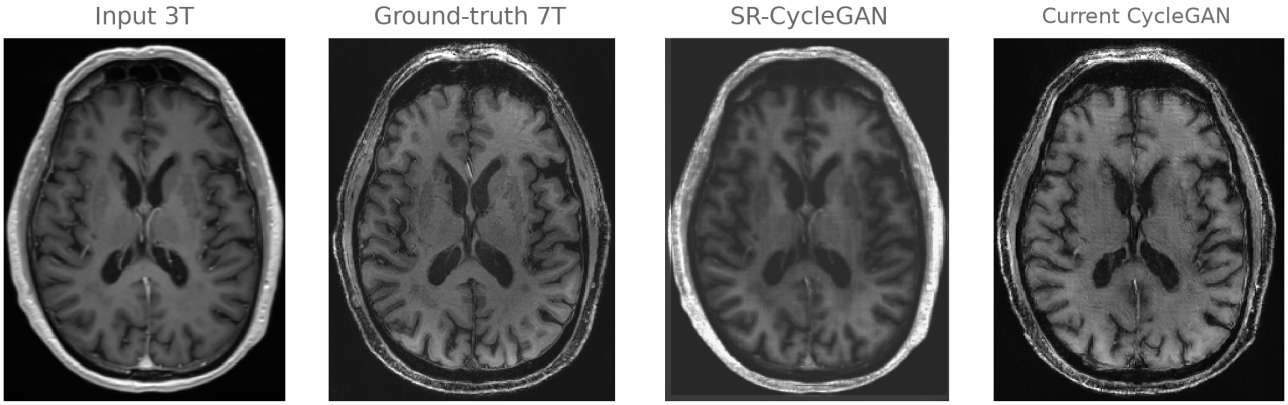


Figure 5.8: Illustration comparison of 7T synthesized MRI for two CycleGAN model. From left to right: 3T MRI input, ground-truth 7T MRI, 7T synthesized using SRCycleGAN generator and 7T synthesized of current CycleGAN.

5.6.3 Perspective

In both tasks, we observe that the reconstructed output at high-resolution levels is quite close to the original input in all aspects, at both the training and testing phases. In general, no additional artifacts appear in the reconstructed images. The proposed method can perform both quality enhancement in 3D space.

Within the scope of the project, we aim to provide an automatically system to produce the UHF MRI images. Our system can take routine brain MRIs to perform the task without any additional process. By synthesizing MRI, we expect our method to produce the output as close to the ground-truth as possible. Synthesizing results from the pre-trained model only takes a few minutes to complete, compared to the usual time of MRI scanners and the current expense for UHF scanners. Hence, it can become a potential research topic for medical image analysis.

5.7 Conclusion

This work presented a hybrid generative model based on CycleGAN to produce realistic UHF-MRI. Advantaged by semi-supervised learning in cycle-consistent architecture, the proposed method can solve synthesis on routine MRI in 3D space. Results from comparison among different methods

demonstrate that the method outperforms current state-of-the-art methods in MRI synthesis to produce UHF-MRI, under both qualitative and quantitative terms. The advantage of cycle-consistent in medical image synthesis is also proven by the convincing results, overcoming the limitation in training data for MRI synthesis.

At this moment, the proposed method works stably on 3D brain MRI, holding a great promise of the research topic in practice. Our objective is to provide an end-to-end system to produce UHF MRI by synthesizing. The network can produce 7T MRIs from routine 3T MRIs without any additional process with a well-trained model. The synthesizing process takes less time than the usual time of MRI scanners, as well as the current expense for UHF scanners. Hence, it can become a potential research topic for medical image analysis. For future work, we will research the functional information and diagnosis on synthesized images to evaluate the research topic comprehensively.

There are different tasks in the fields of MRI synthesis. Among different topics, cross-modality translation is one of research topic that recently receive a lot of attention from community due to its benefits in research and practical. In the next work, we want to examine the performance of CycleGAN model for this task with a comparison to other state-of-the-art methods.

Chapter 6

MRI Cross modality translation

6.1 Introduction

MRI modalities provide complementary information to radiologists for diagnosing, assessing, and planning patient treatment. Different MRI pulse sequences produce different modalities to capture specific characteristics in contrast and function of the scanned area. T1-weighted is used to observe the structure of an object; T2-weighted is utilized for locating tumors; contrast-enhanced T1 (T1c or T1-gado) is favorable for assessment of tumor shape change with its enhanced demarcation around the tumor; or T2 fluid-attenuated inversion recovery (T2-FLAIR) presents the contours of the lesion with water suppression [132, 23]. Taking advantage of modalities by integration can help to explore meaningful information about tissue that facilitates diagnosis and treatment management.

Acquiring a complete multi-modality MRI for an individual patient is challenging due to several factors. There is a specific failure rate due to incorrect machine settings during the scanning process. The mobility of patients during acquisitions also can lead to the appearance of motion artifacts. Moreover, modalities capture different anatomy characteristics, and the relationship between two modalities is highly non-linear; therefore, it is hard to learn the mapping from one modality to another.

In the case of contrast-enhanced T1 acquisition, using contrast agents such as gadolinium is necessary. In general, it was proved safe to be used with a low, non-toxic dose, well-tolerated without any adverse immediate or long-term effects [175]. However, it has been identified as the causative agent in nephrogenic systemic fibrosis in patients with severe kidney failure and for which there is currently no known specific or consistently compelling treatment [27, 181, 97].

Differences in characteristics across modalities in imaging protocols result in a lack of approaches to get consistent image modalities for every patient. Cross-modality translation - a topic aimed at synthesizing a modality from a given modality without real acquisitions, holds great potential for clinical practice. It has been widely investigated in medical image analysis and has achieved initial achievements in recent years.

In the previous work, we presented the medical image synthesis to solve the task of super-resolution and ultra-high field construction using a generative model.

To address the problem of MRI modality in practice, in this work, we mainly focus on implementing different cross-modality generation frameworks to find an optimized method on experi-

mental CHU data, then develop a method with superior results.Within the research objective, the cross-modality models can generate output among different modalities such as T1↔ T2, T1 ↔ T1c, T1 ↔ T2-Flair. However, in this work, we firstly focus on translation between T1 and T1c due to the research value of contrast-enhanced MRI for overall assessment in general and kidney in particular [10]. On the other hand, we have built a dataset of pairs of T1-T1c MRI that cover a wide variety of diseases to train a model that can perform the translation in any case.

Among different generative architectures, we decided to implement two cross-modality frameworks based on GAN to evaluate its effectiveness. The first method in this work is the CycleGAN model. As we have presented above, the benefits of CycleGAN in the synthesis domain are very promising. Regarding medical images, its performance has been demonstrated in previous works on enhancing MRI quality. Here, we take advantage of the proposed method for UHF synthesis by adopting it for cross-modality translation. CycleGAN can complete tasks in both forward and backward directions. Hence it only requires one model to complete this task, while other GAN-based methods usually require two.

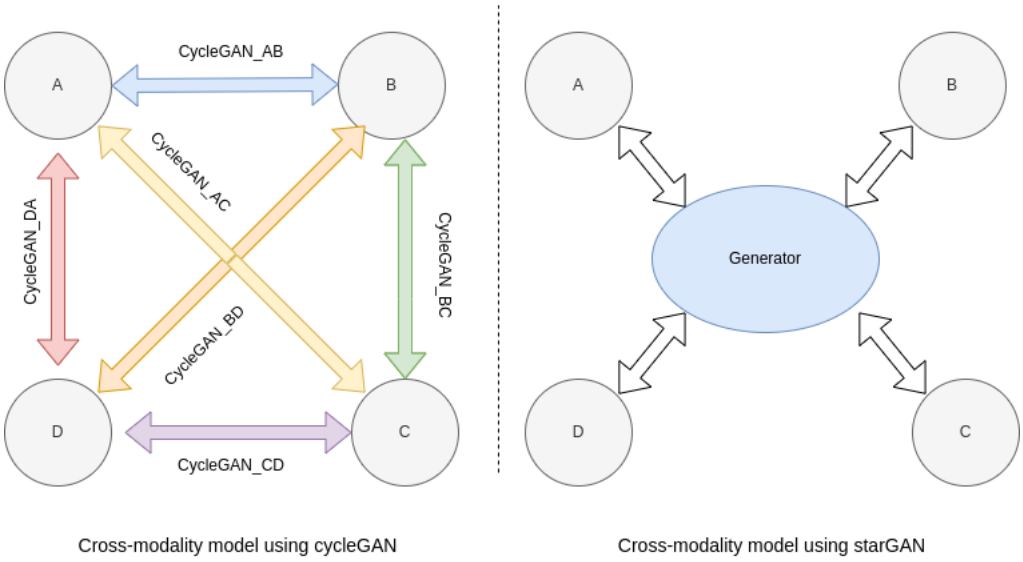


Figure 6.1: Comparison between cross-domain models using CycleGAN and starGAN. To handle multiple domains, there are six CycleGAN model needed for each pair of images, while StarGAN only uses a single generator.

The second framework is StarGAN [42] - a recently novel method for multi-domain image translation. Unlike CycleGAN, StarGAN uses only one unified model with one generator and one discriminator to perform image-to-image translations between multiple domains. For instance, to learn all the mappings between 4 modalities T1, T1c, T2, and T2-Flair, 12 models have to be trained separately using GAN, or six in the case of the CycleGAN. However, StarGAN, only needs a single model. StarGAN aims to overcome the ineffectiveness and inefficiency of the usual GAN model in terms of learning the features among all image domains and their generalizability. We adopted the strategy of StarGAN and specially applied it to multimodal MR image synthesis. Figure 6.1 illustrates the different between starGAN and CycleGAN for multi-modal cross translation.

Experiments are conducted on both research and practical dataset. BraTS dataset is again used in this work because it contains four different modalities including T1 and T1c MRI. On the other hand, we built a practical dataset containing pairs of T1-T1c MRI that cover a wide variety

of diseases.

To sum up, in this phase, our objective is:

- We do a comparative study between methods of the literature and our methods in cross-modality translation among T1-T1c MRI. A comprehensive comparison is provided on research and practical datasets; each has its unique characteristics in terms of data size, patient cohort, and disease status.
- The proposed CycleGAN method that has been applied in previous works, and we would like adopt its architecture for cross-modality translation. Besides, we reimplement the StarGAN in 3D space, with a different configuration to examine its performance on T1-T1c conversion.
- The results demonstrate that our methods obtain a stable result on the research dataset and promising results on the practical dataset. Moreover, experiments have shown that the results of the models are optimizable.

6.2 Related work

At the beginning, multimodal MRI conversion relied on intensity transformation or atlas registration based methods. In atlas registration, images of target modality are reconstructed by the atlas-to-image transformation, which contains pairs of images with different tissue contrasts co-registered and sampled on the same voxel locations in space [172, 35]. However, the performance of these methods is limited only to healthy subjects because an atlas is not usually produced from subjects with diseases [44]. On the other hand, intensity transformation-based algorithms learn the mappings between source and target images based on the intensity of each voxel instead of relying on the strict geometric relationship across different anatomies [91, 92]. Later, statistic learning methods were widely used to replace traditional methods such as dictionary learning [172], nearest neighbor [61], random forests [93].

Along with the rapid development of deep learning, generative models using neural networks have been proposed to solve multi-domain image translation. As a specific synthesis task, the variety of these models can be categorized into three groups: autoencoder, U-net, and GAN. The architectures have been presented in detail in Chapter 3. Especially, GAN-based models have become popular methods and achieved state-of-the-art in several tasks. Research in [163, 142] is the first work using GAN-based architecture for a multi-domain image translation problem. It provided a promising performance to reconstruct output with descriptive ability augmentation of the generator. However, the limit of these GAN-based methods comes from the one-way translation, which requires many resources for complex tasks. In recent years, CycleGAN [250], conditional GAN (cGAN) [85] or recently the starGAN [42] have been released to address the challenge with strong potential.

In the field of medical images, deep generative models have been rapidly applied and achieved excellent performance in multi-modal MRI synthesis. Sevetlidis et al. [183] proposed an autoencoder architecture for one-to-one MRI synthesis using fused feature maps in a hierarchical design in a patch-based method. The encoder represents the source image in the latent space, and then the decoder reconstructs the target image from these representations. Nguyen et al. [212] introduced a location-sensitive deep network to synthesize T2 from given T1 by integrating image intensity feature and spatial information following principled manners. Huang et al. [83] presented a joint

convolutional sparse coding with weakly-supervised training to solve the cross-modality synthesis in 3D medical imaging.

Besides, there are several also GAN studies applying to medical images. Yu et al. [235] applied cGAN to produce a 3D Flair MRI from given T1. Later, they published another method called edge-aware GANs (Ea-GANs) to handle the discrete generation between slices in the 2D synthesis of cGAN. The model adopts the local feature maps at a global level with 3D estimation [236]. Studies in [45, 229] also used cGAN to synthesize between T1 and T2. Models were designed with the addition of pixel-wise and perceptual losses in overall architecture to perform pairs training and the cycle loss for unpaired training. Out et al. [153] introduced a GAN-based approach to reconstruct MR angiography from T1 and T2-weighted MRI. These methods overcome the limitations of unique correlation between the source and target modalities by using two different modalities, encouraging shared latent representations among multiple source images.

Other researchers use multiple inputs to produce single output can be mentioned such as [96] used an autoencoder to synthesize Flair from T1, T2, and diffusion-weighted imaging, [223] applied 3D CNN to produce Flair from given T1, T2, T1 spin-echo, proton density, and double inversion recovery, or [124] proposed a flexibly GAN to take arbitrary subsets of modalities to generate the target modality. Chartsias et al. [34] used the autoencoder to take all the available MRI modalities as input and simultaneously synthesize one or more missing modalities. Recently, Zhou et al. [249] introduced a hybrid-fusion network consisting of modality-specific, multi-modal fusion, and image synthesis subnetworks to learn the correlations among multiple modalities with enhanced multi-level fusion strategy, thus improving the performance of synthesis.

In terms of CycleGAN and starGAN, few studies have been published and achieved promising performance in multi-modality translation. These models can produce more than one modality using a single model to overcome the drawback of one-way translation of previous methods. Besides, both architectures can learn the features among all image domains and their generalizability to different datasets in which images may be labeled partially. For example, [46, 129] directly applied CycleGAN from [250] to synthesize between T1 and T2 MRI; while [191, 122, 44] used the StarGAN with minor modification to solve the synthesis between several modalities. However, current studies using starGAN are implemented on research datasets without questioning its performance on a practical dataset. Thus, we also want to examine its performance on the practical dataset, in which the difficulty is increased.

6.3 Network architecture

Our objective is to synthesize MRI volume between T1 and T1c MRI. Two selected GAN-based methods are starGAN and CycleGAN. Since there is no difference in initial requirement or data pre-processing between UHF synthesis and cross-modality translation task, the architecture of CycleGAN is kept to examine its performance in this task. The details of CycleGAN have been presented in detail in previous sections; therefore, we do not mention its architecture again. We will focus on starGAN architecture to explore this model's requirement and workflow on cross-modality translation.

Image synthesis aims to produce a modality volume to target MRI volumes through the training process. To regularize symbols for presentating architectures, with a pair of MRI volumes, V^{T1} defines the T1-weighted volume while V^{T1c} is presented for the T1 contrast-enhanced volume, which has the same size by a tensor of size $C \times W \times H \times D$.

6.3.1 starGAN

6.3.1.1 Adversarial network architecture

Following the original starGAN [42], the model aims to translate a given MRI modality into multiple modalities with only one generator and one generator. As a GAN-based method [69], the generator works to synthesize images between domains, while the discriminator predicts whether the output is actual or generated. The generator aims at minimizing errors against the discriminator that tries to maximize them.

In general, a generator can learn to produce any output. However, mapping an input to multiple outputs without complementary information is very difficult to define desired output and hard to implement due to the complex training process. To address this issue, the starGAN model uses an input image along with a defined label as input to synthesize an output. The idea of starGAN is quite similar to the CGAN model (in Section 3.3.2). Labels are represented as one-hot encoding vectors to define the modality. During the training process, the generator learns to flexibly generate the input image along with given random labels. However, different from the CGAN model, the generator is also trained to translate the synthesized image back to input by giving the input modality label.

Hence, in general workflow, the generator responds to two different tasks: synthesizing output from an input image and then reconstructing it back to the original modality image using the input label. With the labels, the generator can produce different output types while the reconstruction process ensures performance. The generative process of starGAN has the form of a cycle like CycleGAN.

Regarding the discriminator, it works to distinguish between the real and synthesized samples as a GAN generator and classify the real image corresponding to its modality label. For a given pair of images and labels, the discriminator returns two values: the probability distributions of the image and its modality label. The general workflow of the starGAN is shown in Figure 6.2

6.3.1.2 Loss function

In starGAN [42], the objective loss function is composed of generator and discriminator losses, including three main functions. The generator G aim to translate the input V^{T1} into an V^{T1c} with the target modality label L^{T1c} such that $G(V^{T1}, L^{T1c}) \rightarrow \hat{V}^{T1c} \approx V^{T1c}$. The discriminator D produces probability distributions on image volume and modality label from a given input, for example, $V^{T1} \rightarrow \left\{D_{img}(V^{T1}), D_{label}(V^{T1})\right\}$.

As a GAN method, [69], the adversarial loss $\mathcal{L}_{adv}$ is always used in every model. Besides, via the reconstruction process of the generator, the term of reconstruction loss $\mathcal{L}_{rec}$ are also applied to force the generator to synthesize images that appear to be identical to their inputs corresponding to the domain. Besides, in starGAN [42], a single discriminator works to classify the label along with the given input, and to optimize both D and G. Hence, t be more clear, the value of modality classification loss $\mathcal{L}_{cls}$ are divided into two sub-functions: a loss function to optimize discriminator $\mathcal{L}_{d_cls}$, and a loss function to optimize generator $\mathcal{L}_{g_cls}$.

In general, the objective loss function $\mathcal{L}_{starGAN}$ can be defined as:

$$\mathcal{L}_G = \mathcal{L}_{adv} + \lambda_{cls}\,\mathcal{L}_{g_cls} + \lambda_{rec}\,\mathcal{L}_{rec} \tag{6.1}$$

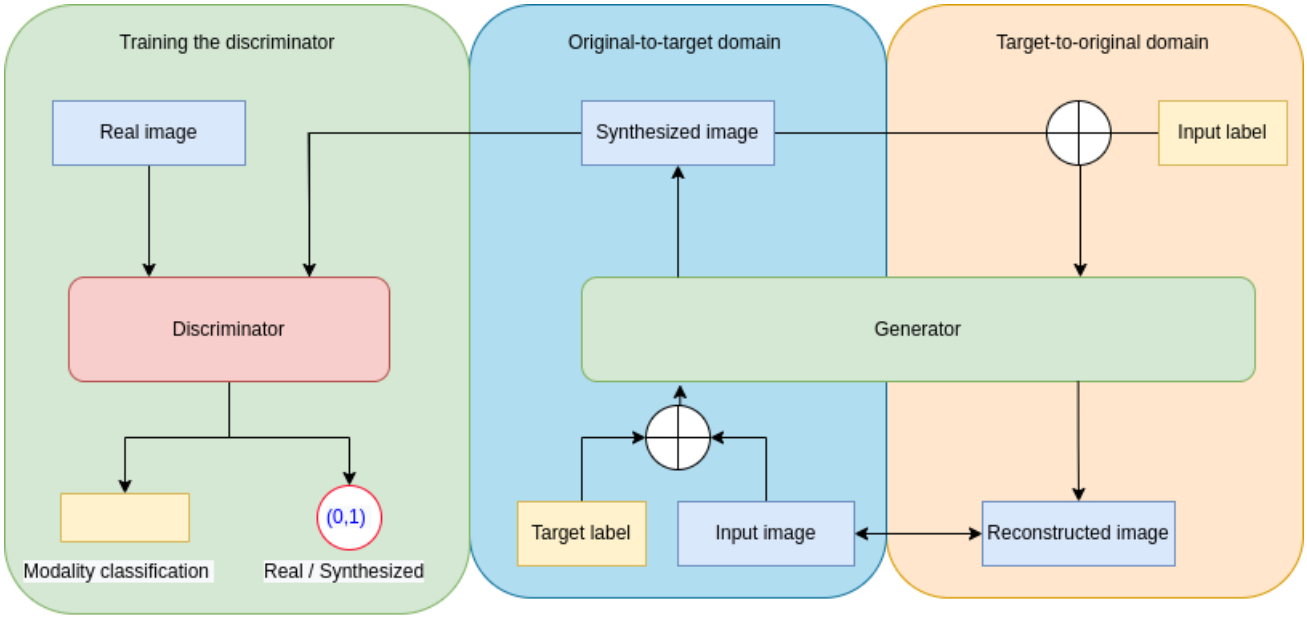


Figure 6.2: Illustration of starGAN schematic, consisting of one generator and one discriminator

$$\mathcal{L}_D = -\mathcal{L}_{adv} + \lambda_{cls}\,\mathcal{L}_{d_cls} \tag{6.2}$$

where λ_{cls} and λ_{rec} are hyper-parameters that of modality classification and reconstruction loss, respectively.

Adversarial loss: The concept of adversarial loss is applied on every GAN-based architecture. Training with adversarial loss solves the min-max problem exposed in the previous section. The adversarial loss of the generator can be presented as:

$$\mathcal{L}_{adv} = E_{V^{T1}}[\log D_{img}(V^{T1})] + E_{V^{T1},L^{T1c}}[\log(1 - D_{img}(G(V^{T1}, L^{T1c})))] \tag{6.3}$$

where G uses input image V^{T1} and the modality label L^{T1c} to generate a synthesized image $G(V^{T1}, L^{T1c}) \rightarrow \hat{V}^{T1c}$ that is expected to look as similar as possible to images V^{T1c}, while D aims at distinguishing between generated samples $\hat{V}^{T1c}$ and real samples V^{T1c}. G minimizes this objective cost against an adversary D that tries to maximize it.

Reconstruction loss: In general, the adversarial loss can control the learning mapping of generators to produce outputs identically distributed as target domains [67]. However, Zhu et al. [250] proved that minimizing only adversarial loss cannot guarantee that translated images preserve the content of their input images while changing only the domain-related part of the

inputs. The terms of reconstruction loss in starGAN are similar to the concept of cycle consistency loss in CycleGAN to reduce the space of possible mapping functions:

$$\mathcal{L}_{rec} = E_{V^{T1},L^{T1c},L^{T1}}[\left\|V^{T1} - G(G(V^{T1}, L^{T1c}), L^{T1})\right\|_1] \tag{6.4}$$

where generator G takes in the translated image $G(V^{T1}, L^{T1c})$ and the modality label L^{T1} as input to reconstruct the $\hat{V}^{T1} \approx V^{T1}$. Thus, in the starGAN model, for each generating process between classes, there always exists the a cycle to secure the generating process between two modality.

Modality Classification Loss: For a given input V^{T1} and a target modality label L^{T1c}, the goal is to translate V^{T1} into an output image V^{T1c} , which is properly classified to the target modality L^{T1c}. To measure the label classification, the discriminator imposes a loss when optimizing both networks during training. The loss of modality classification in starGAN is decomposed into two functions: the modality classification loss for real samples used to optimize the discriminator $\mathcal{L}_{d_cls}$ and another loss of synthesized volumes used to optimize the generator $\mathcal{L}_{g_cls}$.

In detail, the modality classification loss for discriminator is defined as:

$$\mathcal{L}_{d_cls} = E_{V^{T1},L^{T1}}[-logD_{label}(L^{T1}|V^{T1})] \tag{6.5}$$

where the term $D_{label}(L^{T1}|V^{T1})$ represents the probability distribution over modality labels computed by D. Consider that the training data contains the input image and modality label pair (V^{T1}, L^{T1}). With the aim to minimize the objective function, the discriminator learns to classify a real image V^{T1} to its corresponding original modality L^{T1} [42].

On the other hand, the loss function for the modality classification of generator is defined as:

$$\mathcal{L}_{g_cls} = E_{V^{T1},L^{T1c}}[-logD_{label}(L^{T1c}|G(V^{T1}, L^{T1c}))] \tag{6.6}$$

where G tries to minimize this objective to generate images that can be classified as the target label L^{T1c}.

6.3.1.3 Generative network architecture

We built the network based on the original starGAN [42]. The generator uses residual blocks [75] as building units for feature extraction. It has been proven to reduce computational time while ensuring the model performance with fewer parameters. First, the label as a one-hot vector is spatially replicated to have the same size as the input, and then it will be concatenated with the input volumes to form the input of the generator. At the beginning of the network, the first two convolutional layers extract shallow features from the input to reduce size and increase depth. Next, Residual blocks work as building units to synthesize information from feature maps. Two deconvolutional layers are ultilized to reconstruct the feature map back to the original size before generating the synthesized output at the final layers. The details of 3D generator implementation are shown in Figure 6.3.

The structure of discriminators is shown in Figure 6.4. Unlike previous architecture, the starGAN discriminator performs the classification between the input and generated output and

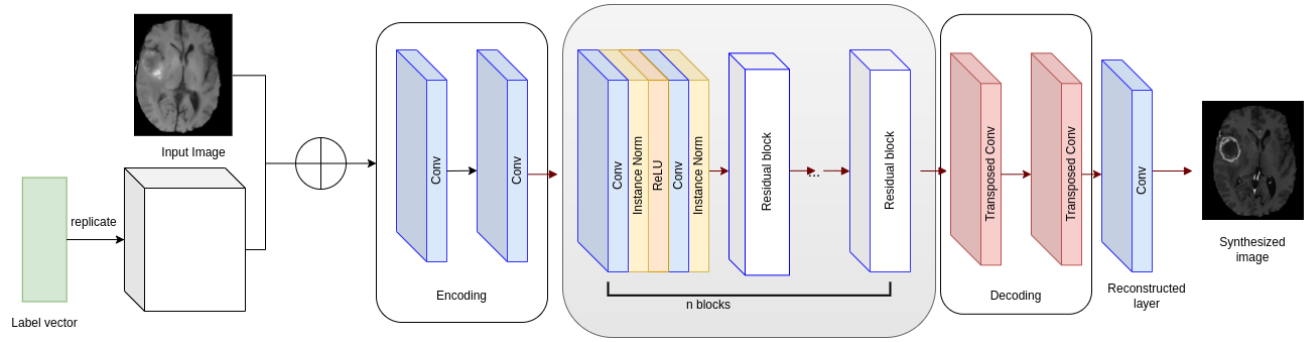


Figure 6.3: The architecture of the generator.

identifies the modality label simultaneously. As usual, a convolutional layer with a single output channel is used to produce values ranging from 0 to 1. Besides, the architecture of the two discriminators is modified with an additional convolutional layer to form the modality classification vector. The discriminator returns two separate outputs for a given input.

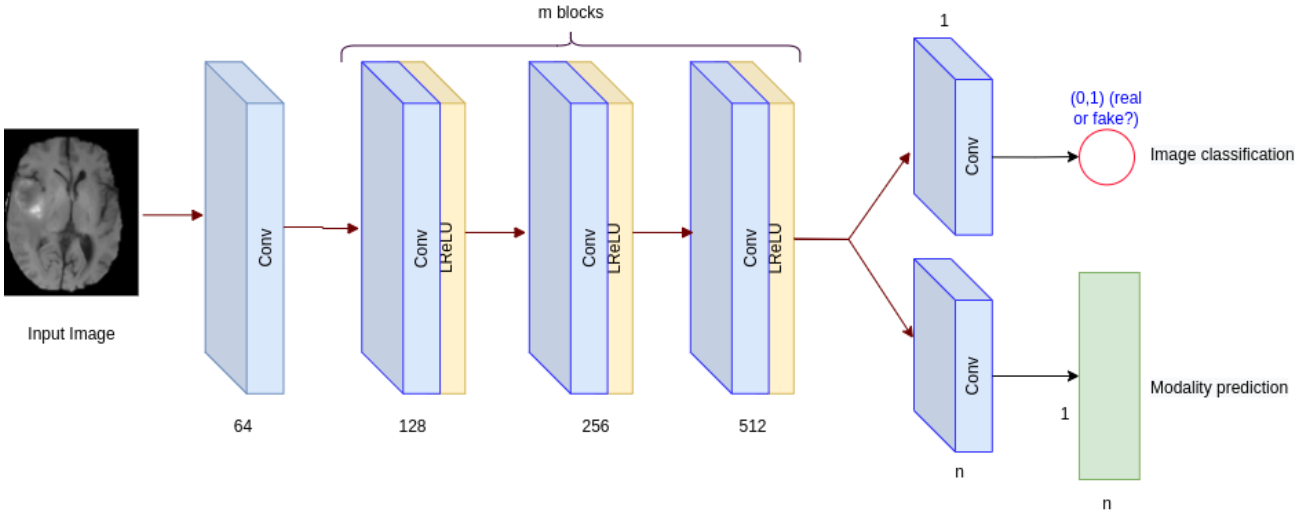


Figure 6.4: Architecture of discriminator.

6.4 Experiments

6.4.1 Dataset

As mentioned above, we want to examine the performance of two different models on cross-modality synthesis on both research and practical data. For this task, we currently focus on T1 and T1c MRI conversion due to its benefits and the scope of the thesis. Then extend to different types of modality.

In terms of the research dataset, again, we use the MICCAI BraTS 2018 [140, 13, 12] because it already has aligned T1-weighted, T2-weighted, T1 contrast-enhanced, and T2-FLAIR. The detail of the BraTS dataset was presented in Section 4.4.1. There are several studies that used the BraTS dataset for synthesis task [44, 230], but we re-implement the selected methods on this dataset for

a comparative purpose.

With the support of CHU Poitiers, we have built another practical dataset (appearing later as the *synCHU dataset*). A total of 290 pairs of T1 and T1c, 3D MP-RAGE brain MRI are collected for this work. All 3T MRI volumes were acquired from a Siemens Magnetom Skyra scanner, at 0.9mm^3 voxel spacing with the field of view $240 \times 288 \times 192$. The variety of samples includes different types of diseases. The number of volume pairs in the training, validation, and test set is 200:40:50, respectively.

Besides, during experiments, we found problems translating subjects with glioma in the synCHU dataset compared to performance on BraTS. Hence, we create a sub-dataset from synCHU that contains glioma subjects and other subjects with a ratio of 50:50 to enhance the performance of translating T1c with glioma. It contains 60 samples, divided into training, validation, and test set with size 40:10:10, respectively.

6.4.2 Pre-processing

Unlike previous work on synthesis, all samples in both dataset have been well aligned, hence we take advantage of aligned pairs to perform these methods. Besides, to pre-process samples for training, we applied standardization and normalization methods to reduce the adverse impact of signal variation across different scanners and sites. Training samples in a modality are standardized with other samples in the same modality using histogram standardization to make voxel values in the same intensity range. Then, all aligned samples are normalized using z-normalization to reduce the computation cost and avoid model divergence. The z-normalization uses terms of mean and standardization to re-scale transform voxel value within range [0, 1].

In terms of augmentation, the variety of synCHU and BraTS datasets has been ensured by the dataset size. We do not need to deploy any techniques on this dataset. on the other hand, we applied techniques such as random rotation and flipping on the subset of the synCHU dataset to extend the size of the dataset, increase variability, and avoid over-fitting before training

6.4.3 Training setup

For each method, we did several experiments to produce T1-T1c MRI on three datasets: BraTS, synCHU, and a subset extracted from the synCHU dataset.

6.4.3.1 CycleGAN

The architecture of model is similar to CycleGAN used for UHF synthesis, where two generators are mirrored with the removal of unnecessary connections between hierarchical features to reduce the dependence of output on input while speeding up the training process. Besides, model complexity will increase along with the increase of depth or input size. Since we have reduced several unnecessary fused connections and the concatenation of the shallow feature, the model configuration can be enhanced to improve the performance. In this work, each generator contains six RDBs, where each block includes three dense blocks in residual connection. Training is executed in pairs, allowing the network to focus on mapping details and optimizing the performance.

Model is also trained on patches. For each batch, patches are randomly extracted into a

maximum size of $64 \times 64 \times 64$ on T1 and T1c volumes. The ADAM optimizer is adapted for optimization. The learning rate starts at $2e^{-4}$ and decays by half every five epochs.

For all datasets, the batch size is also set as four. The learning rate is initialized to $1e^{-4}$, and decay starts after every ten epochs. The training CycleGAN for BraTS, synCHU, and subset dataset takes an average of 10, 12, and 6 hours respectively, on a GPU NVIDIA A100 40GB for 100 epochs.

6.4.3.2 starGAN

The generator is built following the original paper [42]. The generator contains 16 residual blocks, where each block includes two convolutional layers followed by an instance normalization layers and a ReLU activation function in residual connection. The modality labels are pre-defined during pre-processing part. The depth-wise concatenation between input and label is executed during the training phase. The discriminator consists of several convolutional layers followed by the Leaky ReLU activation function to extract the feature. The final layers include two independent convolutional layers. The first convolutional layers return the distributed probability of the given input, while the second layers return a vector with a probability of domain classification.

StarGAN is also trained on patches. For each sample, patches with a size of $64 \times 64 \times 64$ are randomly extracted. The batch size is set to 4. Adam optimizer with a learning rate of $1e^{-4}$ was used for training the model for 100 epochs. On average, the training time of starGAN for BraTS, synCHU, and subset dataset is 15, 18, and 9 hours on GPU NVIDIA A100 40GB.

6.4.4 Evaluation metrics

Similar to the previous tasks, we use peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) metrics to evaluate the image quality between ground-truth and generated MRI volumes in both tasks.

6.5 Results

Within the research objective, we do a comparative study between our proposed CycleGAN and the current baseline starGAN methods for cross-multimodal translation. The BraTS is used as a reference to explore performance on a research dataset. In contrast, synCHU and its subset is considered as practical dataset.

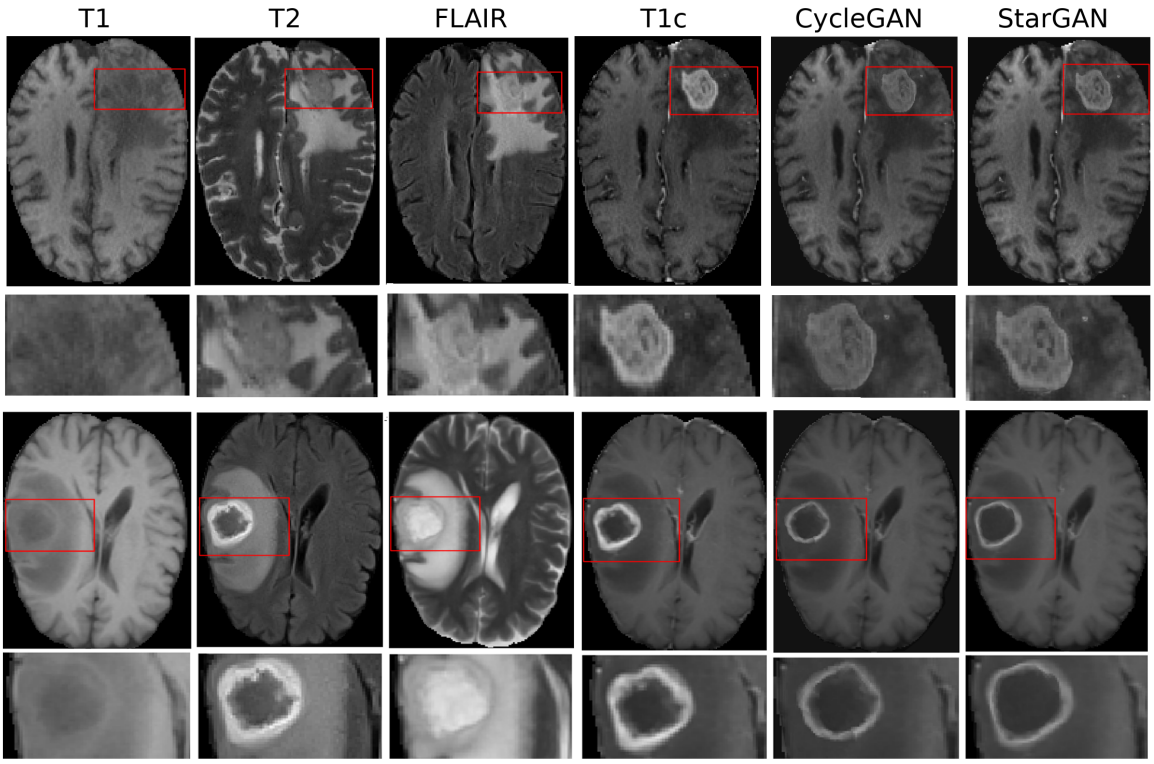


Figure 6.5: Visualization of T1-T1c translation on HGG object in BraTS dataset. In each image, from left to right: T1 input, ground-truth T1c, CycleGAN and starGAN output.

	CycleGAN		starGAN	
	SSIM	PSNR	SSIM	PSNR
T1 to T1c	0.9434	33.76	**0.9682**	**35.92**
T1c to T1	0.9321	31.11	**0.9496**	**32.24**

PSNR, peak signal-to-noise ratio; SSIM, structural similarity index.

Table 6.1: Average value of PSNR (dB) and SSIM for T1-T1c translation on BraTS dataset. Synthesized images generated by CycleGAN and starGAN are compared with corresponding ground-truth MRI for quantitative evaluation.

The performance of CycleGAN and starGAN on BraTS are summarized in Table 6.1 in terms of PSNR and SSIM assessment. It quantitatively shows how starGAN achieves better results than CycleGAN in the BraTS dataset in both directions (T1 to/from T1c). The SSIM/PSNR of starGAN reaches 0.9682/35.92 on T1-to-T1c and 0.9496/32.24 on T1c-to-T1 conversion. The result of our starGAN is quite close to the previous study that also applied starGAN for cross-modality translation [44]. On the other hand, the results of CycleGAN also presented a very competitive result. The SSIM/PSNR of the CycleGAN model reaches 0.9434/33.76 on T1-to-T1c and 0.9321/31.11 on T1c-to-T1, in which we can conclude that the model can work well on research data.

Besides, visualization in Figure 6.6, Figure 6.7 demonstrated that both models can fairly produce T1c from T1 with low distortion of tumour objects. The shape and positions of tumors

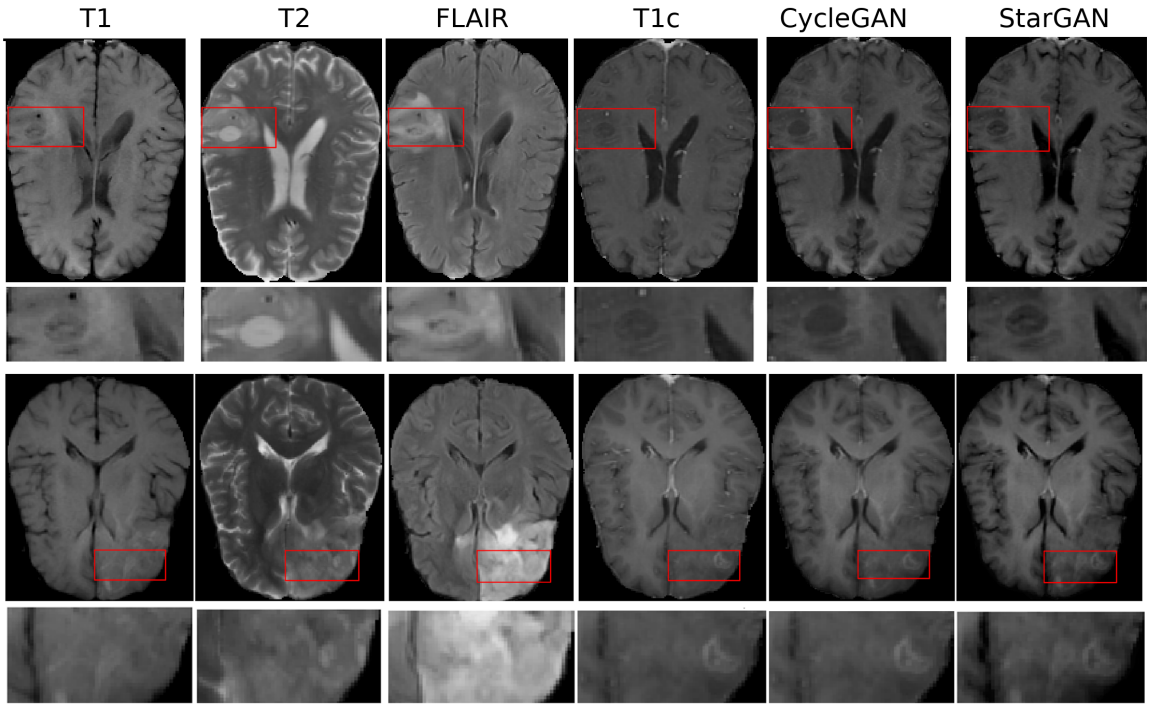


Figure 6.6: Visualization of T1-T1c translation on LGG object in BraTS dataset. In each image, from left to right: T1 input, ground-truth T1c, CycleGAN and starGAN output.

in T1c ground-truth are closely reconstructed on synthesized output.

	CycleGAN		starGAN	
	SSIM	PSNR	SSIM	PSNR
T1 to T1c	**0.8715**	**29.44**	0.8562	27.22
T1c to T1	**0.9021**	**30.70**	0.8677	28.19

PSNR, peak signal-to-noise ratio; SSIM, structural similarity index.

Table 6.2: Average value of PSNR (dB) and SSIM for T1-T1c translation on synCHU dataset. Synthesized images generated by CycleGAN and starGAN are compared with corresponding ground-truth MRI for quantitative evaluation.

Table 6.2 presentes results of two models on the synCHU dataset. The SSIM/PSNR of synthesized T1 for CycleGAN and starGAN are 0.8715/29.44 and 0.8562/27.22, respectively. Here, we notice a performance reduction in producing the translation of both models. In general, the reduction in performance of approaches in practice is usually due to the difference between practical and research data in terms of pre-processing and difficulty.

Figure 6.8 shows synthesized T1c from T1 MRI on non-glioma samples. In fact, in almost all cases, output of both models is quite clear compared to ground-truth with no change in the structure. Moreover, CycleGAN provides a better result on contrast points reconstruction, while it is limited on starGAN.

Besides, we also found that two models have failed to reconstruct T1c from T1 on glioma subjects. Figure 6.9 demonstrated the output of CycleGAN and starGAN on a case with glioma.

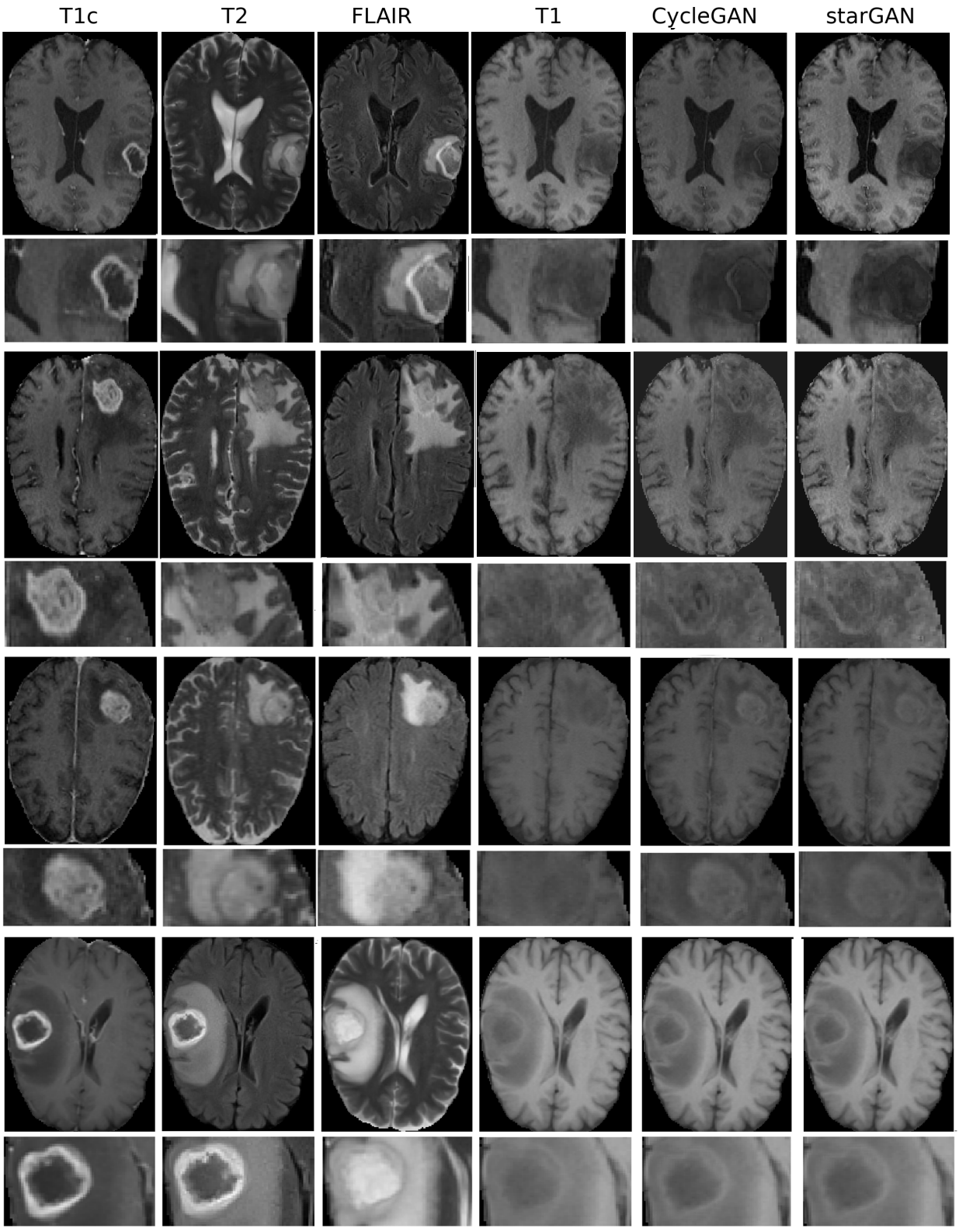


Figure 6.7: Visualization of model performance on BraTS dataset on T1c-T1 conversion. In each image, from left to right: T1 input, ground-truth T1c, CycleGAN and starGAN output.

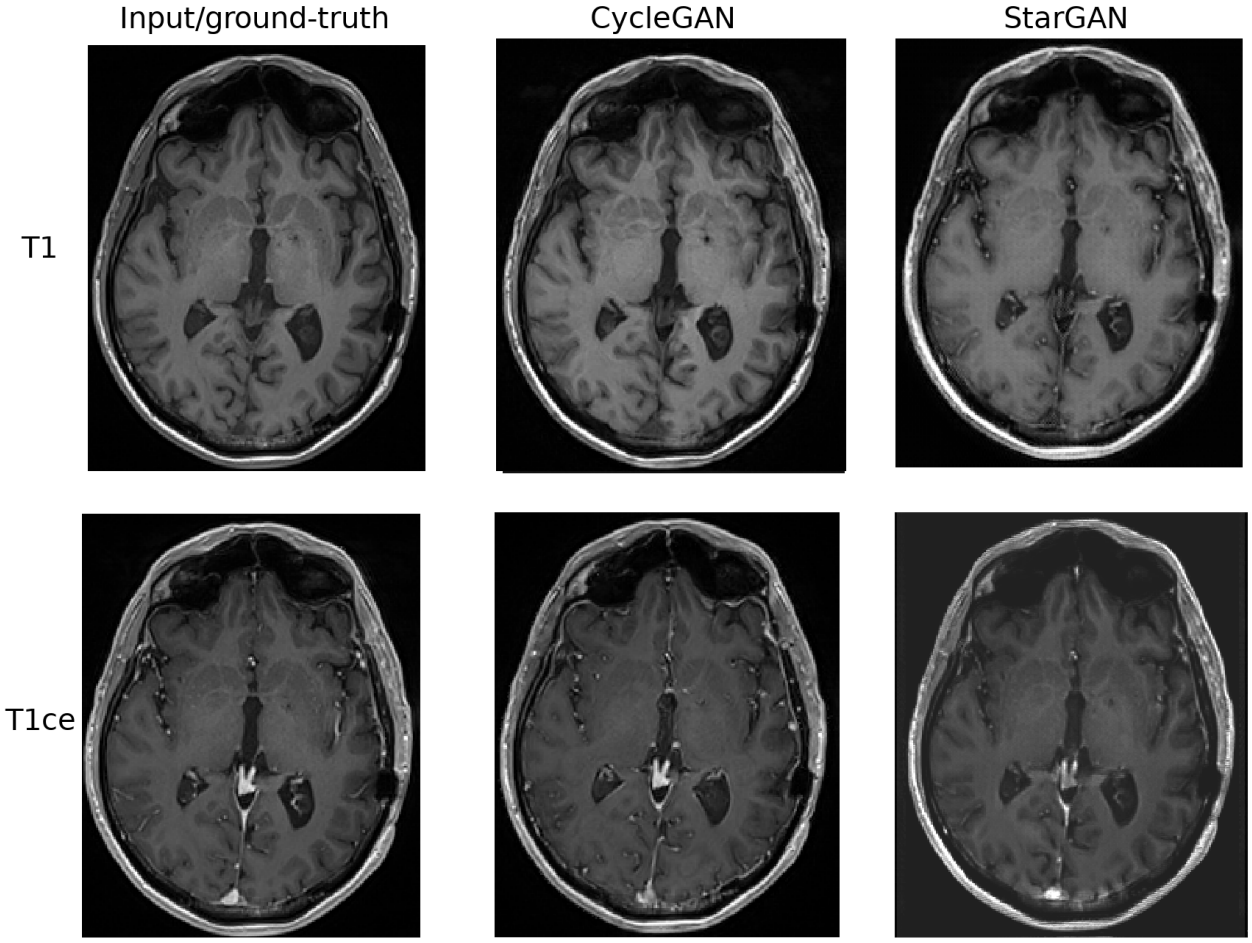


Figure 6.8: Visualization of model performance on synCHU dataset. In each image, from left to right, top to bottom: input/ground-truth T1/T1c MRI, synthesized T1c/T1c generated by CycleGAN and starGAN.

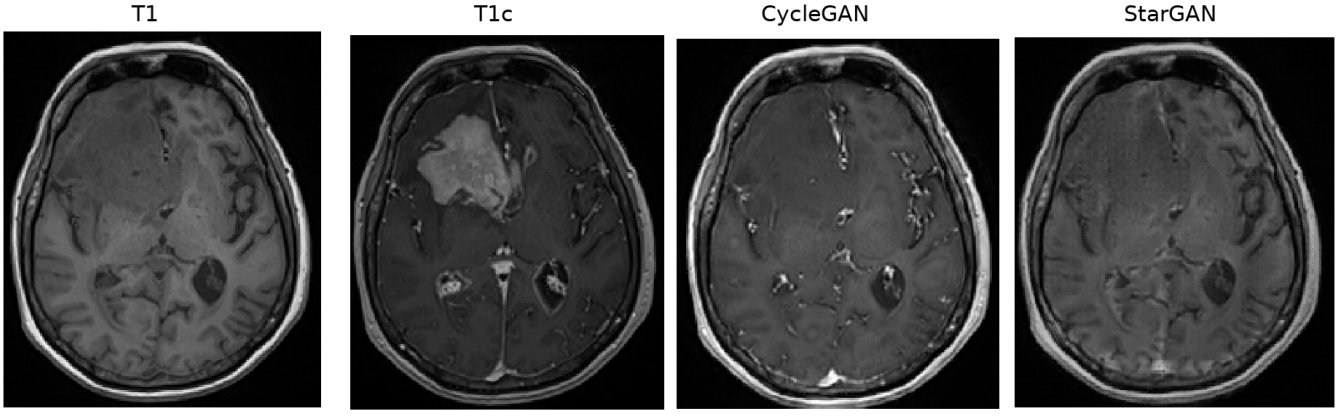


Figure 6.9: Failed recontruction of model on glioma subjects from T1 to T1c. From left to right: T1 MRI input, T1c ground-truth MRI, synthesized output by CycleGAN and starGAN.

As can be seen from the figure, both models cannot recognize the tumour location and shape it in the output.

Through different experiments, we found that the problems come from the unbalance between the number of gliomas and other samples in the synCHU dataset. There are only 10% of samples containing clear tumour on a total of 290 samples which consists of different diseases. Thus, we extracted a sub-dataset from synCHU and re-trained models to focus on MRI with tumour.

	CycleGAN		starGAN	
	SSIM	PSNR	SSIM	PSNR
T1 to T1c	**0.9257**	**32.21**	0.8821	29.53
T1c to T1	**0.9112**	**31.88**	0.8845	28.79

PSNR, peak signal-to-noise ratio; SSIM, structural similarity index.

Table 6.3: Average value of PSNR (dB) and SSIM for T1-T1c translation on subset dataset. Synthesized images generated by CycleGAN and starGAN are compared with corresponding ground-truth MRI for quantitative evaluation.

Table 6.3 presentes the quantitative evaluation of methods on the subset dataset. The model performance has significantly improved, with the SSIM/PSNR reaching 0.9257/32.21 on CycleGAN and 0.8821/29.53 on starGAN. The visualization in Figure 6.10 illustrates that both models could reasonably locate and shape the tumour in synthesized T1c. Although the results still need to be improved, it has been increased a lot from the previous experiments on the entire dataset.

6.6 Discussion

In this work, we compared CycleGAN and starGAN - two recent two state-of-the-art methods for cross-domain translation. Methods aim to learn the similarity and differences between different MRI modalities for synthesis purposes. Models are performed on research (BraTS2018) and practical dataset (synCHU) for evaluation and currently focused on T1 from/to T1c conversion. However, it should be noted that methods can synthesize multiple MRI modalities. Models have presented promising results in the cross-modality translation task. However, limitations are clearly observed in visual assessment and measurement values.

The T1-T1c translation is considered one of the most challenging conversions due to the appearance of contrast-enhanced points in T1c. In the aspect of computer vision, it can appear everywhere in the T1c; hence fully synthesizing T1c modalities need to be carefully studied and researched. That explains why metric evaluation in SSIM/PSNR is usually lower than previous work, such as super-resolution and UHF synthesis. In super-resolution, the predicted output is generally similar to the input in terms of image properties (e.g., brightness, contrast) and structural information, with a significant improvement in spatial resolution and details. Regarding to UHF synthesis, although there might be a mismatch between paired 3T and 7T due to the alignment, in general, the structural information between pairs is similar. In other words, for both tasks, there will be no appearance of new "artifacts" such as contrast points or tumors in the final output like the T1-T1c task. We have compared the performance of starGAN on the BraTS dataset with other studies on the same topics [44, 230] as a reference for our implementation.

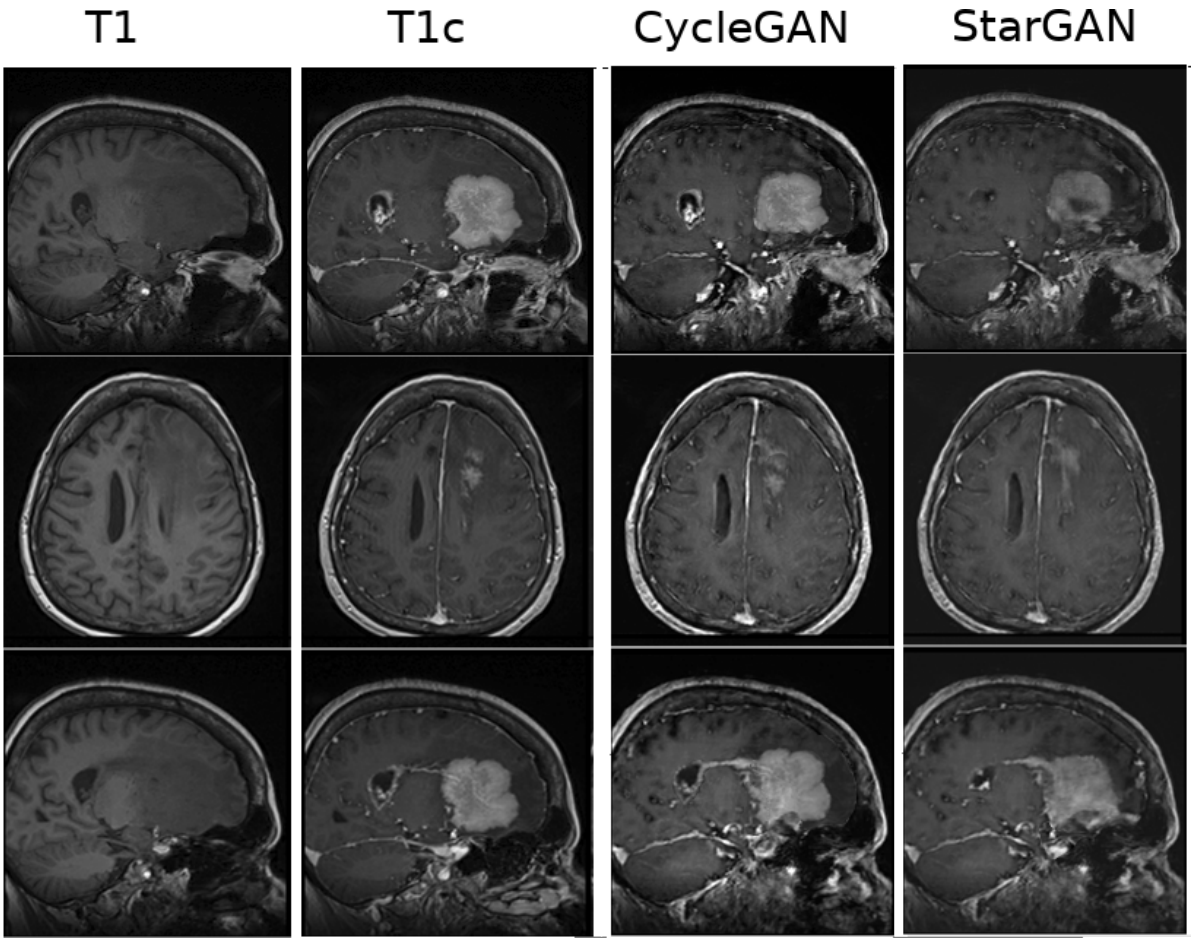


Figure 6.10: Visualization of model performance on subset dataset. From left to right: T1 MRI input, T1c MRI, CycleGAN and starGAN output.

Minor differences in quantitative results might come from configurations in network architecture or training parameters, but the results are close in general.

Besides, both methods can fairly synthesize T1c from/to T1 MRIs. Evaluation in PSNR / SSIM and visualization on the BraTS dataset have shown that starGAN can provide higher performance than CycleGAN. However, the performance of both methods has significantly decreased on the synCHU dataset. We observed detailed results and close to ground truth on non-glioma subjects. The main structure of objects is kept, while contrast points can be reconstructed, especially on the CycleGAN model. However, both models failed to reconstruct samples with glioma.

Through different experiments, we have figured out the problems resulting from the imbalance between glioma and non-glioma samples in the entire dataset. By doing experiments on the subset dataset with more representative glioma samples, we have improved the prediction process from previous work. Both models can reconstruct T1c with the appearance of glioma from the given T1 MRI. Although experimental results are still limited in terms of locations and shape of the tumor compared to the ground truth, we believe that it can be improved in the following experiments. For example, we can use another subset that only contains glioma samples like the BraTS dataset as training data. Besides, the performance of CycleGAN can surpass the starGAN in the synCHU or subset dataset although its performance is lower than the starGAN in BraTS.

In fact, in BraTS, all samples are subjects with glioma, while this number on practical dataset is much lower. Hence, we temporally conclude that the CycleGAN performs better on practical datasets than the starGAN. To explain, the adversarial of CycleGAN are implemented in pairs, which usually can optimize results. It is explicable since the CycleGAN method incorporates cycle-consistent design in forwarding and backward cycles with paired training, taking full advantage of structural information and thus leading to comparable generation results. It has been proved through the MRI synthesis in previous work, where the dataset is limited by size and variety.

On the other hand, starGAN only uses one label to classify the T1 and T1c MRI. It might be too general to define a sample. The potential of starGAN for multi-modality translation is very considerable. With the presentation of modality labels in the architecture, starGAN can generate specific output on corresponding label contents. Furthermore, a starGAN model can use several labels to determine the different properties of a sample. For example, to improve the performance of starGAN on the synCHU dataset, we can use another label along with the modality label to pre-define glioma samples. Then, during the training phase, the model has to consider that there is a second condition to synthesize the desired output. However, doing so will significantly increase the complexity and depth of the network.

At this moment, this work is ongoing. The current step is to compare the performance of two models to examine the performance of our prosed CycleGAN model, the cross-modality synthesis task, while starGAN now is considered the state-of-the-art domain. In future work, we want to increase the current CycleGAN model's performance by attempting different generator architectures. In terms of starGAN, we also want to enhance its performance by modifying the general architecture or re-training the model with multiple labels to specify training samples for details generation. After that, we aim to expand the work to other modalities.

6.7 Conclusion

In this work, we did a comparative study of our proposed CycleGAN and the starGAN for the T1-T1c conversion. Methods have been evaluated using research (BraTS) and practical datasets (synCHU), in which MRI modalities were spatially co-registered. Experiments on research dataset have presented that starGAN achieves better performance compared to CycleGAN, with equivalent results to recent studies in the same topic. On the other hand, CycleGAN provided higher synthetic performance on a practical dataset, where the differences between samples and difficulty are increased.

In general, both methods can enhance the capability of cross-modality MRI translation to tackle the challenges from a clinical context. With the advantage of robust implementation, this topic is very potential in practice where multi-contrasts MRIs are necessary for better diagnosis and treatment planning.

Conclusion

Chronic kidney disease (CKD) describes the gradual loss of renal function. In the context of the increasing CKD as a problem of significant public health, kidney transplantation is an efficient strategy compared to all the strategies evaluated with the benefits of lifetime and treatment cost. In general, CKD patients are known to experience accelerated atherosclerosis, and there are several inflammatories and atherogenic factors due to the increased cardiovascular risk proportional to the increase in serum creatinine, suggesting that renal failure correlates with, if not causes, accelerated vascular and metabolic defects that predispose patients to cardiovascular death. Patients with ESRD benefit from transplantation as early as possible to maximize their potential for extended survival after transplantation. The better outcomes of patients with preemptive transplants and a shorter dialysis time underscore the importance of early referral and evaluation for renal transplantation.

Before transplantation, the determination of the functional status is critical. Standard methods such as estimating the GFR or invasive biopsy are a late indicator of renal impairment. Moreover, the optimal method of preserving the kidney remains an entire problem. If a visual examination is essential, the criteria used to reach the right decision are still vague and challenging to explain. On the other hand, radiology was reported to help with follow-up renal diseases. Medical images such as MRI and CT play an increasingly more critical role in assessing renal function. Especially, MRI was demonstrated to be a powerful tool to evaluate renal tissue by assessing both renal function and structure for both kidneys with structural, functional, and molecular information. The use of imaging techniques to analyze the kidney in different clinical tasks, including that transplantation, is becoming a potential research topic. MRI modalities offer safe, low-cost, non-invasive, clinically available, and short time examination techniques. Those imaging techniques could provide a rapid assessment for both transplanted kidneys and grafts. Thus, developing a robust and non-invasive alternative to imaging is, therefore, a subject with many challenges.

Besides, machine learning, and later deep learning, has been coming into its own, with great adaptability to high dimensionality problems. Machine learning can be defined as a set of algorithms that can learn and improve from experience without being explicitly programmed for a specific task. Learning-based enables the computer to build complex concepts out of simpler concepts. Recent medical imaging and machine learning advantages inspired many innovative researchers to use anatomical and functional imaging for diagnostic assistance. AI models are not approved to replace radiologists medical decision-making; instead, they can assist them in providing optimal diagnoses for their patients.

Within the scope of the thesis, we aim to improve the assessment of renal grafts before removal using medical imaging techniques, focusing on MRI. Besides, the objective of the thesis is to improve kidney assessment to support radiologists making decisions by enhancing the quality of MRI in terms of spatial resolution, field strength, and multi-contrast.

However, since the COVID-19 epidemic took place during most of the thesis work, as well as the insufficiency of available resources, research works are implemented on brain MRI to achieve optimized results. Since we do not focus on any specific disease in general, but to improve the overall quality of MRI, we believe that it can also help improve the quality of kidney assessment. Hence, we decided to use brain MRI instead of kidney MRI to achieve the most optimized results.

There are three main research tasks have been proposed and solved in this thesis, including super-resolution and UHF synthesis for image quality enhancement; and cross-modality translation. In each task, we present an overview of related works used in the domains along with our contribution.

In terms of data, the works are completed on both research and practical dataset. The MICCAI BraTS2018 - a dataset containing 3T MRI volumes with different types of sequences is used as research data. The variety within BraTS is ensured by samples acquired with various clinical protocols and scanners from multiple institutions. It contains several samples, including T1-weighted, T2-weighted, post-contrast T1-weighted, and T2-FLAIR. It is used for super-resolution and cross-domain translation tasks. On the other hand, with the support of MRI systems at CHU Poitiers, we are allowed to extract medical data for research works. Two datasets have been built for the three tasks. The first dataset contains different samples of 3T and 7T MRI, mostly unpaired, and a few samples are paired for quality enhancement tasks. The second dataset contains pairs of 3T MRIs with different modalities (currently T1 and contrast-enhanced T1) used for the cross-modality task.

Chapter 4 presents a method that enhances the spatial resolution on routine 3T MRI to improve diagnosis and assessment. When most current approaches in MRI super-resolution require paired low and high-resolution data for training, which are difficult to obtain due to the limited resource and time computation, the proposed the SRCycleGAN to solve the super-resolution task through either paired/unpaired training based on the self-supervised designs. The architecture of generators is modified from the base model to improve the performance of medical data. Besides, the methods are implemented for both 2D and 3D MRI.

Experiments have been conducted on research and practical MRI datasets to ensure model performance. Results figured out that the proposed methods can work stably on different types of MRI. Evaluation of reconstructed images on both 3T and 7T MRI shows exploitable results with low distortion and detailed texture. The advantage of the approach is that paired data is not required for an efficient training process. Therefore it can be executed on several publicly available MRI datasets, thus overcoming the limitations explained earlier.

Besides, we also compared model performance with different methods in the same domains to have an objective perspective on model performance. Selected methods include the traditional super-resolution (interpolation) and state-of-the-art GAN-based model. The quantitative evaluation shows that the SRCycleGAN is better and more measured than other methods on different scaling factors. Besides, we also conclude that a 3D model can provide better performance than a 2D model in general due to the unity between slices and the entire volume. It also can reduce the appearance of noise/artifacts or even the gap between the object in reconstructed results.

Chapter 5 indicated the quality enhancement process by synthesizing UHF (7T) MRI out of routine (3T) MRI. Synthesis holds great potential by relying on standard MRI data pairs (*e.g.* 3T and 7T) to emulate conversion on a physical level. This work presented a hybrid generative model based on CycleGAN to produce realistic UHF-MRI. Advantaged by semi-supervised learning in cycle-consistent architecture, the proposed method can solve synthesis on routine MRI in 3D

space. The architecture of the generator is modified to work on the synthesis task.

A substantial drawback for MRI synthesis is the difficulty in having reliable training data such as pairs of 3T and 7T MRI content for the same subject acquired simultaneously. With the CHU dataset that has been built from MRI systems at CHU Poitiers, we keep using 3T and 7T MRI to implement the MRI synthesis for the 3T-to-7T conversion task. Since our training data is limited for this task, models are implemented on true data pairs to maximize the performance. Experiments are done on experimental MRI data to evaluate the performance of the dataset. Real 3T and 7T MRI are processed and used for experiments.

Results from a comparison among different methods demonstrate that the method outperforms current state-of-the-art methods in MRI synthesis to produce UHF-MRI under qualitative and quantitative terms. The convincing results also prove the advantage of cycle-consistent medical image synthesis, overcoming the limitation in training data for MRI synthesis. The final models can work stably on 3D brain MRI, holding great promise for the research topic in practice.

In chapter 6, we focus on MRI cross-modality translation task. Models can generate output among different modality such as T1$\leftrightarrow$ T2, T1 $\leftrightarrow$ T1c, T1 $\leftrightarrow$ T2-Flair. However, we currently focus on translation between T1 and T1 contrast-enhanced brain MRI due to its research value for overall assessment in general and for kidney patients in particular.

We do a comparative study between methods in cross-modality translation among T1-T1c MRI. Among all generative architecture that has been proposed, we implement two cross-modality frameworks based on GAN to illustrate its effectiveness on experimental data. The first method in this work is the CycleGAN model. We take advantage of the proposed methods in UHF synthesis by adopting them for cross-modality transformation. The second framework is StarGAN - a recently novel method for image translation. StarGAN uses only one unified model with one generator and one discriminator to perform image-to-image translations between multiple domains. We adopted the strategy of StarGAN and specially applied it to multimodal MR image synthesis. In this work, we build a 3D implementation of StarGAN to solve the T1-T1c conversion task.

The experiments have also been conducted on research and practical datasets. The research dataset results are well-performed compared to the previous study in the same domain. The evaluation is executed on the entire dataset and a subset dataset focusing on a specific task for a practical dataset. Results demonstrated that both selected methods achieved a promising result on usual subjects, while the performance is strongly reduced in samples with glioma due to the unbalance of data. After retraining on the subset dataset, both models have significantly improved the performance with fair reconstruction on samples with tumors.

Chapter 7

Perspectives

Regarding the objective of the thesis, the first future work is to investigate the performance of proposed methods on kidney data. We expect the model to perform well for kidney MRI for different tasks, including super-resolution, UHF synthesis, and T1-T1c conversion. If it can perform well, it can bring a huge value to thesis work in practice. On the other hand, we can apply transfer learning and domain adaptation to current work to expand performance on kidney data instead of re-train all models. As mentioned in section 2.5, models can be fine-tuned with new additional kidney samples for each task. Transfer learning aims to apply a solved problem to a different but related problem. In this case, we consider that MRI for different pathology has the same base properties, such as acquisition and modalities, with the difference in structural information.

In the second task, we want to expand the training dataset size to explore and evaluate the performance of the proposed method. Indeed, the current dataset for UHF is limited. We consider this research topic precious, while the results hold great potential for synthesizing UHF from a given 3T MRI. The quality is significantly enhanced in contrast and structural details, allowing better support for diagnostic aid. Besides, the University Hospital of Poitiers imaging platform, especially via the Telsa MRI 7 and its acquisitions, opens the way for these studies.

In the third task, the work is currently in the comparative study phase, with much opportunity to enhance the performance of both models. Since we have worked on generative models for almost the time of the thesis, we believe that we can improve our current work with several solutions. Later, we can extend the study to different modalities, such as T1 $\leftrightarrow$ T2, T1c $\leftrightarrow$ T2, etc.

Overall, our objective is not to replace radiologists medical decision-making; instead, we aim to assist them in providing optimal diagnoses for their patients. Furthermore, each work of the thesis is completed based on this objective. The processing of images can be implemented independently for each task or linked into a chain based on a different purpose. The model can produce output with desired content from given input without any additional process. All the tasks take less time than the usual time of MRI scanners, as well as the current expense for UHF scanners or the lack of multi-modality resources for a patient. Hence, it can become a potential research topic for medical image analysis. The application can rapidly provide synthesized results in minutes, and the radiologist can evaluate the performance of the algorithm and its agreement with the imaging examination.

Lastly, the extreme objective of clinical workflow is to fill the gap between research and practice. Although we do not want to replace the radiologists in decision making instead of assisting them

in providing optimal personalized care for their patients. However, we hope our work can benefit current MRI acquisition to produce realistic MRI as close to natural as possible. If so, it can overcome the current infrastructure limitation, time consumption, and expense. For example, the 7T MRI is currently the highest strength-field MRI in the commercial. However, it is not popular in clinical and research centers, leading to a lack of UHF resources in diagnosis. If we can have an optimized method with the superior result for ultra-high field and cross-modality translation, we can rapidly produce a complete multi-modality MRI for each patient with less time and expense.

Bibliography

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 126–135, 2017.

[2] Zeynettin Akkus, Alfiia Galimzianova, Assaf Hoogi, Daniel L Rubin, and Bradley J Erickson. Deep learning for brain mri segmentation: state of the art and future directions. *Journal of digital imaging*, 30(4):449–459, 2017.

[3] Guillaume Alain and Yoshua Bengio. What regularized auto-encoders learn from the data-generating distribution. *The Journal of Machine Learning Research*, 15(1):3563–3593, 2014.

[4] Naomi S Altman. An introduction to kernel and nearest-neighbor nonparametric regression. *The American Statistician*, 46(3):175–185, 1992.

[5] Hossein Arabi, Jason A Dowling, Ninon Burgos, Xiao Han, Peter B Greer, Nikolaos Koutsouvelis, and Habib Zaidi. Comparative study of algorithms for synthetic ct generation from mri: consequences for mri-guided radiation planning in the pelvic region. *Medical physics*, 45(11):5218–5233, 2018.

[6] Hossein Arabi and Habib Zaidi. Magnetic resonance imaging-guided attenuation correction in whole-body pet/mri using a sorted atlas approach. *Medical image analysis*, 31:1–15, 2016.

[7] Ali Abbasian Ardakani, Afshin Mohammadi, Bahareh Khalili Najafabad, and Jamileh Abolghasemi. Assessment of kidney function after allograft transplantation by texture analysis. *Iranian journal of kidney diseases*, 11(2):157, 2017.

[8] Martin Arjovsky, Soumith Chintala, and Léon Bottou. Wasserstein generative adversarial networks. In *International conference on machine learning*, pages 214–223. PMLR, 2017.

[9] Devansh Arpit, Yingbo Zhou, Hung Ngo, and Venu Govindaraju. Why regularized auto-encoders learn sparse representation? In *International Conference on Machine Learning*, pages 136–144. PMLR, 2016.

[10] Kyongtae T Bae, Cheng Tao, Fang Zhu, James E Bost, Arlene B Chapman, Jared J Grantham, Vicente E Torres, Lisa M Guay-Woodford, Catherine M Meyers, William M Bennett, et al. Mri-based kidney volume measurements in adpkd: reliability and effect of gadolinium enhancement. *Clinical Journal of the American Society of Nephrology*, 4(4):719–725, 2009.

[11] Khosro Bahrami, Feng Shi, Islem Rekik, Yaozong Gao, and Dinggang Shen. 7t-guided super-resolution of 3t mri. *Medical physics*, 44(5):1661–1677, 2017.

[12] Spyridon Bakas, Hamed Akbari, Aristeidis Sotiras, Michel Bilello, Martin Rozycki, Justin S Kirby, John B Freymann, Keyvan Farahani, and Christos Davatzikos. Advancing the cancer genome atlas glioma mri collections with expert segmentation labels and radiomic features. *Scientific data*, 4(1):1–13, 2017.

[13] Spyridon Bakas, Mauricio Reyes, Andras Jakab, Stefan Bauer, Markus Rempfler, Alessandro Crimi, Russell Takeshi Shinohara, Christoph Berger, Sung Min Ha, Martin Rozycki, et al. Identifying the best machine learning algorithms for brain tumor segmentation, progression assessment, and overall survival prediction in the brats challenge. *arXiv preprint arXiv:1811.02629*, 2018.

[14] Horace B Barlow. Unsupervised learning. *Neural computation*, 1(3):295–311, 1989.

[15] Björn Behr, Jörg Stadler, Henrik J Michaely, Hans-Georg Damert, and Wolfgang Schneider. Mr imaging of the human hand and wrist at 7 t. *Skeletal radiology*, 38(9):911–917, 2009.

[16] Mikhail Belkin and Partha Niyogi. Laplacian eigenmaps for dimensionality reduction and data representation. *Neural computation*, 15(6):1373–1396, 2003.

[17] Aminu K Bello, Adeera Levin, Braden J Manns, John Feehally, Tilman Drueke, Labib Faruque, Brenda R Hemmelgarn, Charles Kernahan, Johannes Mann, Scott Klarenbach, et al. Effective ckd care in european countries: challenges and opportunities for health policy. *American Journal of Kidney Diseases*, 65(1):15–25, 2015.

[18] Shai Ben-David, John Blitzer, Koby Crammer, Alex Kulesza, Fernando Pereira, and Jennifer Wortman Vaughan. A theory of learning from different domains. *Machine learning*, 79(1):151–175, 2010.

[19] Yoshua Bengio, Aaron Courville, and Pascal Vincent. Representation learning: A review and new perspectives. *IEEE transactions on pattern analysis and machine intelligence*, 35(8):1798–1828, 2013.

[20] Yoshua Bengio and Yves Grandvalet. No unbiased estimator of the variance of k-fold cross-validation. *Advances in Neural Information Processing Systems*, 16, 2003.

[21] Yoshua Bengio, Pascal Lamblin, Dan Popovici, and Hugo Larochelle. Greedy layer-wise training of deep networks. *Advances in neural information processing systems*, 19, 2006.

[22] J. Gordon Betts, Peter Desaix, Eddie Johnson, E. Johnson Jody, Oksana Korol, Dean Kruse, Brandon Poe, Wise A. James, Mark Womble, and Kelly A. Young. *Anatomy & Physiology*. OpenStax, 2013.

[23] Richard Bitar, General Leung, Richard Perng, Sameh Tadros, Alan R Moody, Josee Sarrazin, Caitlin McGregor, Monique Christakis, Sean Symons, Andrew Nelson, et al. Mr pulse sequences: what every radiologist wants to know but is afraid to ask. *Radiographics*, 26(2):513–537, 2006.

[24] Andreas K Bitz, Irina Brote, Stephan Orzada, Oliver Kraff, Stefan Maderwald, Harald H Quick, Pedram Yazdanbakhsh, Klaus Solbach, Achim Bahr, Thomas Bolz, et al. An 8-channel add-on rf shimming system for whole-body 7 tesla mri including real-time sar monitoring. In *Proceedings of the 17th Annual Meeting of ISMRM*. Citeseer, 2009.

[25] Felix Bloch. Nuclear induction. *Physical review*, 70(7-8):460, 1946.

[26] Isabelle Bongiovanni, Anne-Line Couillerot-Peyrondet, Cléa Sambuc, Emmanuelle Dantony, Mad-Hélénie Elsensohn, Yoël Sainsaulieu, René Ecochard, and Cécile Couchoud. Évaluation médico-économique des stratégies de prise en charge de l'insuffisance rénale chronique terminale en france. *Néphrologie & Thérapeutique*, 12(2):104–115, 2016.

[27] Alan S Boyd, John A Zic, and Jerrold L Abraham. Gadolinium deposition in nephrogenic fibrosing dermopathy. *Journal of the American Academy of Dermatology*, 56(1):27–30, 2007.

[28] Leo Breiman. Random forests. *Machine learning*, 45(1):5–32, 2001.

[29] Leo Breiman, Jerome H Friedman, Richard A Olshen, and Charles J Stone. *Classification and regression trees*. Routledge, 2017.

[30] Mark A Brown and Richard C Semelka. *MRI: basic principles and applications*. John Wiley & Sons, 2011.

[31] Yuri Burda, Roger Grosse, and Ruslan Salakhutdinov. Importance weighted autoencoders. *arXiv preprint arXiv:1509.00519*, 2015.

[32] Chensi Cao, Feng Liu, Hai Tan, Deshou Song, Wenjie Shu, Weizhong Li, Yiming Zhou, Xiaochen Bo, and Zhi Xie. Deep learning and its applications in biomedicine. *Genomics, proteomics & bioinformatics*, 16(1):17–32, 2018.

[33] Gabriel Chartrand, Phillip M Cheng, Eugene Vorontsov, Michal Drozdzal, Simon Turcotte, Christopher J Pal, Samuel Kadoury, and An Tang. Deep learning: a primer for radiologists. *Radiographics*, 37(7):2113–2131, 2017.

[34] Agisilaos Chartsias, Thomas Joyce, Mario Valerio Giuffrida, and Sotirios A Tsaftaris. Multimodal mr synthesis via modality-invariant latent representation. *IEEE transactions on medical imaging*, 37(3):803–814, 2017.

[35] Min Chen, Amog Jog, Aaron Carass, and Jerry L Prince. Using image synthesis for multi-channel registration of different image modalities. In *Medical Imaging 2015: Image Processing*, volume 9413, page 94131Q. International Society for Optics and Photonics, 2015.

[36] Ricky TQ Chen, Xuechen Li, Roger B Grosse, and David K Duvenaud. Isolating sources of disentanglement in variational autoencoders. *Advances in neural information processing systems*, 31, 2018.

[37] Shupeng Chen, An Qin, Dingyi Zhou, and Di Yan. U-net-generated synthetic ct images for magnetic resonance imaging-only prostate intensity-modulated radiation therapy treatment planning. *Medical physics*, 45(12):5659–5665, 2018.

[38] Yuhua Chen, Anthony G Christodoulou, Zhengwei Zhou, Feng Shi, Yibin Xie, and Debiao Li. Mri super-resolution with gan and 3d multi-level densenet: smaller, faster, and better. *arXiv preprint arXiv:2003.01217*, 2020.

[39] Yuhua Chen, Feng Shi, Anthony G Christodoulou, Yibin Xie, Zhengwei Zhou, and Debiao Li. Efficient and accurate mri super-resolution using a generative adversarial network and 3d multi-level densely connected network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 91–99. Springer, 2018.

[40] Yuhua Chen, Yibin Xie, Zhengwei Zhou, Feng Shi, Anthony G Christodoulou, and Debiao Li. Brain mri super resolution using 3d deep densely connected neural networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 739–742. IEEE, 2018.

[41] Travers Ching, Daniel S Himmelstein, Brett K Beaulieu-Jones, Alexandr A Kalinin, Brian T Do, Gregory P Way, Enrico Ferrero, Paul-Michael Agapow, Michael Zietz, Michael M Hoffman, et al. Opportunities and obstacles for deep learning in biology and medicine. *Journal of The Royal Society Interface*, 15(141):20170387, 2018.

[42] Yunjey Choi, Minje Choi, Munyoung Kim, Jung-Woo Ha, Sunghun Kim, and Jaegul Choo. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 8789–8797, 2018.

[43] Andreas Christ, Wolfgang Kainz, Eckhart G Hahn, Katharina Honegger, Marcel Zefferer, Esra Neufeld, Wolfgang Rascher, Rolf Janka, Werner Bautz, Ji Chen, et al. The virtual family—development of surface-based anatomical models of two adults and two children for dosimetric simulations. *Physics in Medicine & Biology*, 55(2):N23, 2009.

[44] Xianjin Dai, Yang Lei, Yabo Fu, Walter J Curran, Tian Liu, Hui Mao, and Xiaofeng Yang. Multimodal mri synthesis using unified generative adversarial networks. *Medical Physics*, 47(12):6343–6354, 2020.

[45] Salman UH Dar, Mahmut Yurt, Levent Karacan, Aykut Erdem, Erkut Erdem, and Tolga Çukur. Image synthesis in multi-contrast mri with conditional generative adversarial networks. *IEEE transactions on medical imaging*, 38(10):2375–2388, 2019.

[46] Salman UH Dar, Mahmut Yurt, Mohammad Shahdloo, Muhammed Emrullah Ildız, Berk Tınaz, and Tolga Çukur. Prior-guided image reconstruction for accelerated multi-contrast mri via generative adversarial networks. *IEEE Journal of Selected Topics in Signal Processing*, 14(6):1072–1087, 2020.

[47] Anna J Dare, Gavin J Pettigrew, and Kourosh Saeb-Parsy. Preoperative assessment of the deceased-donor kidney: from macroscopic appearance to molecular biomarkers. *Transplantation*, 97(8):797–807, 2014.

[48] Laurens JL De Cocker, Arjen Lindenholz, Jaco JM Zwanenburg, Anja G van der Kolk, Maarten Zwartbol, Peter R Luijten, and Jeroen Hendrikse. Clinical vascular imaging in the brain at 7 t. *Neuroimage*, 168:452–458, 2018.

[49] Beatrice De Coene, Joseph V Hajnal, Peter Gatehouse, Donald B Longmore, Susan J White, Angela Oatridge, JM Pennock, IR Young, and GM Bydder. Mr of the brain using fluid-attenuated inversion recovery (flair) pulse sequences. *American journal of neuroradiology*, 13(6):1555–1564, 1992.

[50] Emily L Denton, Soumith Chintala, Rob Fergus, et al. Deep generative image models using a laplacian pyramid of adversarial networks. *Advances in neural information processing systems*, 28, 2015.

[51] Anna M Dinkla, Jelmer M Wolterink, Matteo Maspero, Mark HF Savenije, Joost JC Verhoeff, Enrica Seravalli, Ivana Išgum, Peter R Seevinck, and Cornelis AT van den Berg. Mr-only brain radiation therapy: dosimetric evaluation of synthetic cts generated by a dilated convolutional neural network. *International Journal of Radiation Oncology* Biology* Physics*, 102(4):801–812, 2018.

[52] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European conference on computer vision*, pages 184–199. Springer, 2014.

[53] Xue Dong, Tonghe Wang, Yang Lei, Kristin Higgins, Tian Liu, Walter J Curran, Hui Mao, Jonathon A Nye, and Xiaofeng Yang. Synthetic ct generation from non-attenuation corrected pet images for whole-body pet imaging. *Physics in Medicine & Biology*, 64(21):215016, 2019.

[54] Ankur M Doshi, Justin M Ream, Andrea S Kierans, Matthew Bilbily, Henry Rusinek, William C Huang, and Hersh Chandarana. Use of mri in differentiation of papillary renal cell carcinoma subtypes: qualitative and quantitative analysis. *American Journal of Roentgenology*, 206(3):566–572, 2016.

[55] John Duchi, Elad Hazan, and Yoram Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of machine learning research*, 12(7), 2011.

[56] Hajar Emami, Ming Dong, Siamak P Nejad-Davarani, and Carri K Glide-Hurst. Generating synthetic cts from magnetic resonance images using generative adversarial networks. *Medical physics*, 45(8):3627–3636, 2018.

[57] Ghaneh Fananapazir, Ramit Lamba, Brittany Lewis, Michael T Corwin, Sima Naderi, and Christoph Troppmann. Utility of mri in the characterization of indeterminate small renal lesions previously seen on screening ct scans of potential renal donor patients. *American Journal of Roentgenology*, 205(2):325–330, 2015.

[58] Thomas M Fiedler, Mark E Ladd, and Andreas K Bitz. Sar simulations & safety. *Neuroimage*, 168:33–58, 2018.

[59] Stig E Forshult. Magnetic resonance imaging–mri–an overview. 2007.

[60] David Foster. *Generative deep learning: teaching machines to paint, write, compose, and play.* O'Reilly Media, 2019.

[61] William T Freeman, Egon C Pasztor, and Owen T Carmichael. Learning low-level vision. *International journal of computer vision*, 40(1):25–47, 2000.

[62] Maayan Frid-Adar, Eyal Klang, Michal Amitai, Jacob Goldberger, and Hayit Greenspan. Synthetic data augmentation using gan for improved liver lesion classification. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 289–293. IEEE, 2018.

[63] Elke R Gizewski, Stefan Maderwald, Jennifer Linn, Benjamin Dassinger, Katja Bochmann, Michael Forsting, and Mark E Ladd. High-resolution anatomy of the human brain stem using 7-t mri: improved detection of inner structures and nerves? *Neuroradiology*, 56(3):177–186, 2014.

[64] ER Gizewski, C Mönninghoff, and M Forsting. Perspectives of ultra-high-field mri in neuroradiology. *Clinical neuroradiology*, 25(2):267–273, 2015.

[65] Xavier Glorot, Antoine Bordes, and Yoshua Bengio. Deep sparse rectifier neural networks. In *Proceedings of the fourteenth international conference on artificial intelligence and statistics*, pages 315–323. JMLR Workshop and Conference Proceedings, 2011.

[66] Enhao Gong, John M Pauly, Max Wintermark, and Greg Zaharchuk. Deep learning enables reduced gadolinium dose for contrast-enhanced brain mri. *Journal of magnetic resonance imaging*, 48(2):330–340, 2018.

[67] Ian Goodfellow. Nips 2016 tutorial: Generative adversarial networks. *arXiv preprint arXiv:1701.00160*, 2016.

[68] Ian Goodfellow, Yoshua Bengio, Aaron Courville, et al. Deep learning book. *MIT Press*, 521(7553):800, 2016.

[69] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Advances in neural information processing systems*, 27, 2014.

[70] John T Guibas, Tejpal S Virdi, and Peter S Li. Synthetic medical images from dual generative adversarial networks. *arXiv preprint arXiv:1709.01872*, 2017.

[71] Dinank Gupta, Michelle Kim, Karen A Vineberg, and James M Balter. Generation of synthetic ct images from mri for treatment planning and patient positioning using a 3-channel u-net trained on sagittal images. *Frontiers in oncology*, 9:964, 2019.

[72] Xiao Han. Mr-based synthetic ct generation using a deep convolutional neural network method. *Medical physics*, 44(4):1408–1419, 2017.

[73] Muhammad Haris, Gregory Shakhnarovich, and Norimichi Ukita. Deep back-projection networks for super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1664–1673, 2018.

[74] Joseph Harms, Yang Lei, Tonghe Wang, Rongxiao Zhang, Jun Zhou, Xiangyang Tang, Walter J Curran, Tian Liu, and Xiaofeng Yang. Paired cycle-gan-based image correction for quantitative cone-beam computed tomography. *Medical physics*, 46(9):3998–4009, 2019.

[75] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[76] James Hensman, Nicolo Fusi, and Neil D Lawrence. Gaussian processes for big data. *arXiv preprint arXiv:1309.6835*, 2013.

[77] Johannes T Heverhagen, Eric Bourekas, Steffen Sammet, Michael V Knopp, and Petra Schmalbrock. Time-of-flight magnetic resonance angiography at 7 tesla. *Investigative radiology*, 43(8):568–573, 2008.

[78] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. 2016.

[79] Geoffrey Hinton, Nitish Srivastava, and Kevin Swersky. Neural networks for machine learning lecture 6a overview of mini-batch gradient descent. *Cited on*, 14(8):2, 2012.

[80] Geoffrey E Hinton and Ruslan R Salakhutdinov. Reducing the dimensionality of data with neural networks. *science*, 313(5786):504–507, 2006.

[81] Geoffrey E Hinton, Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan R Salakhutdinov. Improving neural networks by preventing co-adaptation of feature detectors. *arXiv preprint arXiv:1207.0580*, 2012.

[82] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[83] Yawen Huang, Ling Shao, and Alejandro F Frangi. Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6070–6079, 2017.

[84] Sergey Ioffe and Christian Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In *International conference on machine learning*, pages 448–456. PMLR, 2015.

[85] Phillip Isola, Jun-Yan Zhu, Tinghui Zhou, and Alexei A Efros. Image-to-image translation with conditional adversarial networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1125–1134, 2017.

[86] Mohammadhassan Izadyyazdanabadi, Evgenii Belykh, Michael A Mooney, Jennifer M Eschbacher, Peter Nakaji, Yezhou Yang, and Mark C Preul. Prospects for theranostics in neurosurgical imaging: empowering confocal laser endomicroscopy diagnostics via deep learning. *Frontiers in oncology*, 8:240, 2018.

[87] Abdul Jabbar, Xi Li, and Bourahla Omar. A survey on generative adversarial networks: Variants, applications, and training. *ACM Computing Surveys (CSUR)*, 54(8):1–49, 2021.

[88] Hyungseok Jang, Fang Liu, Gengyan Zhao, Tyler Bradshaw, and Alan B McMillan. Deep learning based mrac using rapid ultrashort echo time imaging. *Medical physics*, 45(8):3697–3704, 2018.

[89] Mark Jenkinson and Stephen Smith. A global optimisation method for robust affine registration of brain images. *Medical image analysis*, 5(2):143–156, 2001.

[90] Cheng-Bin Jin, Hakil Kim, Mingjie Liu, Wonmo Jung, Seongsu Joo, Eunsik Park, Young Saem Ahn, In Ho Han, Jae Il Lee, and Xuenan Cui. Deep ct to mr synthesis using paired and unpaired data. *Sensors*, 19(10):2361, 2019.

[91] Amod Jog, Aaron Carass, Snehashis Roy, Dzung L Pham, and Jerry L Prince. Mr image synthesis by contrast learning on neighborhood ensembles. *Medical image analysis*, 24(1):63–76, 2015.

[92] Amod Jog, Aaron Carass, Snehashis Roy, Dzung L Pham, and Jerry L Prince. Random forest regression for magnetic resonance image synthesis. *Medical image analysis*, 35:475–488, 2017.

[93] Amod Jog, Snehashis Roy, Aaron Carass, and Jerry L Prince. Magnetic resonance image synthesis through patch regression. In *2013 IEEE 10th International Symposium on Biomedical Imaging*, pages 350–353. IEEE, 2013.

[94] Alexia Jolicoeur-Martineau. The relativistic discriminator: a key element missing from standard gan. *arXiv preprint arXiv:1807.00734*, 2018.

[95] Shantanu H Joshi, Antonio Marquina, Stanley J Osher, Ivo Dinov, John D Van Horn, and Arthur W Toga. Mri resolution enhancement using total variation regularization. In *2009 IEEE International Symposium on Biomedical Imaging: From Nano to Macro*, pages 161–164. IEEE, 2009.

[96] Thomas Joyce, Agisilaos Chartsias, and Sotirios A Tsaftaris. Robust multi-modal mr image synthesis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 347–355. Springer, 2017.

[97] Aurang Z Khawaja, Deirdre B Cassidy, Julien Al Shakarchi, Damian G McGrogan, Nicholas G Inston, and Robert G Jones. Revisiting the risks of mri with gadolinium based contrast agents—review of literature and guidelines. *Insights into imaging*, 6(5):553–558, 2015.

[98] VS Khoo and DL Joon. New developments in mri for target volume delineation in radiotherapy. *The British journal of radiology*, 79(special_issue_1):S2–S15, 2006.

[99] Hyunjik Kim and Andriy Mnih. Disentangling by factorising. In *International Conference on Machine Learning*, pages 2649–2658. PMLR, 2018.

[100] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1646–1654, 2016.

[101] Ki Hwan Kim, Won-Joon Do, and Sung-Hong Park. Improving resolution of mr images with an adversarial network incorporating images with different contrast. *Medical physics*, 45(7):3120–3131, 2018.

[102] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

[103] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. *arXiv preprint arXiv:1312.6114*, 2013.

[104] Diederik P Kingma and Max Welling. An introduction to variational autoencoders. *arXiv preprint arXiv:1906.02691*, 2019.

[105] Durk P Kingma, Tim Salimans, Rafal Jozefowicz, Xi Chen, Ilya Sutskever, and Max Welling. Improved variational inference with inverse autoregressive flow. *Advances in neural information processing systems*, 29, 2016.

[106] Andy Kitchen and Jarrel Seah. Deep generative adversarial neural networks for realistic prostate lesion mri synthesis. *arXiv preprint arXiv:1708.00129*, 2017.

[107] Scott W Klarenbach, Marcello Tonelli, Betty Chui, and Braden J Manns. Economic evaluation of dialysis therapies. *Nature Reviews Nephrology*, 10(11):644–652, 2014.

[108] K Kollia, Stefan Maderwald, N Putzki, M Schlamann, JM Theysohn, Oliver Kraff, Mark E Ladd, M Forsting, and I Wanke. First clinical study on ultra-high-field mr imaging in patients with multiple sclerosis: comparison of 1.5 t and 7t. *American Journal of Neuroradiology*, 30(4):699–702, 2009.

[109] Alex Krizhevsky, Ilya Sutskever, and Geoffrey E Hinton. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems*, 25, 2012.

[110] Anja Laader, Karsten Beiderwellen, Oliver Kraff, Stefan Maderwald, Karsten Wrede, Mark E Ladd, Thomas C Lauenstein, Michael Forsting, Harald H Quick, Kai Nassenstein, et al. 1.5 versus 3 versus 7 tesla in abdominal mri: A comparative study. *PLoS One*, 12(11):e0187528, 2017.

[111] Mark E Ladd, Peter Bachert, Martin Meyerspeer, Ewald Moser, Armin M Nagel, David G Norris, Sebastian Schmitter, Oliver Speck, Sina Straub, and Moritz Zaiss. Pros and cons of ultra-high-field mri/mrs for human application. *Progress in nuclear magnetic resonance spectroscopy*, 109:1–50, 2018.

[112] John Lafferty, Andrew McCallum, and Fernando CN Pereira. Conditional random fields: Probabilistic models for segmenting and labeling sequence data. 2001.

[113] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 624–632, 2017.

[114] Paul C Lauterbur. Image formation by induced local interactions: examples employing nuclear magnetic resonance. *nature*, 242(5394):190–191, 1973.

[115] Yann LeCun, Yoshua Bengio, et al. Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, 3361(10):1995, 1995.

[116] Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436–444, 2015.

[117] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017.

[118] John A Lee and Michel Verleysen. *Nonlinear dimensionality reduction*, volume 1. Springer, 2007.

[119] June-Goo Lee, Sanghoon Jun, Young-Won Cho, Hyunna Lee, Guk Bae Kim, Joon Beom Seo, and Namkug Kim. Deep learning in medical imaging: general overview. *Korean journal of radiology*, 18(4):570–584, 2017.

[120] Young Han Lee. Efficiency improvement in a busy radiology practice: determination of musculoskeletal magnetic resonance imaging protocol using deep-learning convolutional neural networks. *Journal of digital imaging*, 31(5):604–610, 2018.

[121] Thomas Martin Lehmann, Claudia Gonner, and Klaus Spitzer. Survey: Interpolation methods in medical image processing. *IEEE transactions on medical imaging*, 18(11):1049–1075, 1999.

[122] Yang Lei, Yabo Fu, Hui Mao, Walter J Curran, Tian Liu, and Xiaofeng Yang. Multi-modality mri arbitrary transformation using unified generative adversarial networks. In *Medical Imaging 2020: Image Processing*, volume 11313, page 1131303. International Society for Optics and Photonics, 2020.

[123] Yang Lei, Joseph Harms, Tonghe Wang, Yingzi Liu, Hui-Kuo Shu, Ashesh B Jani, Walter J Curran, Hui Mao, Tian Liu, and Xiaofeng Yang. Mri-only based synthetic ct generation using dense cycle consistent generative adversarial networks. *Medical physics*, 46(8):3565–3581, 2019.

[124] Hongwei Li, Johannes C Paetzold, Anjany Sekuboyina, Florian Kofler, Jianguo Zhang, Jan S Kirschke, Benedikt Wiestler, and Bjoern Menze. Diamondgan: unified multi-modal generative adversarial networks for mri sequences synthesis. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 795–803. Springer, 2019.

[125] Y Li, Bruno Sixou, and F Peyrin. A review of the deep learning methods for medical images super resolution problems. *IRBM*, 42(2):120–133, 2021.

[126] Yinghua Li, Bin Song, Jie Guo, Xiaojiang Du, and Mohsen Guizani. Super-resolution of brain mri images using overcomplete dictionaries and nonlocal similarity. *IEEE Access*, 7:25897–25907, 2019.

[127] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 136–144, 2017.

[128] Chang Liu, Xi Wu, Xi Yu, YuanYan Tang, Jian Zhang, and JiLiu Zhou. Fusing multi-scale information in convolution network for mr image super-resolution reconstruction. *Biomedical engineering online*, 17(1):1–23, 2018.

[129] Fang Liu. Susan: segment unannotated image structure using adversarial network. *Magnetic resonance in medicine*, 81(5):3330–3345, 2019.

[130] Fang Liu, Hyungseok Jang, Richard Kijowski, Tyler Bradshaw, and Alan B McMillan. Deep learning mr imaging–based attenuation correction for pet/mr imaging. *Radiology*, 286(2):676–684, 2018.

[131] Fang Liu, Poonam Yadav, Andrew M Baschnagel, and Alan B McMillan. Mr-based treatment planning in radiation therapy using a deep learning approach. *Journal of applied clinical medical physics*, 20(3):105–114, 2019.

[132] Hanzhang Lu, Lidia M Nagae-Poetscher, Xavier Golay, Doris Lin, Martin Pomper, and Peter CM Van Zijl. Routine clinical brain mri sequences for use at 3.0 tesla. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 22(1):13–22, 2005.

[133] Xiaoqiang Lu, Zihan Huang, and Yuan Yuan. Mr image super-resolution via manifold regularized sparse learning. *Neurocomputing*, 162:96–104, 2015.

[134] Dan Ma, Vikas Gulani, Nicole Seiberlich, Kecheng Liu, Jeffrey L Sunshine, Jeffrey L Duerk, and Mark A Griswold. Magnetic resonance fingerprinting. *Nature*, 495(7440):187–192, 2013.

[135] Alireza Makhzani and Brendan Frey. K-sparse autoencoders. *arXiv preprint arXiv:1312.5663*, 2013.

[136] José V Manjón, Pierrick Coupé, Antonio Buades, D Louis Collins, and Montserrat Robles. Mri superresolution using self-similarity and image priors. *International journal of biomedical imaging*, 2010, 2010.

[137] José V Manjón, Pierrick Coupé, Antonio Buades, Vladimir Fonov, D Louis Collins, and Montserrat Robles. Non-local mri upsampling. *Medical image analysis*, 14(6):784–792, 2010.

[138] AJ Matas, JM Smith, MA Skeans, B Thompson, SK Gustafson, DE Stewart, WS Cherikh, JL Wainright, G Boyle, JJ Snyder, et al. Optn/srtr 2013 annual data report: kidney. *American Journal of Transplantation*, 15(S2):1–34, 2015.

[139] Maciej A Mazurowski, Mateusz Buda, Ashirbani Saha, and Mustafa R Bashir. Deep learning in radiology: An overview of the concepts and a survey of the state of the art with focus on mri. *Journal of magnetic resonance imaging*, 49(4):939–954, 2019.

[140] Bjoern H Menze, Andras Jakab, Stefan Bauer, Jayashree Kalpathy-Cramer, Keyvan Farahani, Justin Kirby, Yuliya Burren, Nicole Porz, Johannes Slotboom, Roland Wiest, et al. The multimodal brain tumor image segmentation benchmark (brats). *IEEE transactions on medical imaging*, 34(10):1993–2024, 2014.

[141] Philippe Meyer, Vincent Noblet, Christophe Mazzara, and Alex Lallement. Survey on deep learning for radiotherapy. *Computers in biology and medicine*, 98:126–146, 2018.

[142] Mehdi Mirza and Simon Osindero. Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*, 2014.

[143] Swapnil Mishra, Seth Flaxman, Tresnia Berah, Mikko Pakkanen, Harrison Zhu, and Samir Bhatt. pi-vae: Encoding stochastic process priors with variational autoencoders. *arXiv preprint arXiv:2002.06873*, 2020.

[144] Andrew D Missert, Lifeng Yu, Shuai Leng, Joel G Fletcher, and Cynthia H McCollough. Synthesizing images from multiple kernels using a deep convolutional neural network. *Medical physics*, 47(2):422–430, 2020.

[145] Tony CW Mok and Albert Chung. Learning data augmentation for brain tumor segmentation with coarse-to-fine generative adversarial networks. In *International MICCAI Brainlesion Workshop*, pages 70–80. Springer, 2018.

[146] Sebastian Neppl, Guillaume Landry, Christopher Kurz, David C Hansen, Ben Hoyle, Sophia Stöcklein, Max Seidensticker, Jochen Weller, Claus Belka, Katia Parodi, et al. Evaluation of proton and photon dose distributions recalculated on 2d and 3d unet-generated pseudocts from t1-weighted mr head scans. *Acta Oncologica*, 58(10):1429–1434, 2019.

[147] Andrew Ng et al. Sparse autoencoder. *CS294A Lecture notes*, 72(2011):1–19, 2011.

[148] Dong Nie, Roger Trullo, Jun Lian, Li Wang, Caroline Petitjean, Su Ruan, Qian Wang, and Dinggang Shen. Medical image synthesis with deep convolutional adversarial networks. *IEEE Transactions on Biomedical Engineering*, 65(12):2720–2730, 2018.

[149] Vera Novak and Gregory Christoforidis. Clinical promise: clinical imaging at ultra high field. In *Ultra High Field Magnetic Resonance Imaging*, pages 411–437. Springer, 2006.

[150] Ozan Oktay, Wenjia Bai, Matthew Lee, Ricardo Guerrero, Konstantinos Kamnitsas, Jose Caballero, Antonio de Marvao, Stuart Cook, Declan O'Regan, and Daniel Rueckert. Multi-input cardiac image super-resolution using convolutional neural networks. In *International conference on medical image computing and computer-assisted intervention*, pages 246–254. Springer, 2016.

[151] Ozan Oktay, Enzo Ferrante, Konstantinos Kamnitsas, Mattias Heinrich, Wenjia Bai, Jose Caballero, Stuart A Cook, Antonio De Marvao, Timothy Dawes, Declan P O'Regan, et al. Anatomically constrained neural networks (acnns): application to cardiac image enhancement and segmentation. *IEEE transactions on medical imaging*, 37(2):384–395, 2017.

[152] Sven Olberg, Hao Zhang, William R Kennedy, Jaehee Chun, Vivian Rodriguez, Imran Zoberi, Maria A Thomas, Jin Sung Kim, Sasa Mutic, Olga L Green, et al. Synthetic ct reconstruction using a deep spatial pyramid convolutional framework for mr-only breast radiotherapy. *Medical physics*, 46(9):4135–4147, 2019.

[153] Sahin Olut, Yusuf H Sahin, Ugur Demir, and Gozde Unal. Generative adversarial training for mra image synthesis using multi-contrast mri. In *International workshop on predictive intelligence in medicine*, pages 147–154. Springer, 2018.

[154] Jorge Ortiz, Austin Gregg, Xuerong Wen, Farah Karipineni, and Liise K Kayler. Impact of donor obesity and donation after cardiac death on outcomes after kidney transplantation. *Clinical transplantation*, 26(3):E284–E292, 2012.

[155] S Orzada, HH Quick, ME Ladd, A Bahr, T Bolz, P Yazdanbakhsh, K Solbach, and AK Bitz. A flexible 8-channel transmit/receive body coil for 7 t human imaging. In *Proc Intl Soc Mag Reson Med*, volume 17, page 2999, 2009.

[156] Jiahong Ouyang, Kevin T Chen, Enhao Gong, John Pauly, and Greg Zaharchuk. Ultra-low-dose pet reconstruction using generative adversarial network with feature matching and task-specific perceptual loss. *Medical physics*, 46(8):3555–3564, 2019.

[157] Chi-Hieu Pham, Aurélien Ducournau, Ronan Fablet, and François Rousseau. Brain mri super-resolution using deep 3d convolutional networks. In *2017 IEEE 14th International Symposium on Biomedical Imaging (ISBI 2017)*, pages 197–200. IEEE, 2017.

[158] Esben Plenge, Dirk HJ Poot, Monique Bernsen, Gyula Kotek, Gavin Houston, Piotr Wielopolski, Louise van der Weerd, Wiro J Niessen, and Erik Meijering. Super-resolution methods in mri: can they improve the trade-off between resolution, signal-to-noise ratio, and acquisition time? *Magnetic resonance in medicine*, 68(6):1983–1993, 2012.

[159] Edward M Purcell, Henry Cutler Torrey, and Robert V Pound. Resonance absorption by nuclear magnetic moments in a solid. *Physical review*, 69(1-2):37, 1946.

[160] Ning Qian. On the momentum term in gradient descent learning algorithms. *Neural networks*, 12(1):145–151, 1999.

[161] Liangqiong Qu, Shuai Wang, Pew-Thian Yap, and Dinggang Shen. Wavelet-based semi-supervised adversarial learning for synthesizing realistic 7t from 3t mri. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 786–794. Springer, 2019.

[162] Liangqiong Qu, Yongqin Zhang, Shuai Wang, Pew-Thian Yap, and Dinggang Shen. Synthesized 7t mri from 3t mri via deep learning in spatial and wavelet domains. *Medical image analysis*, 62:101663, 2020.

[163] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv:1511.06434*, 2015.

[164] Siva P Raman, Yifei Chen, James L Schroeder, Peng Huang, and Elliot K Fishman. Ct texture analysis of renal masses: pilot study using random forest classification for prediction of pathology. *Academic radiology*, 21(12):1587–1596, 2014.

[165] Ievgen Redko, Emilie Morvant, Amaury Habrard, Marc Sebban, and Younes Bennani. *Advances in domain adaptation theory*. Elsevier, 2019.

[166] Sebastian Regnery, Benjamin R Knowles, Daniel Paech, Nicolas Behl, Jan-Eric Meissner, Paul Windisch, Semi Ben Harrabi, Denise Bernhardt, Heinz-Peter Schlemmer, Mark E Ladd, et al. High-resolution flair mri at 7 tesla for treatment planning in glioblastoma patients. *Radiotherapy and Oncology*, 130:180–184, 2019.

[167] DJ Reich, DC Mulligan, Peter L Abt, Timothy L Pruett, MMI Abecassis, A D'alessandro, Elizabeth A Pomfret, RB Freeman, JF Markmann, Douglas W Hanto, et al. Asts recommended practice guidelines for controlled donation after cardiac death organ procurement and transplantation. *American journal of transplantation*, 9(9):2004–2011, 2009.

[168] Danilo Rezende and Shakir Mohamed. Variational inference with normalizing flows. In *International conference on machine learning*, pages 1530–1538. PMLR, 2015.

[169] Salah Rifai, Pascal Vincent, Xavier Muller, Xavier Glorot, and Yoshua Bengio. Contractive auto-encoders: Explicit invariance during feature extraction. In *Icml*, 2011.

[170] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241. Springer, 2015.

[171] Sam T Roweis and Lawrence K Saul. Nonlinear dimensionality reduction by locally linear embedding. *science*, 290(5500):2323–2326, 2000.

[172] Snehashis Roy, Aaron Carass, and Jerry L Prince. Magnetic resonance image example-based contrast synthesis. *IEEE transactions on medical imaging*, 32(12):2348–2363, 2013.

[173] Andrea Rueda, Norberto Malpica, and Eduardo Romero. Single-image super-resolution of brain mr images using overcomplete dictionaries. *Medical image analysis*, 17(1):113–132, 2013.

[174] David E Rumelhart, Geoffrey E Hinton, and Ronald J Williams. Learning internal representations by error propagation. Technical report, California Univ San Diego La Jolla Inst for Cognitive Science, 1985.

[175] Val M Runge. Safety of approved mr contrast media for intravenous injection. *Journal of Magnetic Resonance Imaging: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 12(2):205–213, 2000.

[176] Tanja C Saat, Eline K van den Akker, Jan NM IJzermans, Frank JMF Dor, and Ron WF de Bruin. Improving the outcome of kidney transplantation by ameliorating renal ischemia reperfusion injury: lost in translation? *Journal of translational medicine*, 14(1):1–9, 2016.

[177] Mehdi SM Sajjadi, Bernhard Scholkopf, and Michael Hirsch. Enhancenet: Single image super-resolution through automated texture synthesis. In *Proceedings of the IEEE international conference on computer vision*, pages 4491–4500, 2017.

[178] Tim Salimans, Ian Goodfellow, Wojciech Zaremba, Vicki Cheung, Alec Radford, and Xi Chen. Improved techniques for training gans. *Advances in neural information processing systems*, 29, 2016.

[179] Irina Sánchez and Verónica Vilaplana. Brain mri super-resolution using 3d generative adversarial networks. *arXiv preprint arXiv:1812.11440*, 2018.

[180] Lawrence K Saul, Kilian Q Weinberger, Fei Sha, Jihun Ham, and Daniel D Lee. Spectral methods for dimensionality reduction. *Semi-supervised learning*, 3, 2006.

[181] Nicola Schieda, Jason I Blaichman, Andreu F Costa, Rafael Glikstein, Casey Hurrell, Matthew James, Pejman Jabehdar Maralani, Wael Shabana, An Tang, Anne Tsampalieros, et al. Gadolinium-based contrast agents in kidney disease: a comprehensive review and clinical practice guideline issued by the canadian association of radiologists. *Canadian journal of kidney health and disease*, 5:2054358118778573, 2018.

[182] Bernhard Schölkopf, Alexander Smola, and Klaus-Robert Müller. Nonlinear component analysis as a kernel eigenvalue problem. *Neural computation*, 10(5):1299–1319, 1998.

[183] Vasileios Sevetlidis, Mario Valerio Giuffrida, and Sotirios A Tsaftaris. Whole image synthesis using a deep encoder-decoder network. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 127–137. Springer, 2016.

[184] Dinggang Shen, Guorong Wu, and Heung-Il Suk. Deep learning in medical image analysis. *Annual review of biomedical engineering*, 19:221–248, 2017.

[185] Feng Shi, Jian Cheng, Li Wang, Pew-Thian Yap, and Dinggang Shen. Lrtv: Mr image super-resolution with low-rank and total variation regularizations. *IEEE transactions on medical imaging*, 34(12):2459–2466, 2015.

[186] Jun Shi, Zheng Li, Shihui Ying, Chaofeng Wang, Qingping Liu, Qi Zhang, and Pingkun Yan. Mr image super-resolution via wide residual networks with fixed skip connection. *IEEE journal of biomedical and health informatics*, 23(3):1129–1140, 2018.

[187] Jun Shi, Qingping Liu, Chaofeng Wang, Qi Zhang, Shihui Ying, and Haoyu Xu. Super-resolution reconstruction of mr image with a novel residual learning network algorithm. *Physics in Medicine & Biology*, 63(8):085011, 2018.

[188] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1874–1883, 2016.

[189] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[190] Wan-Chi Siu and Kwok-Wai Hung. Review of image interpolation and super-resolution. In *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 1–10. IEEE, 2012.

[191] Muhammad Sohail, Muhammad Naveed Riaz, Jing Wu, Chengnian Long, and Shaoyuan Li. Unpaired multi-contrast mr image synthesis using generative adversarial networks. In *International Workshop on Simulation and Synthesis in Medical Imaging*, pages 22–31. Springer, 2019.

[192] Casper Kaae Sønderby, Tapani Raiko, Lars Maaløe, Søren Kaae Sønderby, and Ole Winther. Ladder variational autoencoders. *Advances in neural information processing systems*, 29, 2016.

[193] Elisabeth Springer, Barbara Dymerska, Pedro Lima Cardoso, Simon Daniel Robinson, Christian Weisstanner, Roland Wiest, Benjamin Schmitt, and Siegfried Trattnig. Comparison of routine brain imaging at 3 t and 7 t. *Investigative radiology*, 51(8):469, 2016.

[194] Karl D Spuhler, John Gardus, Yi Gao, Christine DeLorenzo, Ramin Parsey, and Chuan Huang. Synthesis of patient-specific transmission data for pet attenuation correction for pet/mri neuroimaging using a convolutional neural network. *Journal of nuclear medicine*, 60(4):555–560, 2019.

[195] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014.

[196] Lesley A Stevens, Josef Coresh, Tom Greene, and Andrew S Levey. Assessing kidney function—measured and estimated glomerular filtration rate. *New England Journal of Medicine*, 354(23):2473–2483, 2006.

[197] Dominic M Summers, Christopher JE Watson, Gavin J Pettigrew, Rachel J Johnson, David Collett, James M Neuberger, and J Andrew Bradley. Kidney donation after circulatory death (dcd): state of the art. *Kidney international*, 88(2):241–249, 2015.

[198] Jian Sun, Zongben Xu, and Heung-Yeung Shum. Gradient profile prior and its applications in image super-resolution and enhancement. *IEEE Transactions on Image Processing*, 20(6):1529–1542, 2010.

[199] Kenji Suzuki. Overview of deep learning in medical imaging. *Radiological physics and technology*, 10(3):257–273, 2017.

[200] Ying Tai, Jian Yang, and Xiaoming Liu. Image super-resolution via deep recursive residual network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3147–3155, 2017.

[201] Joshua B Tenenbaum, Vin de Silva, and John C Langford. A global geometric framework for nonlinear dimensionality reduction. *science*, 290(5500):2319–2323, 2000.

[202] Philippe Thévenaz, Thierry Blu, and Michael Unser. Interpolation revisited [medical images application]. *IEEE Transactions on medical imaging*, 19(7):739–758, 2000.

[203] James H Thrall, Xiang Li, Quanzheng Li, Cinthia Cruz, Synho Do, Keith Dreyer, and James Brink. Artificial intelligence and machine learning in radiology: opportunities, challenges, pitfalls, and criteria for success. *Journal of the American College of Radiology*, 15(3):504–508, 2018.

[204] S Tillet, Sébastien Giraud, PO Delpech, Raphael Thuillier, V Ameteau, JM Goujon, B Renelier, L Macchi, T Hauet, and G Mauco. Kidney graft outcome using an anti-xa therapeutic strategy in an experimental model of severe ischaemia–reperfusion injury. *Journal of British Surgery*, 102(1):132–142, 2015.

[205] M Tonelli, N Wiebe, G Knoll, A Bello, S Browne, D Jadhav, S Klarenbach, and J Gill. Systematic review: kidney transplantation compared with dialysis in clinically relevant outcomes. *American journal of transplantation*, 11(10):2093–2109, 2011.

[206] Tong Tong, Gen Li, Xiejie Liu, and Qinquan Gao. Image super-resolution using dense skip connections. In *Proceedings of the IEEE international conference on computer vision*, pages 4799–4807, 2017.

[207] Angel Torrado-Carvajal, Javier Vera-Olmos, David Izquierdo-Garcia, Onofrio A Catalano, Manuel A Morales, Justin Margolin, Andrea Soricelli, Marco Salvatore, Norberto Malpica, and Ciprian Catana. Dixon-vibe deep learning (divide) pseudo-ct synthesis for pelvis pet/mr attenuation correction. *Journal of nuclear medicine*, 60(3):429–435, 2019.

[208] Sébastien Tourbier, Xavier Bresson, Patric Hagmann, Jean-Philippe Thiran, Reto Meuli, and Meritxell Bach Cuadra. An efficient total variation algorithm for super-resolution in fetal brain mri with adaptive regularization. *NeuroImage*, 118:584–597, 2015.

[209] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Instance normalization: The missing ingredient for fast stylization. *arXiv preprint arXiv:1607.08022*, 2016.

[210] Lale Umutlu, Stephan Orzada, Sonja Kinner, Stefan Maderwald, Irina Brote, Andreas K Bitz, Oliver Kraff, Susanne C Ladd, Gerald Antoch, Mark E Ladd, et al. Renal imaging at 7 tesla: preliminary results. *European radiology*, 21(4):841–849, 2011.

[211] Jan Willem Van Dalen, Eva EM Scuric, Susanne J Van Veluw, Matthan WA Caan, Aart J Nederveen, Geert Jan Biessels, Willem A Van Gool, and Edo Richard. Cortical microinfarcts detected in vivo on 3 tesla mri: clinical and radiological correlates. *Stroke*, 46(1):255–257, 2015.

[212] Hien Van Nguyen, Kevin Zhou, and Raviteja Vemulapalli. Cross-domain synthesis of medical images using efficient location-sensitive deep network. In *International Conference on Medical Image Computing and Computer-Assisted Intervention*, pages 677–684. Springer, 2015.

[213] Eric Van Reeth, Ivan WK Tham, Cher Heng Tan, and Chueh Loo Poh. Super-resolution in magnetic resonance imaging: a review. *Concepts in Magnetic Resonance Part A*, 40(6):306–325, 2012.

[214] Vladimir Vapnik. Principles of risk minimization for learning theory. *Advances in neural information processing systems*, 4, 1991.

[215] J Thomas Vaughan, Carl J Snyder, Lance J DelaBarre, Patrick J Bolan, Jinfeng Tian, Lizann Bolinger, Gregor Adriany, Peter Andersen, John Strupp, and Kamil Ugurbil. Whole-body imaging at 7t: preliminary results. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 61(1):244–248, 2009.

[216] John Thomas Vaughan, Michael Garwood, CM Collins, W Liu, Lance DelaBarre, Gregor Adriany, P Andersen, H Merkle, R Goebel, MB Smith, et al. 7t vs. 4t: Rf power, homogeneity, and signal-to-noise comparison in head images. *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, 46(1):24–30, 2001.

[217] Pascal Vincent, Hugo Larochelle, Isabelle Lajoie, Yoshua Bengio, Pierre-Antoine Manzagol, and Léon Bottou. Stacked denoising autoencoders: Learning useful representations in a deep network with a local denoising criterion. *Journal of machine learning research*, 11(12), 2010.

[218] Jiancong Wang, Yuhua Chen, Yifan Wu, Jianbo Shi, and James Gee. Enhanced generative adversarial network for 3d brain mri super-resolution. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 3627–3636, 2020.

[219] Tonghe Wang, Yang Lei, Yabo Fu, Jacob F Wynne, Walter J Curran, Tian Liu, and Xiaofeng Yang. A review on medical imaging synthesis using deep learning and its clinical applications. *Journal of Applied Clinical Medical Physics*, 22(1):11–36, 2021.

[220] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European conference on computer vision (ECCV) workshops*, pages 0–0, 2018.

[221] Zhaowen Wang, Ding Liu, Jianchao Yang, Wei Han, and Thomas Huang. Deep networks for image super-resolution with sparse prior. In *Proceedings of the IEEE international conference on computer vision*, pages 370–378, 2015.

[222] Angela C Webster, Evi V Nagler, Rachael L Morton, and Philip Masson. Chronic kidney disease. *The lancet*, 389(10075):1238–1252, 2017.

[223] Wen Wei, Emilie Poirion, Benedetta Bodini, Stanley Durrleman, Olivier Colliot, Bruno Stankoff, and Nicholas Ayache. Flair mr image synthesis by using 3d fully convolutional networks for multiple sclerosis. In *ISMRM-ESMRMB 2018-Joint Annual Meeting*, pages 1–6, 2018.

[224] Jeremy West, Dan Ventura, and Sean Warnick. Spring research presentation: A theoretical foundation for inductive transfer. *Brigham Young University, College of Physical and Mathematical Sciences*, 1(08), 2007.

[225] Jelmer M Wolterink, Anna M Dinkla, Mark HF Savenije, Peter R Seevinck, Cornelis AT van den Berg, and Ivana Išgum. Deep mr to ct synthesis using unpaired data. In *International workshop on simulation and synthesis in medical imaging*, pages 14–23. Springer, 2017.

[226] Wenchuan Wu and Karla L Miller. Image formation in diffusion mri: a review of recent technical developments. *Journal of Magnetic Resonance Imaging*, 46(3):646–662, 2017.

[227] Lei Xiang, Qian Wang, Dong Nie, Lichi Zhang, Xiyao Jin, Yu Qiao, and Dinggang Shen. Deep embedding convolutional neural network for synthesizing ct image from t1-weighted mr image. *Medical image analysis*, 47:31–44, 2018.

[228] Chih-Yuan Yang, Chao Ma, and Ming-Hsuan Yang. Single-image super-resolution: A benchmark. In *European conference on computer vision*, pages 372–386. Springer, 2014.

[229] Qianye Yang, Nannan Li, Zixu Zhao, Xingyu Fan, EI-Chao Chang, Yan Xu, et al. Mri image-to-image translation for cross-modality image registration and segmentation. *arXiv preprint arXiv:1801.06940*, 2018.

[230] Qianye Yang, Nannan Li, Zixu Zhao, Xingyu Fan, Eric I Chang, Yan Xu, et al. Mri cross-modality image-to-image translation. *Scientific reports*, 10(1):1–18, 2020.

[231] Qingsong Yang, Pingkun Yan, Yanbo Zhang, Hengyong Yu, Yongyi Shi, Xuanqin Mou, Mannudeep K Kalra, Yi Zhang, Ling Sun, and Ge Wang. Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*, 37(6):1348–1357, 2018.

[232] Xiao Yang, Roland Kwitt, Martin Styner, and Marc Niethammer. Quicksilver: Fast predictive image registration–a deep learning approach. *NeuroImage*, 158:378–396, 2017.

[233] Jong Chul Ye. Compressed sensing mri: a review from signal processing perspective. *BMC Biomedical Engineering*, 1(1):1–17, 2019.

[234] Xin Yi, Ekta Walia, and Paul Babyn. Generative adversarial network in medical imaging: A review. *Medical image analysis*, 58:101552, 2019.

[235] Biting Yu, Luping Zhou, Lei Wang, Jurgen Fripp, and Pierrick Bourgeat. 3d cgan based cross-modality mr image synthesis for brain tumor segmentation. In *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*, pages 626–630. IEEE, 2018.

[236] Biting Yu, Luping Zhou, Lei Wang, Yinghuan Shi, Jurgen Fripp, and Pierrick Bourgeat. Sample-adaptive gans: Linking global and local mappings for cross-modality mr image synthesis. *IEEE transactions on medical imaging*, 39(7):2339–2350, 2020.

[237] G Zaharchuk, E Gong, M Wintermark, D Rubin, and CP Langlotz. Deep learning in neuroradiology. *American Journal of Neuroradiology*, 39(10):1776–1784, 2018.

[238] Kun Zeng, Hong Zheng, Congbo Cai, Yu Yang, Kaihua Zhang, and Zhong Chen. Simultaneous single-and multi-contrast super-resolution for brain mri images based on a convolutional neural network. *Computers in biology and medicine*, 99:133–141, 2018.

[239] Wei Zeng, Jie Peng, Shanshan Wang, and Qiegen Liu. A comparative study of cnn-based super-resolution methods in mri reconstruction and its beyond. *Signal Processing: Image Communication*, 81:115701, 2020.

[240] Di Zhang, Jiazhong He, Yun Zhao, and Minghui Du. Mr image super-resolution reconstruction using sparse representation, nonlocal similarity and sparse derivative prior. *Computers in biology and medicine*, 58:130–145, 2015.

[241] Kuan Zhang, Haoji Hu, Kenneth Philbrick, Gian Marco Conte, Joseph D Sobek, Pouria Rouzrokh, and Bradley J Erickson. Soup-gan: Super-resolution mri using generative adversarial networks. *arXiv preprint arXiv:2106.02599*, 2021.

[242] Qianqian Zhang, Haifeng Wang, Hongya Lu, Daehan Won, and Sang Won Yoon. Medical image synthesis with generative adversarial networks for tissue recognition. In *2018 IEEE International Conference on Healthcare Informatics (ICHI)*, pages 199–207. IEEE, 2018.

[243] Yongqin Zhang, Pew-Thian Yap, Liangqiong Qu, Jie-Zhi Cheng, and Dinggang Shen. Dual-domain convolutional neural networks for improving structural information in 3 t mri. *Magnetic resonance imaging*, 64:90–100, 2019.

[244] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018.

[245] Can Zhao, Aaron Carass, Blake E Dewey, and Jerry L Prince. Self super-resolution for magnetic resonance images using deep networks. In *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*, pages 365–368. IEEE, 2018.

[246] Shengjia Zhao, Jiaming Song, and Stefano Ermon. Infovae: Information maximizing variational autoencoders. *arXiv preprint arXiv:1706.02262*, 2017.

[247] Xiaole Zhao, Yulun Zhang, Tao Zhang, and Xueming Zou. Channel splitting network for single mr image super-resolution. *IEEE Transactions on Image Processing*, 28(11):5649–5662, 2019.

[248] Jiannan Zheng, Shun Miao, Z Jane Wang, and Rui Liao. Pairwise domain adaptation module for cnn-based 2-d/3-d registration. *Journal of Medical Imaging*, 5(2):021204, 2018.

[249] Tao Zhou, Huazhu Fu, Geng Chen, Jianbing Shen, and Ling Shao. Hi-net: hybrid-fusion network for multi-modal mr image synthesis. *IEEE transactions on medical imaging*, 39(9):2772–2781, 2020.

[250] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In *Proceedings of the IEEE international conference on computer vision*, pages 2223–2232, 2017.

[251] Fuzhen Zhuang, Zhiyuan Qi, Keyu Duan, Dongbo Xi, Yongchun Zhu, Hengshu Zhu, Hui Xiong, and Qing He. A comprehensive survey on transfer learning. *Proceedings of the IEEE*, 109(1):43–76, 2020.