
# Unsupervised Simultaneous Denoising and Cross-Modality Synthesis of Medical Image via Generative Adversarial Networks

*Author:*
Junyu Jiang

*Supervisor:*
Dr. Charith Abhayaratne

*A thesis submitted in fulfilment of the requirements for the degree of Master of Philosophy*

*in the*

Department of Electronic and Electrical Engineering
Faculty of Engineering
The University of Sheffield

3rd February 2026

# Declaration of Authorship

I, Junyu Jiang, declare that this thesis entitled, 'Unsupervised Simultaneous Denoising and Cross-Modality Synthesis of Medical Image via Generative Adversarial Networks' and the work presented in it are my own. I confirm that:

- This work was done wholly while in candidature for a research degree at this university.
- Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
- Where I have consulted the published work of others, this is always clearly attributed.
- Where I have quoted from the work of others, the source is always given. With the exception of such quotations, this thesis is entirely my own work.
- I have acknowledged all main sources of help.
- Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Signed: Junyu Jiang

Date: 03/02/2026

# *Abstract*

Multi-modality medical imaging provides complementary anatomical and functional information that is essential for accurate diagnosis and treatment planning. In clinical practice, however, the acquisition of multiple high-quality imaging modalities is often constrained by long scanning times, high financial cost, and patient discomfort. Medical images are also frequently degraded by noise introduced during acquisition, particularly under low-dose or high-speed scanning conditions. Although denoising and cross-modality synthesis are closely related in real-world workflows, existing methods commonly treat them as independent problems.

This thesis presents an unsupervised unified framework for the simultaneous denoising of medical images and cross-modality image synthesis. A Multi-Channel Asymmetric Residual Generator (MARG) is proposed to enhance noise suppression while preserving fine anatomical structures. In addition, a Dual-Channel Joint Discriminator (DCD) is developed to improve modality translation by jointly enforcing local structural consistency and pixel-level intensity distribution. These components are integrated into a single adversarial learning framework that enables concurrent denoising and cross-modality synthesis.

Experiments conducted on low-dose chest CT and multi-modal brain MRI datasets demonstrate that the proposed methods outperform state-of-the-art unsupervised baselines in terms of PSNR, SSIM, and MS-SSIM. For simultaneous denoising and synthesis tasks, the proposed framework achieves improved image fidelity and structural consistency compared with conventional sequential processing strategies across multiple noise levels.

By reducing reliance on paired training data and mitigating error accumulation associated with multi-stage pipelines, the proposed approach offers a practical and scalable solution for enhancing image quality and modality availability in clinical environments. The framework has the potential to improve diagnostic reliability, workflow efficiency, and overall patient experience.

# *Acknowledgements*

On the occasion of the completion of this thesis, I would like to thank my supervisor, family, and friends for their expectations and encouragement. At this moment, I cannot find the right words to express my deepest and most sincere gratitude.

First of all, I would like to thank my supervisor, Dr. Charith Abhayaratne, without whom this work would not have been possible. He is responsible, has excellent professional knowledge, and has given me a lot of valuable advice and help throughout the thesis writing process. I learned new knowledge and skills every time I met with him, which helped me tremendously in the whole research period.

I would like to thank my wife, He Xiaoqian. She took care of my life, accompanied, supported, and encouraged me, and provided me with the motivation to keep working hard. It would have been difficult for me to stay in the process without her.

I would also like to thank my daughter, Ophelia Jiang, whose birth has given me even more motivation to complete my studies. Her little figure and laughter are a comfort and support to me every day.

I would like to thank my parents and my in-laws. Thank you for your silent support, understanding, trust, and expectation over the years, which have been the driving force for me to move forward. Now that this thesis is finally finished, I have finally reached our goal.

I would also like to thank my best friend and colleague, Roy Xue, for all the help and support he has given me.

Last but not least, thanks again to all those who have cared for me and helped me.

# Contents

# List of Figures

# List of Tables

# Abbreviations

| | |
|---|---|
| **AAE** | **A**dversarial **A**uto **E**ncoder |
| **ACGAN** | **A**uxiliary **C**lassifier **G**enerative **A**dversarial **N**etworks |
| **AGE** | **A**dversarial **G**enerator **E**ncoder |
| **ALI** | **A**dversarially **L**earned **I**nference |
| **ANN** | **A**rtificial **N**eural **N**etworks |
| **Bi-GAN** | **Bi**directional **G**enerative **A**dversarial **N**etworks |
| **CAT** | **C**omputerised **A**xial **T**omography |
| **CGAN** | **C**onditional **G**enerative **A**dversarial **N**etworks |
| **CLT** | **C**entral **L**imit **T**heorem |
| **CMS** | **C**ross **M**odality **S**ynthesis |
| **CMS-DN** | **C**ross **M**odality **S**ynthesis and **D**e**N**oising |
| **CNN** | **C**onvolutional **N**eural **N**etworks |
| **CMS** | **C**ross **M**odality **S**ynthesis |
| **CT** | **C**omputed **T**omography |
| **DCD** | **D**ual **C**hannel joint **D**iscriminator |
| **DCE** | **D**ynamic **C**ontrast **E**nhanced |
| **DCGAN** | **D**eep **C**onvolutional **G**enerative **A**dversarial **N**etworks |
| **DeBiGAN** | **De**noising **Bi**directional **G**enerative **A**dversarial **N**etworks |
| **DWI** | **D**iffusion **W**eighted **I**mages |
| **EM** | **E**arth **M**over |
| **FL** | **F**uzzy **L**ogic |
| **GA** | **G**enetic **A**lgorithms |

| | |
|---|---|
| **GANs** | **G**enerative **A**dversarial **N**etworks |
| **HR** | **H**igh **R**esolution |
| **JS** | **J**ensen **S**hannon |
| **LDCT** | **L**ow **D**ose **C**omputed **T**omography |
| **LMMSE** | **L**inear **M**inimum **M**ean **S**quare **E**rror |
| **LR** | **L**ow **R**esolution |
| **MAR** | **M**ulti-channel **A**symmetric **R**esidual |
| **MARG** | **M**ulti-channel **A**symmetric **R**esidual **G**enerator |
| **MLP** | **M**ulti **L**ayer **P**erceptron |
| **MRA** | **M**agnetic **R**esonance **A**ngiography |
| **MRI** | **M**agmetic **R**esonance **I**maging |
| **MSSSIM** | **M**ulti **S**cale **S**tructural **Sim**ilarity |
| **NDCT** | **N**ormal **D**ose **C**omputed **T**omography |
| **NLM** | **N**on **L**ocal **M**eans |
| **PD-w** | **P**roton **D**ensity **w**eighted |
| **PET** | **P**ositron **E**mission **T**omography |
| **PSNR** | **P**eak **S**ignal to **N**oise **R**atio |
| **SGD** | **S**tochastic **G**radient **D**escent |
| **SPECT** | **S**ingle **P**hoton **E**mission **C**omputed **T**omography |
| **SR** | **S**uper **R**esolution |
| **SRGAN** | **S**uper **R**esolution **G**enerative **A**dversarial **N**etworks |
| **SSIM** | **S**tructural **Sim**ilarity |
| **T1-w** | **T1 w**eighted |
| **T2-w** | **T2 w**eighted |
| **US** | **U**ltra **S**ound |
| **VAE** | **V**ariational **A**uto **E**ncoders |
| **WGAN** | **W**asserstein **G**enerative **A**dversarial **N**etworks |
| **WSM** | **W**avelet **S**ub-band coefficient **M**ixing |

# Chapter 1

# Introduction

Medical imaging has gained increasing importance in the fields of disease treatment, clinical diagnosis, and medical research. Medical imaging commonly generates images using positron emission tomography (PET), computed tomography (CT), and magnetic resonance imaging (MRI), among others. The acquisition of high-quality images is achieved at the expense of costly equipment, patient comfort, and long scanning times. The presence of such uncertainties might result in incomplete recordings as a consequence of data corruption, loss, noise, or image artefacts.

Despite the significant contributions of multi-modality medical imaging to disease prevention, detection, and treatment, in addition to the ongoing advancements in technology aimed at enhancing human health and well-being, the estimation of modality transformations based on anatomical and functional contrasts between scans remains a challenging problem. Although prior studies have examined the fact that combining multimodality with high-quality medical images has numerous benefits, existing approaches often address cross-modality synthesis and image denoising as separate tasks. As a result, there are limited methods that sufficiently in-

tegrate these two processes within a single framework.

The objective of this thesis is to investigate medical image denoising and cross-modality synthesis, and to explore their simultaneous implementation using an integrated model. As this thesis focuses on medical image denoising and cross-modality synthesis, the medical image noise and modality will be introduced in detail in the next section. In order to have a visual impression of medical image noise and different modalities, Fig. 1.1 shows some examples of medical image noise, and Fig. 1.2 illustrates representative examples of different medical image modalities.

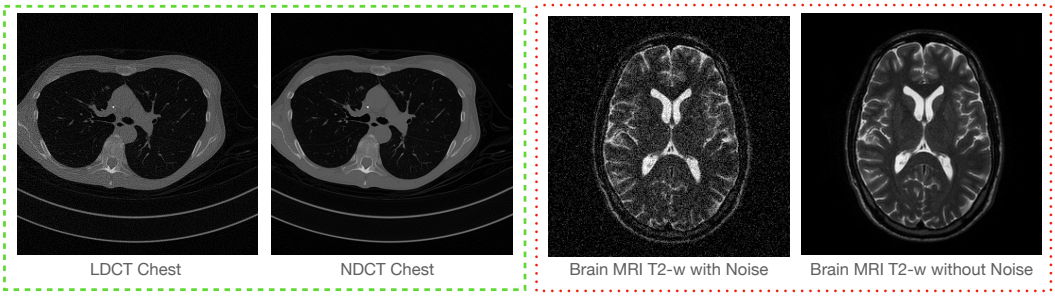


**Figure 1.1: Examples of medical image noise (from left to right): LDCT Chest image (With noise), NDCT Chest image (Without noise), Brain MRI T2-weighted image (With noise), Brain MRI T2-weighted image (Without noise)**

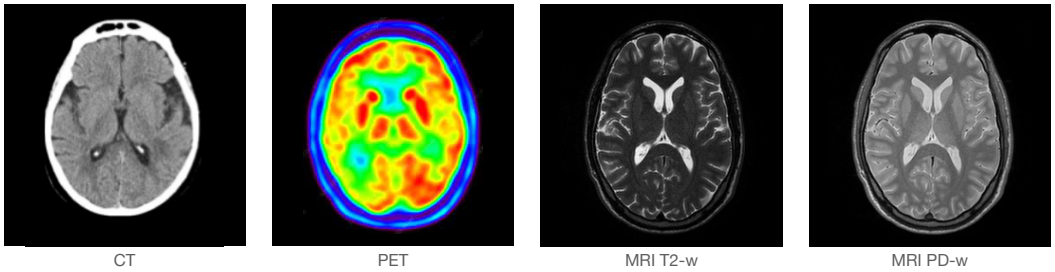


**Figure 1.2: Examples of different medical image modalities (from left to right): Brain CT image, Brain PET image, Brain MRI T2-weighted image, Brain MRI PD-weighted image.**

## 1.1 Medical Image Noise

In recent decades, medical imaging and diagnostic techniques have undergone rapid development and have evolved into an integral component of disease diagnosis. Medical images provide an internal view of the human body by conveying information about the heart, brain, chest, and other vital organs. A multitude of mathematical algorithms can be implemented on medical images to support the identification of pathological abnormalities. Nonetheless, any loss of critical information during medical image acquisition could potentially lead to adverse clinical outcomes. The primary challenge encountered during the medical imaging procedure is acquiring images while minimising the loss of crucial information.

There is a strong likelihood that the acquired images may be tainted by artefacts or noise during the acquisition and/or subsequent processing stages. Noise is characterised by the stochastic variation of pixel intensity values, which is particularly detrimental to objects that are small in size and possess a low contrast ratio. In contrast to conventional images, the contrast of medical images is generally low, making them more susceptible to noise. Noise can therefore result in image degradation and hinder the identification of diseases. Therefore, medical image denoising has become an essential and widely adopted pre-processing stage in medical image processing. Unlike natural image noise, the majority of medical images exhibit signal-dependent noise, which is difficult to remove using conventional natural image denoising methods.

Medical image noise can be caused by various factors, all of which negatively affect image quality and effectiveness. Sensors are used by medical imaging instruments to detect a variety of signals, such as ultrasound, X-rays, and magnetic resonance signals. The inherent sensor disturbances, referred to as sensor noise, can be sporadic or systematic and may compro-

mise image accuracy. In X-ray imaging, the image noise level is influenced by the magnitude of the radiation dose. An inadequate radiation dose may result in noise amplification, while an excessive radiation dose may increase the patient's radiation exposure.

Motion-related artefacts may arise as a consequence of patient respiration or movement during specific medical imaging procedures. These artefacts hinder image interpretation and analysis and can be treated as noise. Images are also susceptible to degradation under low-contrast conditions. Low contrast may arise due to tissue homogeneity or inappropriate selection of imaging parameters. In addition, certain medical imaging equipment may be affected by systematic noise, which is typically associated with the equipment itself or with specific imaging process factors. For example, systematic noise can be observed as ringing artefacts in CT imaging.

Electromagnetic interference may also cause image noise when a medical imaging device operates in a disturbed electromagnetic environment. Scattered radiation is a significant contributor to noise in X-ray imaging and can cause background degradation. This radiation is generated when matter undergoes scattering rather than transmission. MRI images can be affected by magnetic field inhomogeneities, gradient coil inhomogeneities, and patient motion, among other factors, all of which can result in artefacts and noise.

## 1.2 Medical Image Modalities

Medical imaging consists of a variety of modalities, which include X-ray computed tomography, ultrasound, positron emission tomography, magnetic resonance imaging, and electrical impedance tomography, among others. The various modalities of an image refer to images obtained using vari-

ous imaging methods or techniques, each of which can provide information about distinct anatomical or functional aspects of the subject being imaged.

X-ray imaging employs X-ray radiation to penetrate matter and record images, which can be used to detect internal structures such as the chest and abdomen. CT imaging uses X-ray radiation to produce a three-dimensional tomographic image of the object by rotating a scan. It is commonly used in medical diagnosis to detect abnormalities and internal structures. Ultrasound imaging employs high-frequency sound waves to create images and is commonly used in obstetrics and examinations of the heart, the liver, and the kidneys, enabling the observation of organ movement in real time. PET imaging employs radioactive tracers to measure metabolic activity and provide information about organ and tissue function. SPECT stands for single-photon emission computed tomography. SPECT is commonly used to create images of the heart, bones, and brain using a radioactive tracer.

MRI utilises a powerful magnetic field and non-ionising radio radiation to generate high-resolution images of soft tissues, such as the brain, muscles, and organs. These various imaging modalities have a wide range of applications in medicine and scientific research, and they offer a variety of methods for obtaining diverse information about the object, thereby contributing to medical diagnosis and scientific research. The proper selection and combination of these modalities can provide more comprehensive information to support accurate diagnoses.

Recent studies have focused on recovering the missing modalities that may exist in various MRI modalities in order to capture diverse characteristics of the underlying anatomy. Since this study focuses on the cross-modality synthesis of brain MRI, the details presented below are mainly concentrated on the different modalities of MRI.

MRI is used to generate cross-sectional views of the physiological and

anatomical processes of the human organism. The advantage of MRI is that it is non-ionising, with no harmful ionising radiation for the patient. MRI uses strong magnetic fields to generate the image acquisitions, which can provide anatomical and functional diagnostic information. The contrast between the different modalities of MRI is determined by the number and magnetic properties of hydrogen nuclei. Clinicians and researchers can acquire comprehensive information from multi-modality MRI, which is beneficial for the characterisation of lesions.

MRI provides the soft tissues with anatomical information that is used to examine brain activity. The contrast of MRI can be modified through the execution of acquisition sequences that possess different weightings. Proton density-weighted (PD-w), T1-weighted (T1-w), and T2-weighted (T2-w) are the three principal parameters.

T1-w images are more sensitive to the T1 relaxation time in the tissue, emphasising the contrast of tissue types. In T1-w images, fat usually appears as a high signal (bright), while water and other soft tissues are relatively dark. T1-w images are commonly used to display anatomical structures, such as identifying tissue types, observing anatomical details, and evaluating the characteristics of tumours. T2-w images are more sensitive to T2 relaxation time in tissues and emphasise the contrast of the water distribution. In T2-w images, water molecules usually appear high-signal (bright), while fat and other tissues are relatively dark. T2-w images are commonly used for detecting oedema, inflammation, multiple sclerosis, and related conditions. It is also useful for displaying the anatomical structures of the brain and the spinal cord.

PD-w images emphasise the density of protons in tissues, which is sensitive to the distribution of water and the type of tissue. In PD-w images, the contrast between water and other tissues is usually relatively homogeneous. PD-w images are widely used in joint imaging, particularly for the detection

and assessment of soft tissues in joints. It is also used to assess water distribution in neural and other tissues. These three modalities are often used together to obtain more comprehensive image information. The physician will select the appropriate modality or combination of modalities to support diagnostic and therapeutic decision-making based on clinical needs. The use of different MRI modalities in combination helps to integrate multiple aspects of tissue information and improve understanding of disease and anatomy.

## 1.3 Challenges

Medical image denoising is a complex problem with multiple challenges. There are many different types of noise that can affect medical images, including system noise, photoreceptor noise, motion noise, and other noise sources. Different noise types require different denoising methods. To reduce radiation exposure, medical images are usually acquired at low radiation doses, which leads to a decrease in the signal-to-noise ratio (SNR), making noise more significant. Under low-dose conditions, effective noise removal becomes increasingly difficult.

During denoising, a balance must be achieved between noise suppression and the preservation of fine anatomical structures, as excessive denoising may remove clinically relevant details and adversely affect diagnosis. In addition, medical images often contain complex anatomical patterns and small-scale structures that are highly sensitive to noise, placing further demands on denoising algorithms in terms of robustness and generalisation.

Beyond denoising, medical image cross-modality synthesis introduces additional challenges. Different imaging modalities exhibit substantial variations in contrast, resolution, and noise characteristics. Ensuring structural

consistency while synthesising realistic target modalities is non-trivial, particularly in the presence of noisy inputs. Moreover, synthesis models must generalise well across diverse datasets and imaging conditions.

In certain clinical application scenarios, such as surgical navigation and intraoperative guidance, there is also a strong demand for real-time or near real-time synthesis performance. Therefore, synthesis algorithms are required to generate high-quality images within limited computational time, further increasing the complexity of model design.

### 1.3.1 Research Gap

Despite extensive research on medical image denoising and cross-modality synthesis, most existing studies continue to treat these two problems as independent tasks. In conventional medical image processing pipelines, denoising is typically performed as a pre-processing step, followed by cross-modality synthesis in a separate stage. Such sequential strategies fail to explicitly capture the interaction between noise characteristics and modality-specific representations, which can limit performance under low-dose or noisy imaging conditions.

In addition, many cross-modality synthesis approaches rely on paired or well-aligned multi-modality data, which are costly and difficult to obtain in real clinical settings. Although unpaired learning frameworks partially alleviate this requirement, their robustness to noise degradation remains limited. These challenges highlight the need for integrated approaches that jointly address medical image denoising and cross-modality synthesis within a unified learning framework.

To address the above limitations, this thesis introduces a series of methodological contributions aimed at jointly modelling noise characteristics and

cross-modality relationships in medical imaging.

## 1.4 Contributions

In this thesis, we have invented several models to address the challenges of multi-modality medical image acquisition and noise degradation in acquired images. Specifically, the main contributions of this thesis are summarised as follows:

1. A new multi-channel asymmetric residual block and generator architecture were proposed for medical image denoising. Based on this design, an unsupervised low-dose CT chest image denoising framework was developed, achieving state-of-the-art performance.

2. A dual-channel discriminator structure was constructed to provide adversarial feedback to the generator from both local structural information and pixel-level space. This design achieved state-of-the-art results in an unsupervised brain MRI cross-modality synthesis task.

3. A pioneering framework for simultaneous medical image denoising and cross-modality synthesis was proposed. By integrating the multi-channel asymmetric residual generator and the dual-channel discriminator, a unified network architecture for dual-task learning was developed. A dedicated experimental evaluation procedure for simultaneous multitasking was designed, and state-of-the-art performance was achieved under this framework.

## 1.5 Thesis Outline

This thesis begins with the background and significance of the study, and the current challenges in medical image processing. The rest of this thesis is organised as follows:

Chapter 2 is an introduction to the related work, which begins with Generative Adversarial Networks and their variants. Subsequently, a detailed analysis of their utilisation in medical image denoising and cross-modality synthesis is presented.

Chapter 3 provides a detailed description of low-dose chest CT image denoising. The unsupervised bidirectional adversarial network is introduced and utilised to handle the denoising task of LDCT chest images.

Chapter 4 focuses on unsupervised brain MRI cross-modality synthesis using generative adversarial networks. A dual-channel discriminator network structure is proposed to obtain state-of-the-art results.

Chapter 5 describes and discusses the possibility of simultaneous cross-modality synthesis and denoising of medical images. From a practical point of view, the multi-channel asymmetric residual generator structure is employed as the generator, and the dual-channel discriminator structure is employed as the discriminator, to create a new network structure for performing both tasks simultaneously. The effectiveness of the model is demonstrated through extensive experimental evaluation.

In Chapter 6, the entire research is summarised. Additionally, it discusses potential directions for future work.

# Chapter 2

# Related Work

In this chapter, the Generative Adversarial Networks (GANs) and their variants have been introduced, followed by a specific review of their application in cross-modality medical image synthesis and medical image denoising. In the field of cross-modality synthesis of medical images, GANs and their variants are employed to transform images from one modality to another. Such transformations can have important applications in medical imaging, including converting CT images to MRI images or PET images to CT images, as well as synthesising from one MRI modality to another. By training the generator network, the model acquires knowledge of the mapping relationships between modalities and generates realistic synthetic images. GANs are also extensively applied to medical image denoising. Noise is a prevalent issue affecting the quality and dependability of medical images. Image denoising can be accomplished by training a generator network that transforms noisy images into clear ones. Typically, these methods incorporate loss functions, regularisation techniques, and reconstruction constraints to enhance the effect of denoising. In addition to traditional GANs, several variants and enhanced network structures have been applied to cross-modality synthesis and image denoising tasks. Conditional GANs

(cGANs) [103] introduce conditional information, which permits generators and discriminators to be subject to additional constraints and guidance. In the absence of paired training data, CycleGAN [178] and UNIT [86] can be used to train a cross-modality synthesis model. Several studies have also investigated the application of techniques like multi-scaling, attention mechanisms, and residual connectivity. These reviews provide a clear understanding of the applications and prospects of GANs in the fields of medical image denoising and cross-modality synthesis.

## 2.1 Generative Adversarial Networks

Since 2014, when Ian Goodfellow introduced Generative Adversarial Networks (GAN) [45], it has become a popular research topic that has been extensively investigated, and numerous algorithms have been proposed. GANs are constructed from two different models, which have been referred to as a generator and a discriminator. These two models are commonly realised through neural networks. They can be achieved through the application of any type of distinguishable system that maps data from one domain to another. The generator makes an effort to obtain an accurate representation of the distribution of real examples before generating new data examples. In most cases, the discriminator is a binary classifier, and its job is to differentiate between generated examples and real examples in the most precise manner possible. Minimax optimisation is the type of challenge that arises when trying to optimise GANs. The optimisation procedure concludes at a saddle point, which is the minimum of the generator and the maximum of the discriminator. Alternatively, the objective of optimisation is to attain Nash equilibrium [119]. Then, it is reasonable to conclude that the generator has effectively captured the actual distribution of real examples. The architecture of GANs is illustrated in Fig 2.1. The generator $G$

is used to generate data that is as close as possible to the true distribution. The purpose of the discriminator $D$ is to distinguish between the data $G(z)$ generated by the generator $G$ and the real data $x$. Typically, the input to the generator is a random noise vector $z$. Through the generator $G$, $z$ is mapped into a new vector space, resulting in a generated multi-dimensional vector $G(z)$. In general, a discriminator $D$ is a binary classifier that takes in the generated data $G(z)$ and the real data $x$, and outputs a discriminant based on the probability that the received sample is real or fake. When the discriminator does not detect that the received data is true or generated data, it means that the generator $G$ has been well trained and is capable of generating relatively real data.

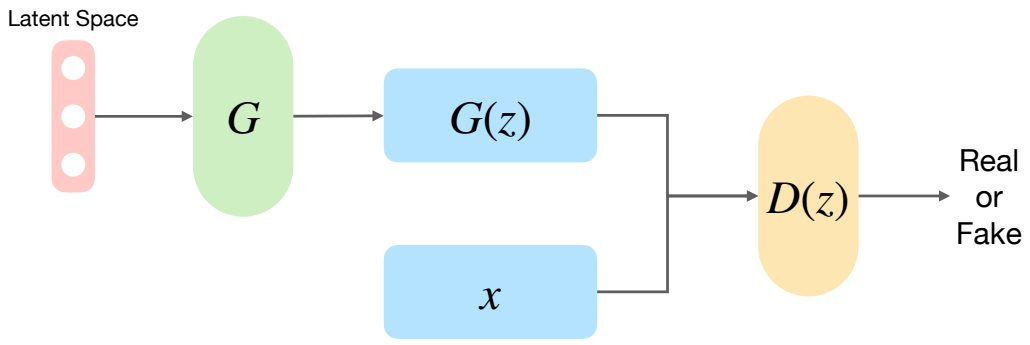


**Figure 2.1: The architecture of GAN**

## 2.2 Variations of Generative Adversarial Networks

Due to the rise of GANs, an increasing number of researchers have proposed improved models based on the original theory of GANs. These GAN

models can be divided into two categories: architecture-optimisation-based GANs and objective-function-optimisation-based GANs. To this day, GANs are still undergoing rapid development. In the course of this development, many representative models have been proposed. A selection of representative GAN models based on the above two categories will be introduced in this section.

### 2.2.1 Architecture Optimisation Based GAN

#### Convolution Based GAN

Initially, a Multi-Layer Perceptron (MLP) mechanism was used for the generator and discriminator network structure. However, MLP is inferior to Convolutional Neural Networks (CNN) [77] in terms of extracting image features. Redford *et al.* [116] combined CNN with GAN, thus proposing Deep Convolutional Generative Adversarial Networks (DCGAN). By substituting the fully connected layers in the generator with deconvolutional layers, this approach achieves strong performance in image generation tasks. This strategy replaces the fully connected layers in the generator with deconvolutional layers, thereby improving image generation capability.

#### Condition Based GAN

Model collapse is a serious problem for GAN, which is primarily due to the generator's input consisting of unconstrained random noise $z$. This may result in the collapse of the model. To address this problem, Mirza and Osindero [103] developed Conditional Generative Adversarial Networks (CGAN). The principle is to incorporate a conditional variable $c$ in the generator and discriminator, as well as augment the model with supplementary restrictive

information that impacts the data generation procedure. The structure of the CGAN is shown in Fig 2.2. The generator input is the stochastic noise vector $z$ and the conditional variable $c$. The same conditional variable $c$ also controls the real sample $X$, and then the real sample $X$ and $G(z|c)$ (from the generator) are the inputs of the discriminator.

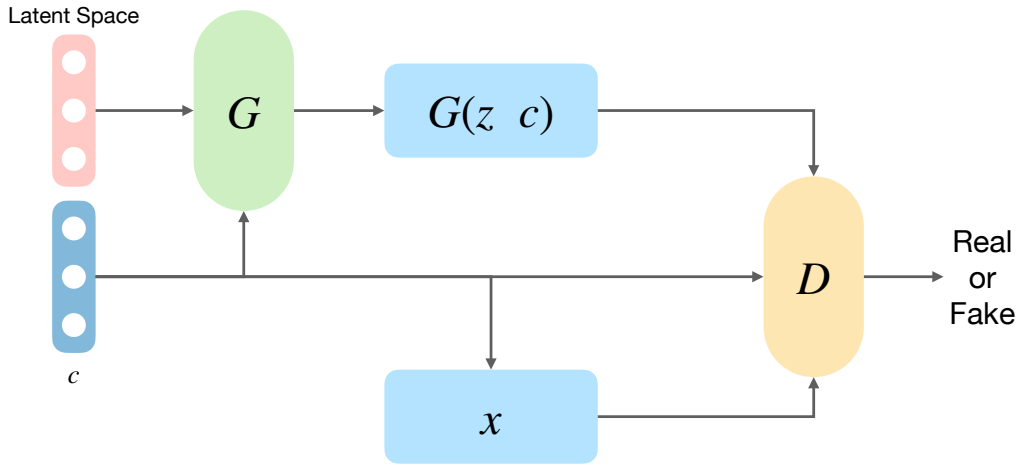


**Figure 2.2: The architecture of CGANs**

In the period after CGAN was proposed, its great potential was discovered. So in the following years, many improved models were proposed based on CGAN; two of the classical models are reviewed here as examples. The first one is InfoGAN, which has been proposed by Chen *et al.* [23]. The network structure of InfoGAN has been shown in Fig 2.3. InfoGAN introduces mutual information to make the process of generation more manageable. Maximising the mutual information can strengthen the correlation between the generated data $x$ and the latent code $c$. The InfoGAN generator is similar to CGAN, with the exception that the latent code $c$ is required to be discovered through training. The discriminator has been substantially modified so that the output condition variable $Q(c|x)$ includes a supplemen-

tal network $Q$ along with the original GANs' discriminator. This mutual information constraint increases the plausibility of the generated data. The second one is Auxiliary Classifier GAN (ACGAN), which was proposed by Odena *et al.* [113]. Compared to InfoGAN, the discriminator of ACGAN does not include the conditional variable $c$. An alternative classifier will be implemented to display the probability associated with the class labels. Subsequently, the class prediction accuracy is enhanced by adjusting the loss function. The condition variable $c$ will not be added to the discriminator, and another classifier will be used to display the probability over the class labels. The loss function is then modified to enhance the likelihood of class prediction accuracy.

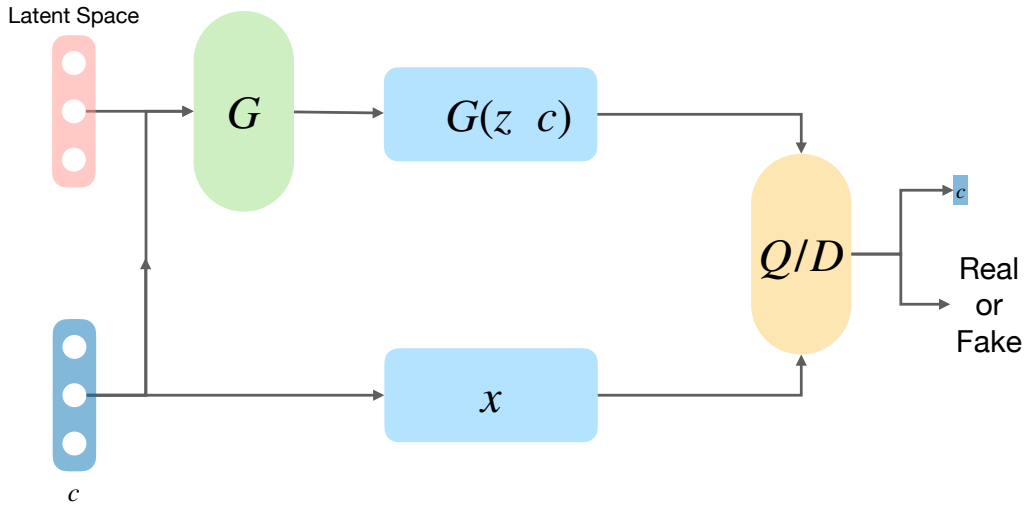


**Figure 2.3: The architecture of InfoGANs**

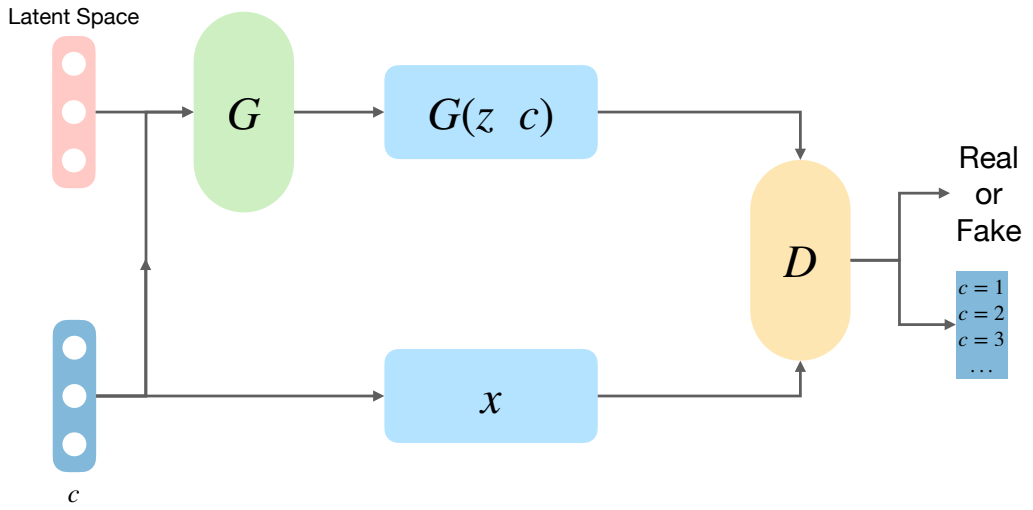


**Figure 2.4: The architecture of ACGANs**

### Autoencoder Based GAN

An autoencoder consists of two main components: an encoder and a decoder. The encoder transforms the input into a hidden representation through dimensionality reduction, which is then passed to the decoder for reconstruction. Since the autoencoder training process does not require labels, it is an unsupervised learning method. Autoencoders have been widely utilised in combination with latent variable modelling for generative tasks in recent years.

Nonetheless, the uneven distribution of the hidden layers in the obtained encoder in the specified space results in a significant number of distribution gaps. In response to these drawbacks, Makhzani *et al.* [93] proposed the Adversarial Autoencoder (AAE), which is based on the idea of combining an autoencoder with an adversarial network. In order to guarantee that the prior distribution has no gaps, AAE imposes an arbitrary prior distribution on the hidden distribution arrived at by the encoder, thus

allowing the decoder to reconstruct meaningful samples from an arbitrary fraction.

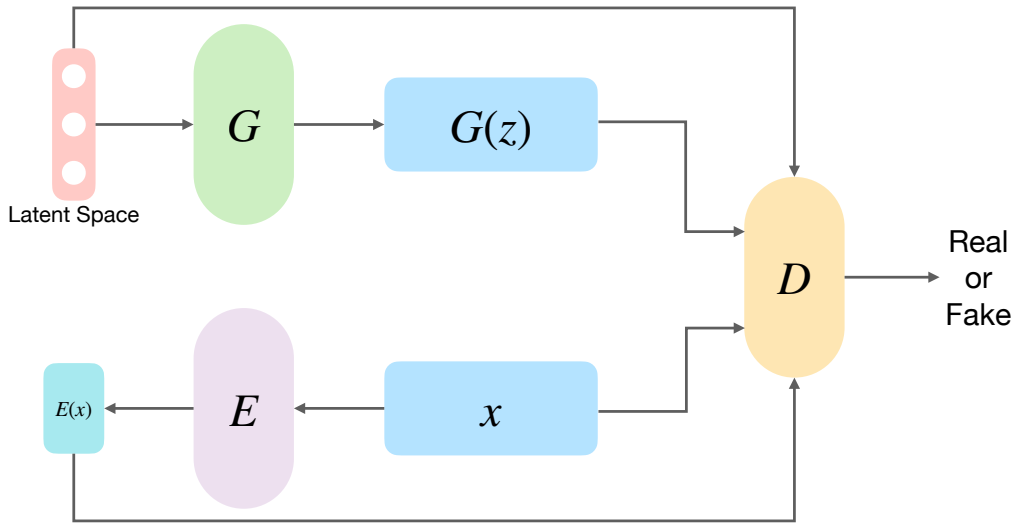


**Figure 2.5: The architecture of BIGANs**

After AAE has been proposed, some modules combine AAE and GAN together, adding the encoder to GAN. Donahue *et al.* [35] introduced a novel approach called Bidirectional Generative Adversarial Networks (BiGAN), which significantly improves the quality of generated samples. By adding an additional encoder that inversely projects the generated data distribution into the latent space, additional features can be extracted for the discriminator. Meanwhile, another similar model, Adversarially Learned Inference (ALI), was proposed by Dumoulin *et al.* [37]. In terms of network structure, ALI and BiGAN are almost identical. And both methods train generators and encoders simultaneously. Among the approaches that combine both AAE and GAN, the one that differs from the two mentioned above is the Adversarial Generator-Encoder Network (AGE), which was proposed by Ulyanov *et al.* [137]. The AGE does not require the participation of a discriminator, and the adversarial network is functioning as an inter-

mediary between the encoder and the generator. In this network structure, the function of the generator is to minimise the difference between the latent noise distribution and the generated data distribution. The function of the encoder is to maximise the difference between the latent noise and the generated data distribution, while minimising the difference with the real sample distribution. In combination with the reconstruction loss function, the problem of model collapse was successfully avoided.

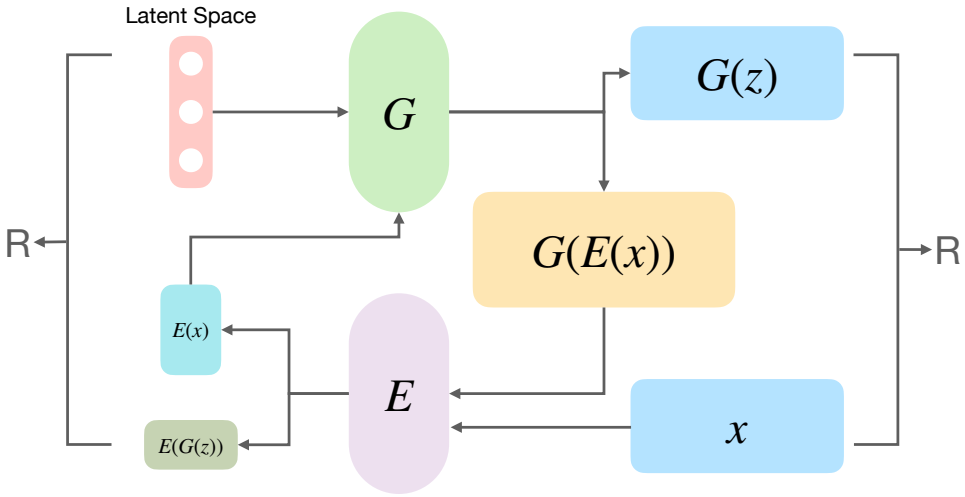


**Figure 2.6: The architecture of AGE**

### 2.2.2 Objective Function Optimisation Based GAN

As mentioned previously, the stability of GANs has been a non-negligible drawback, and a great deal of effort has been made by researchers to improve GAN stability. In this section, methods that improve the stability of GANs by optimising their objective functions are reviewed. In the original GAN, the Jensen-Shannon (JS) divergence was used to minimise the generator loss function. To enhance the stability of GAN, some studies [99, 112, 173] utilised different divergences to construct the objective func-

tion.

Some other methods are using different regularisations to enhance the stability of GAN. Che *et al.* [17] used two regularisers to stabilise the learning process. The problem of gradient disappearance has also been affecting the training of GAN. The cause of gradient disappearance is that there is no or only minimal overlap between the distribution of the real data and the generated data, in which case the divergence becomes a constant and so the gradient becomes zero. To address this issue, Arjovsky *et al.* [5] introduced the Wasserstein Generative Adversarial Networks (WGAN), which demonstrated that the Earth-Mover (EM) distance provides superior gradient behaviours in comparison to alternative distance metrics. In WGAN, the weight clipping technique was implemented to enforce the Lipschitz constraint; additionally, a new loss function was developed to address the challenge of the unstable training procedure. However, there is a possibility that the WGAN will fail to converge or generate some unreasonable results as a consequence of the weight clipping technique. Gulrajani *et al.* [49] proposed the WGAN-GP, which added a gradient penalty to the WGAN so that the Lipschitz constraint has been enforced and the performance is better than WGAN. Meanwhile, the WGAN-GP permits more stable training of various GANs architectures with minimal hyper-parameter tuning. Based on the WGAN-GP, Petzka *et al.* [114] introduced WGAN-LP, which appended an additional penalty term to further strengthen the Lipschitz constraint. This makes the training even more stable.

In addition, the original GAN presumes that the discriminator is capable of modelling infinitely without constraints on the distribution of the real sample. This can readily result in overfitting and a lack of generalisation ability. To restrict the infinite modelling capacity of GANs, the loss function obtained by minimisation of the objective function was constrained to a space satisfying Lipschitz continuous functions [115]. Therefore, the two

major problems affecting GAN training, mode collapse and vanishing gradient, are basically addressed by [115] and [5] with Lipschitz regularity.

## 2.3 Applications of GANs

Data generation is GANs' primary use as generative models. That is, learn from real sample distributions, then generate samples to match them. Some applications of GANs will be introduced in this section.

Since medical image denoising and cross-modality synthesis can be considered as image translation, we first introduce some applications of GANs in the field of image translation. Image translation is actually the conversion of image content from one domain to another. Based on conditional GAN, Isola *et al.* [65] proposed pix2pix for supervised image translation. Although this method achieved good results in the image translation task, its ability to process high-resolution images is insufficient. In order to solve this problem, pix2pixHD [143] improved the pix2pix method to enhance the generation ability so that it can generate $2048 \times 1024$ high-resolution images. However, since the pix2pix based method strictly requires that the image source domain and target domain are spatially paired, it is very difficult to obtain such paired data. Especially in the field of medical image processing, obtaining strictly spatially paired images requires high human resources and time costs. Therefore, it is crucial to find a method that does not require paired data. To solve this problem, DualGAN [160], DiscoGAN [71], and CycleGAN [178] were proposed almost at the same time, and they all adopt the idea of cycle consistency, which does not require strict spatial pairing of the source domain and target domain, and unpaired data can be used to train the mapping between the two domains. All of the above methods are based on image translation between two domains, and these models cannot solve the problem of image translation between multiple domains.

So Choi *et al.* [26] proposed StarGAN to solve the image translation problem between multiple domains and achieved good results.

GANs are widely used not only in the field of image translation but also in image super-resolution. Since medical image super-resolution is one of our future work directions, here we briefly introduce some image super-resolution related work from GANs. Ledig *et al.* [78] proposed Super-Resolution Generative Adversarial Networks (SRGAN), which can generate images with four times higher resolution than the original input image. However, the texture information generated by this model is not realistic enough, and it generates noise and artifacts while generating high-resolution images. To address this problem, Wang *et al.* [144] improved the SRGAN and proposed an Enhanced-SRGAN, which enhances SRGAN in terms of network structure, adversarial loss, and perceptual loss, respectively, so that it generates high-resolution images with effects beyond SRGAN.

In addition to this, GANs have been applied to generate high-quality realistic image samples such as faces, natural scenes, and artworks [14, 58, 120, 136, 169, 170]. GANs can also be used to generate text, dialogue systems, and translations [9,66,80]. In the field of natural language processing, GANs have also made some contributions [82, 85, 106, 166]. In addition to applications in image processing and language processing, GANs are also used in tasks such as video generation, video restoration, and video prediction. It is capable of generating realistic video sequences, video super-resolution processing, etc. The applications of GANs in various domains have demonstrated their great potential in generating and processing data, but they also face some challenges, such as training stability, pattern collapse, and other problems. With the continuous development of research, it is believed that GAN will achieve wider and far-reaching applications in more fields.

## 2.4 Medical Image Denoising

With the development of computing capability and data storage over the last two decades, medical imaging has gained tremendous popularity. Similar to other imaging techniques, medical images are affected by noise and artifacts. Noise can be introduced for a variety of reasons, most of which are random or white noise that is evenly distributed. In some cases, the imaging device mechanism or the signal processing algorithm also introduces additional noise components. Due to the specific characteristics of medical images, noise may affect the identification and analysis of diseases, which can lead to serious consequences if accurate clinical decisions are compromised by noise. Consequently, medical image denoising has become a compulsory pre-processing step for subsequent medical image analysis.

The most common medical imaging modalities include Computed Tomography (CT), Magnetic Resonance (MR) imaging, Positron Emission Tomography (PET), and Ultrasound (US). Since most medical image noise is signal-dependent, there are significant differences between medical image denoising methods and natural image denoising techniques. The noise affecting medical images varies across modalities. The following section describes the noise characteristics of different medical imaging modalities. As this thesis focuses on CT and MR imaging, only these two modalities are discussed.

Computed Tomography (CT) images, also known as Computerised Axial Tomography (CAT) scans or X-ray Computed Tomography (X-ray CT). The principle of operation is to use X-rays to penetrate the object from different angles and then reconstruct each cross-sectional image into a three-dimensional image. CT imaging, on the other hand, makes use of ionising radiation, and the radiation dose continues to accumulate over time. LDCT images, which stand for low-dose computed tomography, are produced with

the intention of lessening the effects of ionising radiation [32, 135]. On the other hand, the image quality is positively correlated with the radiation dose, so a high radiation dose is required to obtain a high-quality image, and lowering the radiation dose also reduces the image quality. At the same time, the reduction in image quality leads to an increase in noise. Studies have shown that a twofold reduction in radiation dose results in a $\sqrt{2}$ increase in noise in the image [13]. As the reliability of the disease diagnosis requires the ratio of noise amplitude and the relevant tissue contrast to be sufficiently large, the radiation dose needs to be guaranteed at a certain intensity in order to achieve this requirement. This makes the diagnosis of disease using LDCT images extraordinarily difficult. Low-dose scanning protocols frequently result in an increase in noise and non-stationary streak artifacts, which leads to a reduction of the reconstructed image quality [24]. The primary source of noise in a CT scan is the quantum noise, which is supposed to CT projection noise follows a mixed Poisson-Gaussian distribution [32]. CT image reconstruction involves various post-processing steps, which makes it hard to determine the CT image noise statistics. Based on the Central Limit Theorem (CLT) [135], assuming that the noise in each voxel follows a Gaussian distribution, the noise model used in the image-based denoising methods was simplified [33, 46, 109]. The method of calculating the voxels of a CT image is to add up the different projections, which results in an additive Gaussian statistic for the noise of the CT image [53].

Magnetic resonance (MR) imaging generates the inside of the body image by combing radio waves and magnetic fields, which is superior to other forms of medical imaging modalities in terms of examining the internal body structures. The principle of MR imaging uses the way that billions of hydrogen atoms in the body respond to a magnetic field. All positively charged hydrogen atoms are nearly uniformly aligned in the direction of the magnetic field as it travels through the body. A radio frequency wave pulse is then used to influence the alignment of the protons. The energy will be

released when protons revert to their initial positions. Using the received signal to generate a grey matter diagram, thus creating a cross-sectional image. In general, to model the noise distribution, each quadrature noise is assumed to be an independent, zero-mean white Gaussian noise [55]. In MR images, the noise power is constant for each voxel, so that the noise follows a stationary Rician distribution [3, 48].

Medical image denoising methods can be classified into two categories based on the stage at which denoising is applied: during image acquisition and post-acquisition processing. Acquisition-stage denoising enhances the imaging system with specialised hardware or processing modules to reduce noise, while post-acquisition denoising relies on digital image processing techniques applied to reconstructed images.

Various methods are frequently employed to reduce the noise of medical images, including diffusion filters [167], Gaussian averaging, mean, median [47], and Lee [79]. To denoise particular forms of medical image noise, a variety of filters have been devised with attributes that augment their denoising capability [1, 81, 105]. However, these methods result in the disappearance of low-contrast minor lesions and noise, which has resulted in the inapplicability of these methods to the diagnosis of diseases. Compared to those filters, the performance of non-linear filters in medical image enhancement is superior. The bilateral filter [134] and median filter [89] are two of the most common non-linear filters. Although the computational complexity of these methods is relatively low, they tend to blur the image. Non-local means filters [34, 95] have been shown to have good denoising effects in MR images, but they have a higher computational complexity. Multi-scale analyses use time-frequency analysis to achieve the aim of image denoising, which also has low computational complexity. Considerable effort has been dedicated to the application of multi-scale analysis in image denoising [11, 33, 151]. However, most wavelet thresholding methods are

hampered by the fact that the threshold chosen does not correspond to the signal and noise component distributions at various scales. In order to solve this issue, based on Bayesian theory, Tian *et al.* [133] proposed nonlinear estimators. Although the above problems are solved to some extent, the high computational complexity is still an unavoidable drawback. In addition, some medical image denoising algorithms use soft computing principles, for example, Genetic Algorithms(GA), Artificial Neural Networks(ANN), and Fuzzy Logic(FL) [7]. However, these methods require extensive quantities of logical thought and are extremely complex.

As different imaging modalities exhibit distinct noise distributions, a unified denoising approach for multimodal medical images is difficult to achieve. An effective MR image denoising algorithm aims to reduce noise while preserving spatial resolution and structural details. Maintaining edge integrity and robustness to artifacts is essential for diagnostic applications. Common MRI denoising techniques include low-rank approximation, self-similarity-based methods, filtering approaches, sparsity-based algorithms, and transform-domain techniques [96, 98, 108, 131].

For MR image denoising, Henkelman *et al.* [55] proposed averaging, spatial filtering, and temporal filtering. Due to the spatial filtering that caused image blurring, [3, 48] proposed anisotropic diffusion filtering and its derivatives. Golshan *et al.* [43] found that the problem of denoising could be solved by using the Linear Minimum Mean Square Error (LMMSE) estimation. Later, the non-local LMMSE estimation [130] has been proposed to solve the issue in LMMSE that cannot exploit 3D MR data redundancy. While such spatial domain techniques have had some success, the transform domain methods have also been used for denoising Rician distributed noise. As one of the pioneering methods in transform domain approaches, Coupe *et al.* [28] proposed Wavelet Sub-band Coefficient Mixing (WSM). However, its performance is limited when dealing with smooth, curved edges. Non

Local Means (NLM) filter [15] is a popular method in image denoising that takes into account the inherent redundancy of image patterns. Due to the promising denoising results achieved by the NLM filter, a number of variants of the NLM filter have been proposed based on this approach and have achieved state-of-the-art results [8, 29, 95, 97, 98, 165].

With the rise of GANs, an increasing number of studies have started to adopt it for medical image denoising. Yang *et al.* [159] proposed a GAN-based method with Wasserstein distance and perceptual loss (WGAN), which does not apply pixel-level MSE loss but instead employs VGG to extract features from the image. As it is more in accordance with human visual intuition, WGAN reduces unnecessary smoothness and enhances image quality. Armanious *et al.* [6] introduced MedGAN for the task of medical image translation, which combines the adversarial framework of GAN and a new loss function. MedGAN allows the discriminator to penalise the differences between the translated image and the original image, achieving good results on the task of PET denoising.

In addition, the great success of Transformer in the field of natural language processing has led to its introduction into image processing as well. Although Transformer has achieved great success in natural language processing, the image processing-based Transformer is still in a relatively early stage. There are some studies that have used Transformer for medical image denoising tasks and have also achieved good performance [91, 138]. There are also some Transformer-based methods [19, 40, 139, 161, 168, 174] applied to natural image denoising that may inspire the subsequent research on medical image denoising.

The above review focuses on representative methods and is not intended to provide an exhaustive survey.

## 2.5 Medical Image Cross-Modality Synthesis

Medical image cross-modality synthesis is a technique that algorithmically converts medical images from one imaging modality to another. In medical imaging, different imaging modalities provide different information, and sometimes a patient may have received only one of these imaging examinations. In order to obtain more comprehensive information, cross-modality synthesis techniques can generate missing modality images, which provide clinicians with a more comprehensive assessment of the patient's condition by providing more information. One of the main goals of this technique is to create mappings between different imaging modalities such that the synthetic image visually preserves the structure and characteristics of the original information. Typically, deep learning methods, especially Generative Adversarial Networks (GANs), among others, are widely used for the task of cross-modality synthesis of medical images. These methods are able to learn complex mapping relationships that make the transformation from one modality to another more accurate and realistic.

Application scenarios include converting X-ray images to CT or MRI images, converting CT images to MRI images, and other similar tasks. Cross-modality synthesis of medical images helps physicians obtain consistent information between different imaging modalities, thereby improving diagnostic accuracy and confidence in clinical decision-making.

Consider the history of medical image cross-modality synthesis. From an evolutionary perspective, the development of natural image-to-image translation does indeed provide essential guidelines and insights for medical image cross-modality synthesis. In the background of medical image cross-modality synthesis, dictionary learning [2] plays an important role. Based on dictionary learning, Roy *et al.* [124] proposed a method that trains a source domain dictionary and a target domain dictionary, using the input

image to identify similar patches in the source domain dictionary, then extracts the corresponding patches in the target domain dictionary to synthesise the target domain data. Inspired by this, by imposing a graph Laplacian constraint, Huang *et al*. [60] enhanced the quality of the synthesised image. Isola *et al*. [65] proposed pix2pix, which significantly impacted the field of supervised cross-modality synthesis. Since then, most approaches to supervised cross-modality synthesis have adopted the model associated with pix2pix or its variants. Maspero *et al*. [101] utilised pix2pix to synthesise CT images from MRI. Olut *et al*. use the steerable filter loss on pix2pix to synthesise magnetic resonance angiography from T1-w and T2-w images. In the same way that pix2pix methods have achieved in the field of supervised cross-modality synthesis, CycleGAN [178] has played a crucial role in the field of unsupervised cross-modality synthesis. Hiasa *et al*. [56] introduced a gradient consistency loss in CycleGAN to optimise the edge map of the synthesised image. Zhang *et al*. [172] introduced implicit shape constraints to the process of image translation and utilised two segmentation networks that separate the respective image from semantic labels. Unlike Zhang *et al*. [172]'s method, Chen *et al*. [18] trained the segmentation network offline in advance and fixed the segmentation network during the training of image translation networks.

As stated before, after Transformer was applied to the field of image processing, an increasing number of researchers have used it, including in the medical image synthesis [30, 57, 70, 75, 90, 122, 129, 148]. In future research, we will explore more possibilities by combining the structure of GANs and Transformer.

## 2.6 Summary and Research Gap

This chapter begins by introducing Generative Adversarial Networks (GANs) and some of their variants. It then reviews related methods for medical image denoising and cross-modality synthesis. Based on the review presented in this chapter, it is evident that existing approaches can be broadly categorised into traditional methods and deep learning-based methods. In both medical image denoising and cross-modality synthesis, deep learning-based approaches demonstrate notable advantages over traditional techniques.

Firstly, deep learning-based models are capable of learning complex image features and noise patterns from large-scale datasets, thereby achieving improved performance and generalisation in denoising and cross-modality synthesis tasks. Moreover, these models can learn noise representations and perform noise removal in an end-to-end manner, without the need for manually designed feature extractors, which simplifies the overall processing pipeline. While traditional approaches typically rely on hand-crafted features and predefined models, deep learning-based methods depend less on data preprocessing and feature engineering, enabling more effective utilisation of raw data.

In addition, GAN-based methods are able to generate synthetic images that are more realistic and closer to real medical images. Furthermore, GAN-based approaches can model complex noise distributions and generate denoised images, which is beneficial for handling noise originating from different sources. Importantly, some GAN-based methods are capable of learning from unlabelled data, which is particularly valuable in medical image processing, where acquiring paired datasets is often difficult.

However, despite these advances, the majority of existing methods are designed to address a single task independently. For example, medical im-

age denoising methods typically focus on single-modality noise removal, while cross-modality synthesis methods often assume noise-free training data and neglect noise degradation in the input images. In practical medical image acquisition scenarios, factors such as noise and limited resolution are unavoidable. Consequently, methods that perform denoising or cross-modality synthesis in isolation are based on idealised assumptions.

Moreover, synthesising a noise-free multi-modality medical image usually requires multiple sequential processing steps, which can lead to cumulative image quality degradation. This observation highlights a significant research gap in the simultaneous processing of multiple medical image tasks. We posit that employing an integrated model capable of performing denoising and cross-modality synthesis simultaneously could maximise image realism while minimising quality loss. This study represents an exploration and initial attempt toward addressing this challenge.

Based on the identified research gap, the following chapter focuses on the design and evaluation of an unsupervised framework for low-dose CT image denoising.

# Chapter 3

# Unsupervised Low-Dose Chest CT Image Denoising Bidirectional Adversarial Networks

As motivated by the research gap identified in the previous chapter, medical image denoising is a crucial stage in medical image processing. Noise is an unwanted or arbitrary disturbance in an image that can be caused by a number of factors. In medical images, noise may be introduced due to equipment limitations, radiation, irradiation dose, patient movement, etc. The image noise may affect the accurately diagnose, degrade image quality, and result in inaccurate diagnostic results. Therefore, medical image denoising is crucial for enhancing diagnostic precision and image quality. In this chapter, we create a Denoising Bidirectional Adversarial Network (DeBiGAN), which utilises the innovative design Multi-channel Asymmetric Residual (MAR) block structure that makes up the Multi-channel Asymmetric Residual Generator (MARG) to solve the problem of LDCT image denoising using unsupervised learning and achieve state-of-the-art results.

## 3.1 Introduction

The process of medical image denoising is to remove the noise or undesirable artifacts from the acquired image while retaining the relevant information. Computerised Tomography (CT) is one of the medical image modality that can produce detailed images of many structures inside the body. However, the obtained CT images are often degraded by noise, which can compromise the accuracy of diagnosis and treatment. The traditional methods used for medical image denoising are previously introduced, which include the Gaussian filter, the median filter, and others. There are more advanced methods, such as wavelet-based denoising, non-local means denoising, and deep learning-based denoising. Wavelet-based denoising involves decomposing the image into different frequency bands and filtering out the noise in each band separately. Non-local means denoising methods utilise the redundant information of an image to remove the noise while maintaining the fundamental structure. Deep learning-based denoising trains neural networks to learn the noise feature from a large, noisy, and clear medical image dataset. As an essential step in medical image processing, medical image denoising can improve the accuracy and reliability of medical diagnoses and treatments.

As was previously stated, CT plays a crucial role in disease diagnosis, but the wide use of CT is increasing public concerns about its safety, as the X-ray radiation can cause irreparable harm to humans and may cause cancer. Over the past few decades, decreasing CT radiation has been widely acknowledged in CT-related research. However, as the radiation dose is reduced, noise and artifacts are introduced to the reconstructed images, which compromises the ensuing diagnostic and other tasks significantly. Removing the noise from the LDCT image is the straightforward solution to the issue. Consequently, extensive research is dedicated to LDCT denoising al-

gorithms.

The denoising algorithms for LDCT can be categorised into three groups, which are sinogram filtration, iterative reconstruction, and image post-processing. In contrast to routine CT, the LDCT scanner produces noisy sinogram data. Sinogram filtration-based approaches [94, 141, 142] apply denoising prior to image reconstruction. Iterative reconstruction approaches utilise the prior information of the image domain [117, 153], for example, total variation [175] and dictionary learning [155]. In contrast to the prior two categories, the post-processing image approach worked on the publicly available images, which removed patient confidentiality. Non-local means [92] and block-matching 3-D [41] are two traditional methods of image post-processing. These methods resulted in the loss of critical structural details and over-smoothed the denoised LDCT image. Numerous medical applications have been advanced by the accelerated development of deep learning techniques. The LDCT image denoising also obtained impressive results by deep learning techniques [20, 21, 83, 125, 126, 140, 150, 152, 159]. In the deep learning-based denoising method, loss function and network structure are two essential components. In terms of loss function, MSE is the simplest one; however, empirical evidence suggests a negative correlation with human perception of image quality [45, 147]. Therefore, alternative loss functions have been explored in the context of LDCT image denoising, which include $l_1$ loss, perceptual loss, adversarial loss, or mixed loss functions, among others.

The reason that GANs' success is due to adversarial loss forces the generated image distribution to fit the real image distribution, which is notably effective for image generation tasks. On the other hand, adversarial loss has been demonstrated to be the most effective due to its ability to dynamically assess the similarity of denoised and normal-dose images throughout the training phase. In this study, we adopt adversarial loss to force the denoised

image distribution to fit the normal-dose image distribution.

Due to the difficulty and cost of acquiring paired medical images, in order to solve the problem in practice, this method uses unsupervised learning. Thus, pairs of data are not required for the training process. Unsupervised learning is a viable option when acquiring paired data is challenging or costly. In the context of LDCT image denoising, acquiring large quantities of paired data is difficult, so the unsupervised learning method is advantageous. Using unsupervised learning, the model can directly discover patterns and representations from unpaired data. Inspired by Cyclegan [178], we designed a network architecture with dual generators and dual discriminators to implement unsupervised learning without paired data. By employing dual generators, a greater diversity of image transformations and operations can be accomplished. Each generator is capable of conducting a unique transformation or producing a distinct output. This design makes the model more adaptable and capable of performing a broader variety of tasks. Dual discriminators can provide additional feedback signals and loss functions to guide the model's training. Each discriminator can concentrate on a unique feature or aspect and provide more feedback, thereby facilitating model learning and enhancement.

In the proposed method, Unsupervised Low-Dose Chest CT Image Denoising Bidirectional Adversarial Network (DeBiGAN), we created a new generator architecture: Multi-channel Asymmetric Residual Generator (MARG), which is demonstrated in Fig 3.2. This is an innovative generator structure that introduces multi-channel asymmetric residual (MAR) block, thus providing better flexibility and expressiveness. This design uses different residual modules in the hierarchical links of the generator to increase the variability and diversity of the network. The structure of the MAR is shown in Fig 3.3. The original residual networks address gradient disappearance and model degradation by incorporating shortcuts and residual blocks, allowing

for the training and optimisation of deeper networks. ResNet performs well on a variety of tasks, but it has a number of drawbacks. Due to the fact that the structure of ResNet is relatively complex and contains a large number of skip connections and residual blocks, the network is difficult to interpret. Understanding and interpreting the network's decision-making process can become challenging at deeper network levels. The depth and complexity of ResNet can result in longer training durations and greater resource consumption. In certain circumstances, the use of deeper ResNet structures is required to obtain better performance, which may increase the training and inference time and computational cost. Deeper ResNet models may be susceptible to overfitting when using smaller datasets or fewer training samples. This is due to the fact that deeper networks have more parameters and complexity, necessitating more training data to avoid overfitting issues. The depth and width parameters in ResNet must be meticulously tailored to achieve optimal performance. Choosing the appropriate depth and width can necessitate extensive experimentation and hyper-parameter optimisation, thereby compounding the difficulty of model design and training.

The proposed Multi-channel Asymmetric Residual (MAR) structure solves some of the above problems that existed in the original ResNet. The network structure is robust and generalisable even with small training data sets, helping to avoid overfitting problems. The use of a multi-channel asymmetric structure can increase the network's feature extraction process's adaptability. In a conventional symmetric structure, the feature extraction procedure is identical for each channel, and each channel has the same weight parameters. In contrast, in an asymmetric structure, each channel can have distinct weight parameters, allowing different channels to extract features from input data in a differentiated manner. By utilising a multi-channel asymmetric structure, the network can learn the weight assignment between channels based on the characteristics of the input data and the correlation between channels. This adaptability can assist the network in cap-

turing the diversity and complexity of the input data and extracting more robust and distinct features. In addition, the asymmetric structure permits varying levels of feature weighting for various channels, enabling the network to place a greater emphasis on extracting the most pertinent and informative features from the input data. By adaptively modifying the weights of the various channels, the network can make better use of the information in the input data and increase the discriminative and generalising power of the features. In conclusion, the adaptability of the multi-channel asymmetric structure for feature extraction enables the network to better accommodate the characteristics and complexity of the input data. This flexibility contributes to the improvement of network performance and provides greater versatility and expressiveness for a variety of tasks.

## 3.2 Method

To formally describe the proposed unsupervised LDCT denoising framework, this section first introduces the underlying GAN formulation and relevant preliminaries.

### 3.2.1 Preliminaries

As previously mentioned, a generator and a discriminator comprise the Generative Adversarial Networks (GANs). The objective function of GANs is:

$$\min_G \max_D V(D,G) = \mathbb{E}_{x\sim p_{\text{data}}(x)}[\log D(x)] + \mathbb{E}_{z\sim p_z(z)}[\log(1-D(G(z))], \quad (3.1)$$

where $G$ is the generator, $D$ is the discriminator, $x$ is the vectorised image, $p_{\text{data}}$ is the real data distribution, $z$ is the latent random vector, and $p_z$ is the

uniform noise distribution.

The generator $G$ is trained to mimic the actual image via mapping $z$ sampled from $p_z$ to $p_{\text{data}}$. The discriminator $D$ is optimised to distinguish if the image is from generated $G(z) \sim P_z$ or from the real $x \sim P_{data}$. Specifically, provided the vectorised image $x$, the generator $G$ and discriminator $D$ can be trained alternatively. That is, fix $G$'s parameters, optimise $D$. Then fix $D$'s parameters and optimise $G$. The global optimality is $p_g = p_{data}$. When $G$ and $D$ have been well trained, they will have the ability to make $p_g$ converge to $p_{data}$, which means the global optimality has been achieved. As mentioned in the previous chapter, there are numerous variants of GAN that improve its stability while adding different attributes. For example, Wasserstein GAN [5], WGAN with gradient penalty (WGAN-GP) [49], and least-squares GANs [99] have been developed to improve the stability of GANs' training. UNIT [86], DiscoGAN [71], DualGAN [160], and CycleGAN [178] have dual generators and dual discriminators, which utilise the cycle-consistency property to achieve unsupervised image translation.

From the perspective of LDCT image denoising, the purpose of the generator is to produce photorealistic denoised images. The discriminator attempts to differentiate between denoised images and real normal-dose CT images.

### 3.2.2 Problem Formulation

Building upon the GAN framework introduced above, this section formulates the unsupervised low-dose CT image denoising problem addressed in this study.

In this chapter, we propose an unsupervised method for medical image denoising. The goal is to perform the Low-Dose CT image denoising

without supervision of paired images. The source domain is the LDCT images, which can be expressed as $\mathscr{X} = \{\boldsymbol{X}_n\}_{n=1}^{S} \in \mathbb{R}^{i \times j \times k \times S}$. The target domain is the denoised Normal-Dose CT (NDCT) images, which is $\mathscr{Y} = \{\boldsymbol{Y}_n\}_{n=1}^{T} \in \mathbb{R}^{i \times j \times k \times T}$. Where $\boldsymbol{X}$ and $\boldsymbol{Y}$ are the input images, $i$ and $j$ mean the image dimension of axial view, $k$ is the number of the image stacks alone on the z-axis, $S$ and $T$ indicate the numbers of subjects in the source domain and the target domain. Similar to the current dual GANs' learning, with two generators and two discriminators, the first generator, $G$, is used to map from the source domain to the target domain, and the other generator, $F$, is used to map from the target domain to the source domain. In contrast to the two generators, two discriminators, $D_G$ and $D_F$ are constructed to distinguish the generated images from $G$ and $F$ respectively. Therefore, the process of these two mappings is formulated as: $\mathscr{X} \rightarrow \mathscr{Y} : \hat{\boldsymbol{Y}} = G(\boldsymbol{X})$ and $\mathscr{Y} \rightarrow \mathscr{X} : \hat{\boldsymbol{X}} = F(\boldsymbol{Y})$. Fig 3.1 demonstrated the proposed method, which will be introduced in detail in the experiments section.

### 3.2.3 Loss Function

Following the problem formulation, this section defines the loss functions used to optimise the proposed unsupervised LDCT denoising model.

To remove the noise of an image $\boldsymbol{X}_n$ in $\mathscr{X}$, the function $G : \mathscr{X} \rightarrow \mathscr{Y}$ has been trained with the expected output $\hat{\boldsymbol{Y}}_n = G(\boldsymbol{X}_n)$. Then the discriminator $D_G$ calculated the likelihood of the image $\boldsymbol{X}_n$ that has been sampled from the target domain $\mathscr{Y}$. Conversely, the generator $F$ is trained to map an image $\boldsymbol{Y}_n$ in the domain $\mathscr{Y}$ to an image in the domain $\mathscr{X}$. The discriminator $D_F$ gives the corresponding possibility that the output image $\hat{\boldsymbol{X}}_n = F(\boldsymbol{Y}_n)$ is from the source domain $\mathscr{X}$. Based on the process described above, combined with the original GAN's adversarial losses [45], the two mapping functions can be combined and expressed as:

$$\begin{aligned}\mathcal{L}_{ba}(G,F,D_G,D_F) =& \mathbb{E}_{\boldsymbol{X}\sim p_{\text{data}}(\boldsymbol{X})}[\log(1-D_G(G(\boldsymbol{X})))] + \mathbb{E}_{\boldsymbol{Y}\sim p_{\text{data}}(\boldsymbol{Y})}[\log D_G(\boldsymbol{Y})] + \\ & \mathbb{E}_{\boldsymbol{Y}\sim p_{\text{data}}(\boldsymbol{Y})}[\log(1-D_F(F(\boldsymbol{Y})))] + \mathbb{E}_{\boldsymbol{X}\sim p_{\text{data}}(\boldsymbol{X})}[\log D_F(\boldsymbol{X})],\end{aligned} \tag{3.2}$$

where $\mathcal{L}_{ba}$ is the bi-adversarial loss. Eq.(3.2) establishes a closed loop between two adversarial losses, thereby extending GANs into a bidirectional learning process. A typical characteristic of unsupervised bidirectional learning is forcing both learning processes to generate fake input from each other. To express this process in terms of the formula is: generate $\boldsymbol{X}_c$ for the process of $\mathscr{X} \rightarrow \mathscr{Y}$, and generate $\boldsymbol{Y}_c$ for the process of $\mathscr{Y} \rightarrow \mathscr{X}$. Therefore, $\boldsymbol{X}_c = F(\hat{\boldsymbol{Y}}) = F(G(\boldsymbol{X}))$, and $\boldsymbol{Y}_c = F(\hat{\boldsymbol{X}}) = G(F(\boldsymbol{Y}))$. Inspired by some contemporary works [71, 86, 160, 178], in order to regularise the two mappings, the cycle-consistency constraint (Eq. 3.3) has been utilised.

$$\mathcal{L}_c = \mathbb{E}_{\boldsymbol{Y}\sim p_{\text{data}}(\boldsymbol{Y})} \| \boldsymbol{Y} - F(G(\boldsymbol{Y})) \|_1 + \mathbb{E}_{\boldsymbol{X}\sim p_{\text{data}}(\boldsymbol{X})} \| \boldsymbol{X} - F(G(\boldsymbol{X})) \|_1 . \tag{3.3}$$

Eq. 3.3 utilise the $l_1$ distance to measure the difference between the input data and the reconstructed data. Some research [160, 178] show that the $l_1$ loss has can avoid blurriness.

The structural information of medical images is crucial, as it typically contains organisational information, which is essential for medical diagnosis and analysis. In order to keep the denoised image as realistic as possible in terms of structural information, we add a constraint to reduce the structural changes in the pre- and post-denoised images.

$$\mathcal{L}_s = \mathbb{E}_{\boldsymbol{X}\sim p_{\text{data}}(\boldsymbol{X})} \| \boldsymbol{X} - G(\boldsymbol{X}) \|_2^2 + \mathbb{E}_{\boldsymbol{Y}\sim p_{\text{data}}(\boldsymbol{Y})} \| \boldsymbol{Y} - F(\boldsymbol{Y}) \|_2^2 . \tag{3.4}$$

It is to ensure that the generated images retain the anatomical structure information while still meeting the optimisation requirements of other loss functions to obtain more medically interpretable and quality results in medical image denoising.

Correspondingly, the full objective of the proposed model can be modified as follows:

$$\mathcal{L}(\mathscr{X},\mathscr{Y}) = \lambda_{ba}\mathcal{L}_{ba}(G,F,D_G,D_F) + \lambda_c\mathcal{L}_c + \lambda_s\mathcal{L}_s, \tag{3.5}$$

where $\lambda_{ba}, \lambda_c$, and $\lambda_s$ determine the relative weight of the three objectives separately.

## 3.3 Experiments

To evaluate the proposed method, this section presents the experimental setup and network architecture used to assess the unsupervised LDCT denoising model.

### 3.3.1 Network Structures

The proposed method (DeBiGAN) is constructed by two generators and two discriminators, as shown in Fig 3.1, where the input LDCT image has been generated into an NDCT image through the MARG($G$), and the MARG($F$) is used to generate the reconstructed LDCT image. The discriminator $D_G$ is constructed to distinguish the input LDCT image from the reconstructed LDCT image. The discriminator $D_F$ is created to distinguish between the original NDCT image and the generated NDCT image.

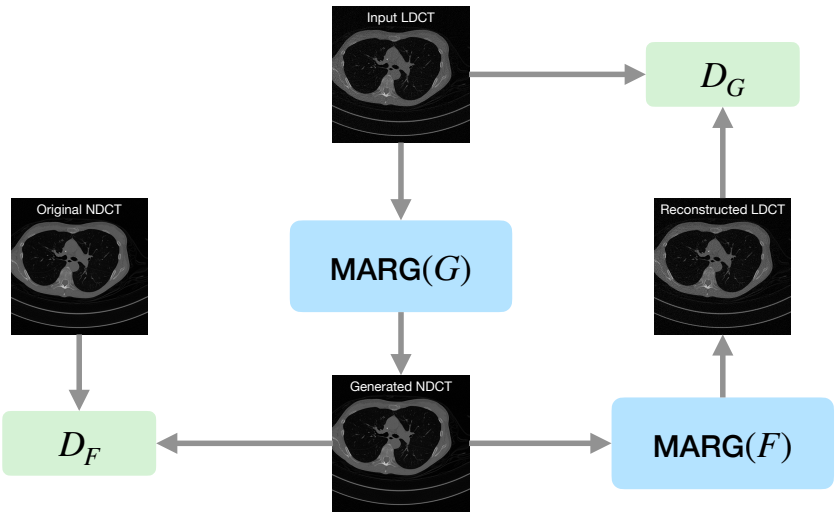


**Figure 3.1: The network architecture of the proposed DeBiGAN**

The structure of innovative MARG is shown in Fig 3.2. The initial block of MARG is composed of 3 convolution layers; the strides are 1, 2, and 2. The first convolution layer has a $7 \times 7$ kernels, which provides a larger field of perception, meaning that each output pixel point can perceive a larger portion of the input image. This allows the network to better comprehend the global structure and semantics of the input data by capturing features and contextual information on a larger scale. A larger convolution kernel provided a greater representational capacity, which was capable of learning more complex and detailed feature representations, thereby enhancing the network's ability to discriminate and generalise input data. However, larger convolutional kernels increase computational effort. This may increase training and inference costs in terms of time and resources. Using larger convolution kernels may result in a loss of local detail information. A larger perceptual field may obscure or smooth out local image details, thereby diminishing the perception of fine-grained characteristics. Therefore, the rest of the convolution layers use $3 \times 3$ kernels, which can considerably reduce the number of network parameters when compared

to larger convolutional kernels. This is due to the fact that a $3 \times 3$ kernel only has nine weighting parameters, whereas a larger convolutional kernel requires more parameters. Reducing the number of parameters aids in reducing the computational complexity and memory usage of the model, as well as preventing overfitting. Multiple $3 \times 3$ convolutional layers stacked together can attain a perceptual field size comparable to that of a larger convolutional kernel. By layering multiple $3 \times 3$ convolutional layers, each layer can learn a distinct feature representation scale. This design enhances the network's representational capacity, allowing it to better capture the multi-scale characteristics and intricate patterns of the input data. The $3 \times 3$ convolutional kernel incorporates additional non-linear operations, thereby enhancing the network's ability to model input data. Multiple $3 \times 3$ convolutional layers applied successively can form more complex non-linear transformations, thereby enhancing the expressiveness and discriminative ability of the network. The translational isotropy of the input data is preserved when a $3 \times 3$ convolutional kernel is utilised. This indicates that the $3 \times 3$ convolutional kernel can learn and detect features at various locations, regardless of where they appear in the input image. This enables the network to acquire representations of features with greater translational invariance, thereby enhancing generalisation to the input data. Batch normalisations and ReLU nonlinearity are added in between the convolutional layers, which form Convolution-Batch Normalisations-ReLU. After that, 4 MAR blocks (Fig3.3) connected linearly. In each of the MAR blocks, the convolutional layer filters have been fixed at 256, and the kernel size is $3 \times 3$. Following the 4 MAR blocks, there are two transposed-convolutional layers with 128 and 64 filters, respectively. Between these two transposed-convolutional layers, the batch normalisations and ReLU are inserted, with a structure of TransposedConv-BatchNorm-ReLU. This is followed by a convolutional layer with a filter size of 1 and a stride of 1. Finally, a tanh layer is connected to the output. The generator structure is demonstrated in Fig

3.2.

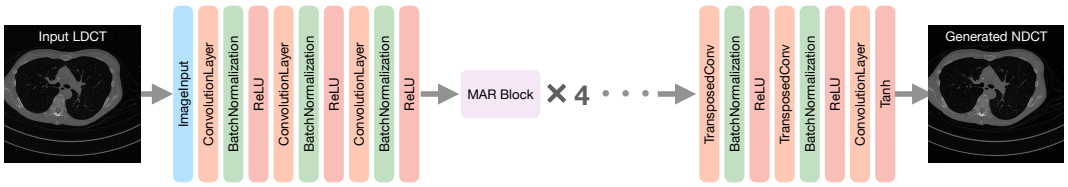


**Figure 3.2: Generator Structure of DeBiGAN**

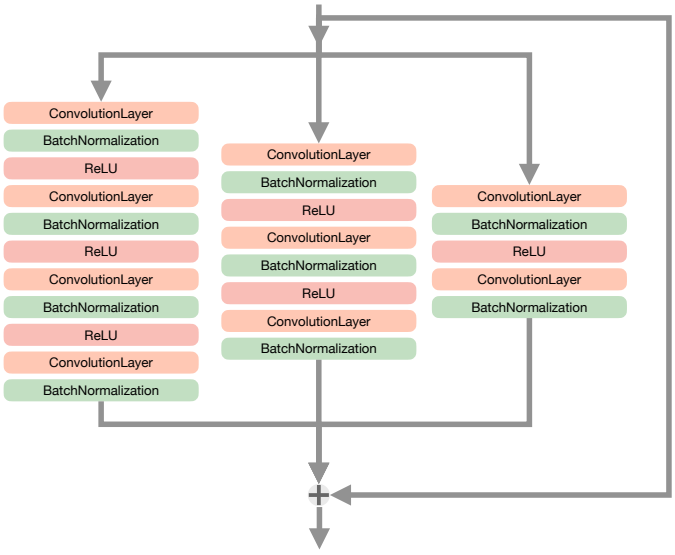


**Figure 3.3: MAR block structure**

For the discriminator in this work, we adopt the Markovian PatchGAN previously investigated in [65, 80, 160], which is a convolutional neural net-

work structure for image generation and image processing tasks, particularly for image evaluation and discrimination. PatchGAN utilised a local perceptual field strategy in the discriminator of the GAN to evaluate the credibility of the generated image by discriminating local image regions, which is based on the principle of dividing the input image into multiple overlapping segments and then classifying each patch independently. Typically, these sections are squares and can overlap to enhance the perception of image detail. By discriminating between local regions of an image, PatchGAN is able to assess the veracity of an image in greater detail. It is capable of capturing the local details and texture information of an image and discriminating the generated image block by block for various regions, thereby providing more precise feedback. Compared to global discrimination of the entire image, PatchGAN's local discrimination strategy reduces the network's computational complexity. By discriminating only small pieces, the network's number of parameters and computational effort are comparatively small, thereby improving training and inference efficiency. The patch size is overlappingly set to $70 \times 70 \times 70$, and the discriminator is trained using a stack of Convlution-BatchNorm-Leaky ReLU layers. Similar to [54, 65, 160], the discriminator is executed convolutionally throughout all pixels, and the final results are obtained by averaging all responses.

### 3.3.2 Experimental Setup

In accordance with the described network architecture, this subsection details the datasets, training configuration, and evaluation protocols used in the experiments.

### Datasets

Low-dose CT image and projection dataset [102] is a public dataset from The Cancer Imaging Archive (TCIA) [107]. The dataset consists of 299 clinically performed patient CT scans of three types: noncontrast head CT scans, low-dose noncontrast chest scans, and contrast-enhanced abdomen CT scans. The size of the dataset is 1.32 $TB$, which contains 299 subjects, 13,009,241 files, and 3 clinical reports. The projection data is stored in DICOM-CT-PD format, and the image data has been stored in DICOM format. The head CT scans and the abdomen CT scans are provided at 25% of the full dose, and the chest CT scans are provided at 10% of the full dose. The image size of the dataset is $512 \times 512$.

### Training Details

The proposed model is trained to update the generators $G, F$ and discriminators $D_G, D_F$ in alternating phases for each batch. The initial learning rate is set to $2 \times 10^{-4}$, utilise the stochastic gradient descent. The batch size is 16. The Adam solver [72] is used for optimisation. The gradient decay factor is set to 0.5 and the squared gradient decay factor is set to 0.999. Data augmentation is performed by rotation, stretching and reflection the training data, eight $128 \times 128$ patches are randomly selected as training data in each training image. Shuffle the training data before each training epoch, and shuffle the validation data before each validation. To balance the influence between objectives, we set $\lambda_{ba} = 1, \lambda_c = 10$, and $\lambda_s = 0.5$. Perform 100 Epochs of training to end up with a trained model.

**Evaluation Metrics**

Peak Signal-to-Noise Ratio (PSNR), Structural Similarity (SSIM), and Multi-Scale Structural Similarity (MSSSIM) are adopted as the evaluation criteria. PSNR is a widely employed metric used to evaluate the quality of videos and images. It calculates the extent of disparity between the original signal and the signal that has undergone compression, encoding, or transmission. PSNR is frequently employed to evaluate the degree of distortion in an image or video and to assess the performance of digital image compression algorithms. As the PSNR value increases, the disparity between the processed and unprocessed signals diminishes, resulting in a reduction in quality degradation. However, PSNR fails to consistently reflect the perception of image or video quality as perceived by the human eye. Sometimes an image with a high PSNR value may not actually be superior due to the possibility that the human visual system's perception of image structure and details is disregarded. Therefore, a combination of other evaluation metrics is necessary. SSIM is an image quality evaluation metric that calculates the degree of similarity between two images by considering their structure, contrast, and luminance. Drawing inspiration from the mechanisms of the human visual system, SSIM endeavours to emulate the way in which the human eye interprets an image. The SSIM value range is from 0 to 1, with 1 indicating that the two images are identical. SSIM is a relatively straightforward image similarity measurement that is commonly used in the fields of image processing, compression, and reconstruction. However, SSIM is not always accurate in its assessment of the quality of images, especially when dealing with images that are subject to noise, compression, or distortion. MSSSIM is a metric for assessing image quality that extends and improves upon SSIM. MSSSIM aims to more accurately measure the similarity between images by introducing a multi-scale analysis that takes into account information about the structure of an image at different scales, which is

done by decomposing the image and comparing its respective brightness, contrast, and structure at multiple scales. These values are then weighted and averaged to arrive at a final similarity metric. MSSSIM is able to capture the structural similarity between images more comprehensively because it considers not only the overall image features but also the detailed information of the image at different scales. This allows MSSSIM to provide more accurate results in evaluating distortion, compression, and noise images.

### Qualitative Evaluations

To demonstrate the effectiveness of the proposed approach in producing photorealistic denoised outcomes that retain accurate details, we compare it with five baseline methods, which are RED-CNN [20], WGAN-VGG [159], CPCE-2D [126], CNCL [42], and DU-GAN [63]. The results are shown in Fig 3.4.

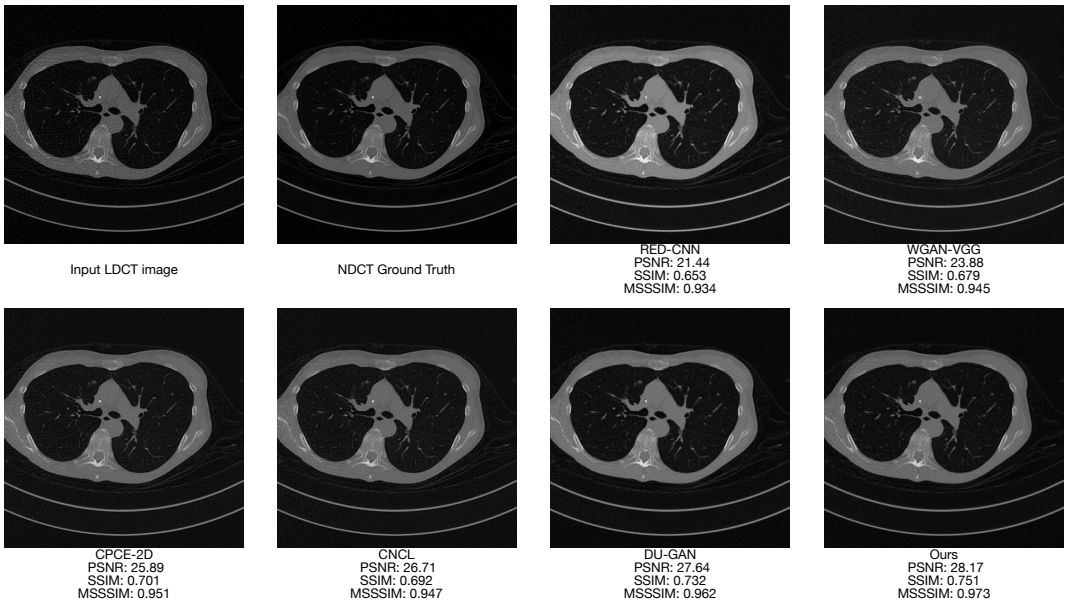


**Figure 3.4: Results of the comparison with the five baseline methods**

### Quantitative Evaluations

For quantitative evaluation, we use PSNR, SSIM, and MSSSIM as the evaluation criteria. The comparison results are shown in Table 3.1. As can be seen from this table, our method achieved state-of-the-art results. To visualise the comparison more intuitively, we made box-plot for each of the three metrics. Fig 3.5 shows the PSNR box-plot, Fig 3.6 illustrates the box-plot of SSIM, and Fig 3.7 presents the SSIM box-plot results. It is clear from these figures that our method has the superior performance compare to the baseline methods. The average PNSR of our method reaches 28.17, and even the closest DU-GAN still has a gap of 0.53. At the same time SSIM and MSSSIM have the highest mean values, reaching 0.751 and 0.973, respectively. Combining the three figures, the proposed method is outperformance than RED-CNN, WGAN-VGG, CPCE-2D, and CNCL, but there are some overlapping regions in the values of these evaluation metrics with the DU-GAN method. It is mainly because of the presence of different images on the test dataset, some of which have small regions of interest, in which case higher evaluation metrics value are obtained, so there are cases where the highest point of the DU-GAN exceeds the average of our method, but as a more comparative metric, its average is a better demonstration of the model's performance.

LDCT Chest Denoising

| Metric (avg.) | RED-CNN | WGAN-VGG | CPCE-2D | CNCL | DU-GAN | Ours |
|---|---|---|---|---|---|---|
| PSNR (dB) | 21.44 | 23.88 | 25.89 | 26.71 | 27.64 | **28.17** |
| SSIM | 0.653 | 0.679 | 0.701 | 0.692 | 0.732 | **0.751** |
| MSSSIM | 0.934 | 0.945 | 0.951 | 0.947 | 0.962 | **0.973** |

**Table 3.1: Quantitative comparisons of different methods on the testing dataset. The best results are marked in bold.**

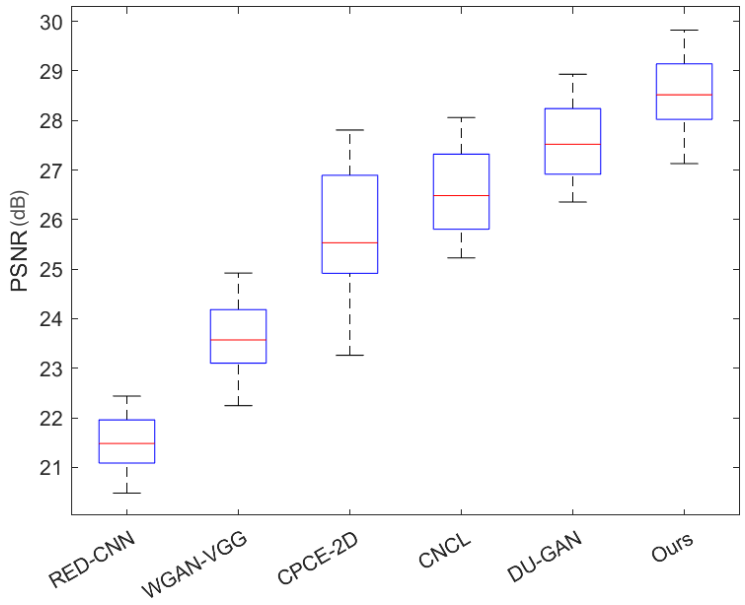


**Figure 3.5: Visual presentation of PSNR for the baseline methods and our method.**

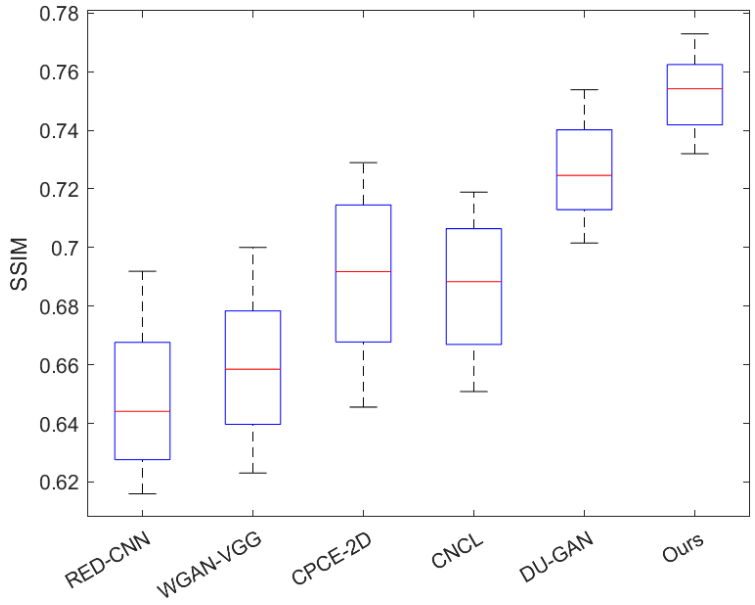


**Figure 3.6: Visual presentation of SSIM for the baseline methods and our method.**

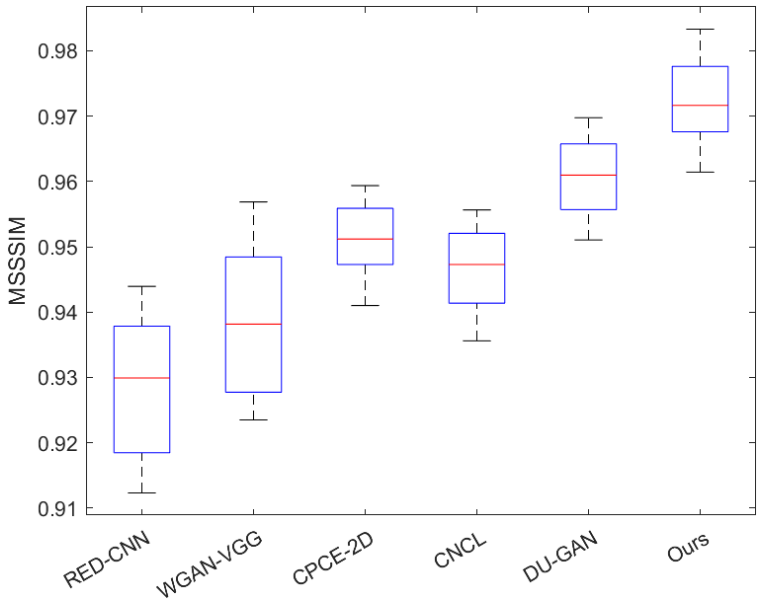


**Figure 3.7: Visual presentation of MSSSIM for the baseline methods and our method.**

### Ablation Studies

To further analyse the contribution of individual components in the proposed model, ablation studies are conducted in this subsection.

The proposed method's ablation study has been conducted to demonstrate its superior performance. The proposed method incorporates multiple components; the ablation study will be conducted independently from various vantage points. Since our method creates an innovative MARG generator structure, we began with this structure and performed ablation studies to demonstrate the effectiveness of the structure compared to the common residual network structure, and a comparison is made with the reduced-channel MARG to demonstrate the necessity and effectiveness of a multi-channel structure. In details, the MARG structure will be replaced by a normal residual network, middle-channel-removed MAR block (Fig 3.8), right-channel-removed MAR block (Fig 3.9), respectively. In order to make

fair comparisons, we maintained the consistency of the experimental setup. In Fig 3.10, the results of the comparison between the common residual network and the MARG structure with some of its channels removed are presented. The result demonstrated the superiority of the MARG structure in comparison to ordinary residual networks, and it also proved the necessity of the multi-channel design.

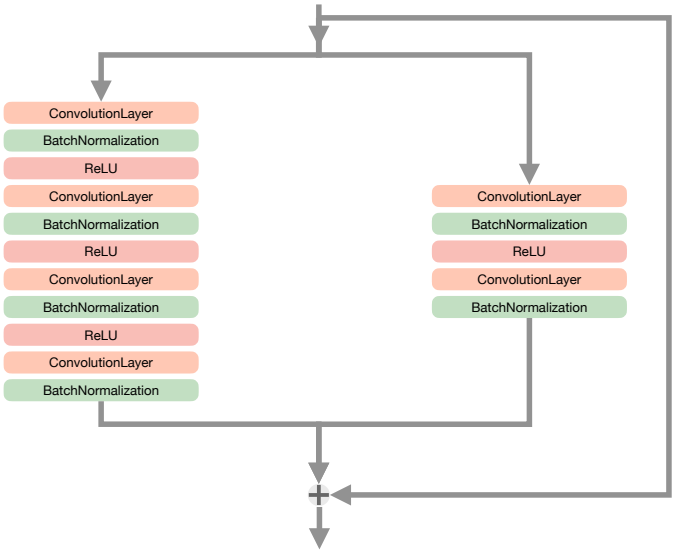


**Figure 3.8: The MARG structure with middle channel removed**

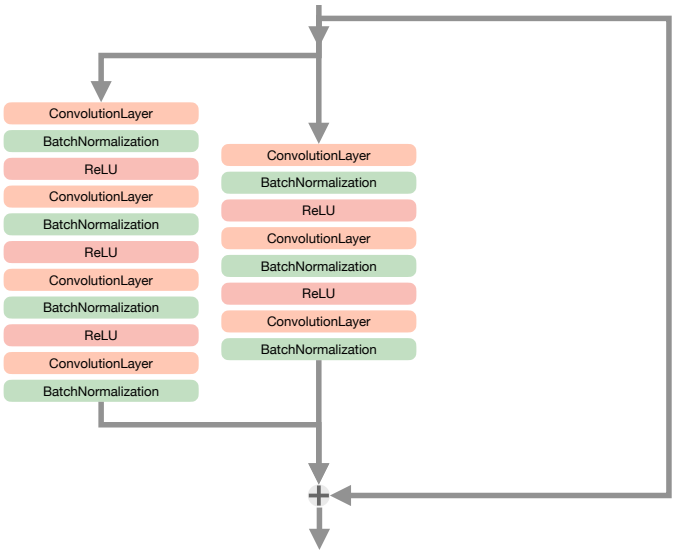


**Figure 3.9: The MARG structure with right channel removed**

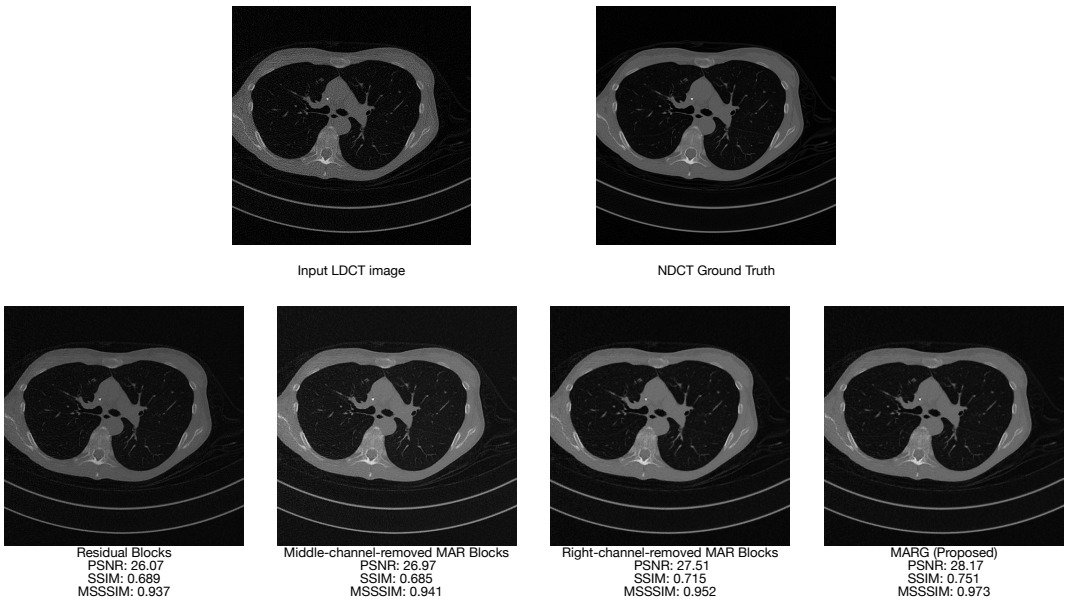


**Figure 3.10: Ablation study between the common residual network and the MARG structure with some of its channels removed comparison results.**

The effectiveness of the asymmetric structures in MARG will be demonstrated next. We compared all asymmetric channel network structures with

the same symmetric structure in its entirety. In the MARG structure, we employ distinct numbers of convolutional blocks for various channels. namely 4, 3, and 2 convolutional blocks. The validity of this asymmetric structure will be demonstrated by 3 different scenarios: First, replacing the 3 convolutional blocks channel and 2 convolutional blocks channel by 4 convolutional channels (Fig 3.11); Second, replacing the 4 convolutional blocks channel and 2 convolutional blocks channel by 3 convolutional channels (Fig 3.12); Third, replacing the 4 convolutional blocks channel and 3 convolutional blocks channel by 2 convolutional channels (Fig 3.13). Fig 3.14 illustrates the results of the comparison with the three replacement structures described above, which demonstrate the superiority of our proposed MARG structure over symmetric multi-channel residual networks.

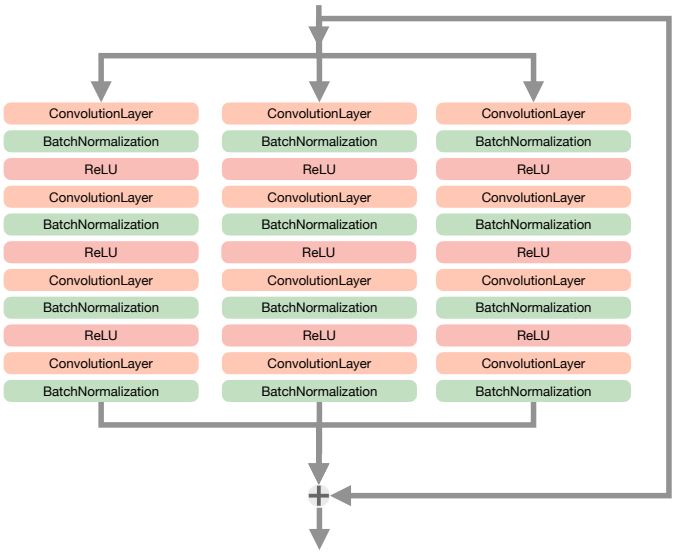


**Figure 3.11: The structure of 4 Convolutional Blocks Symmetric Residual Blocks**

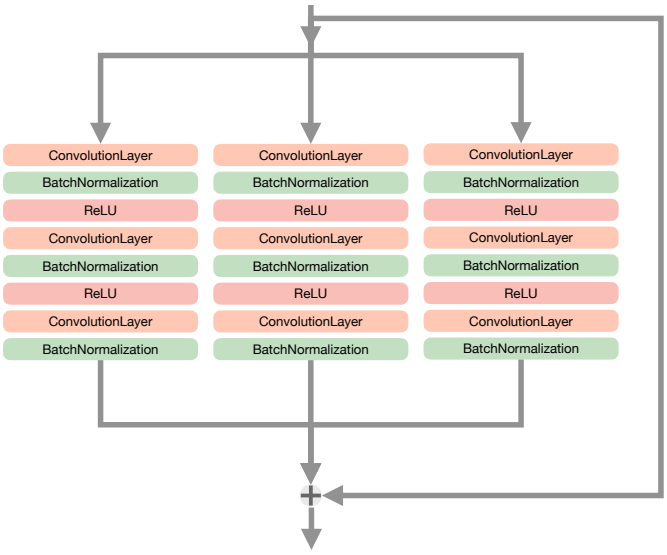


**Figure 3.12: The structure of 3 Convolutional Blocks Symmetric Residual Blocks**

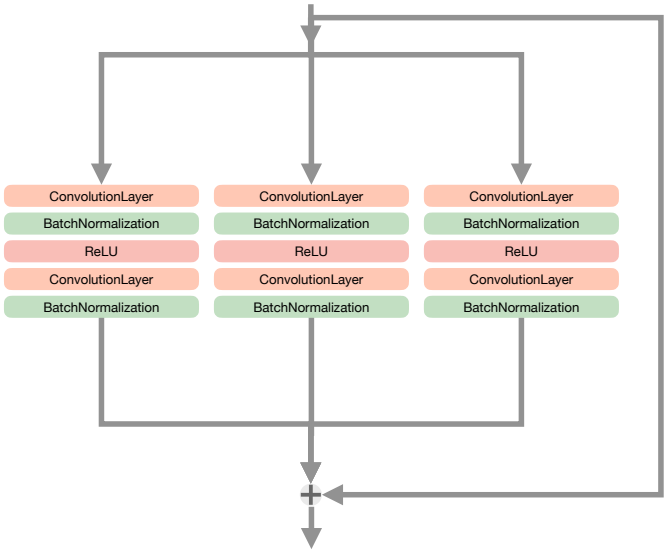


**Figure 3.13: The structure of 2 Convolutional Blocks Symmetric Residual Blocks**

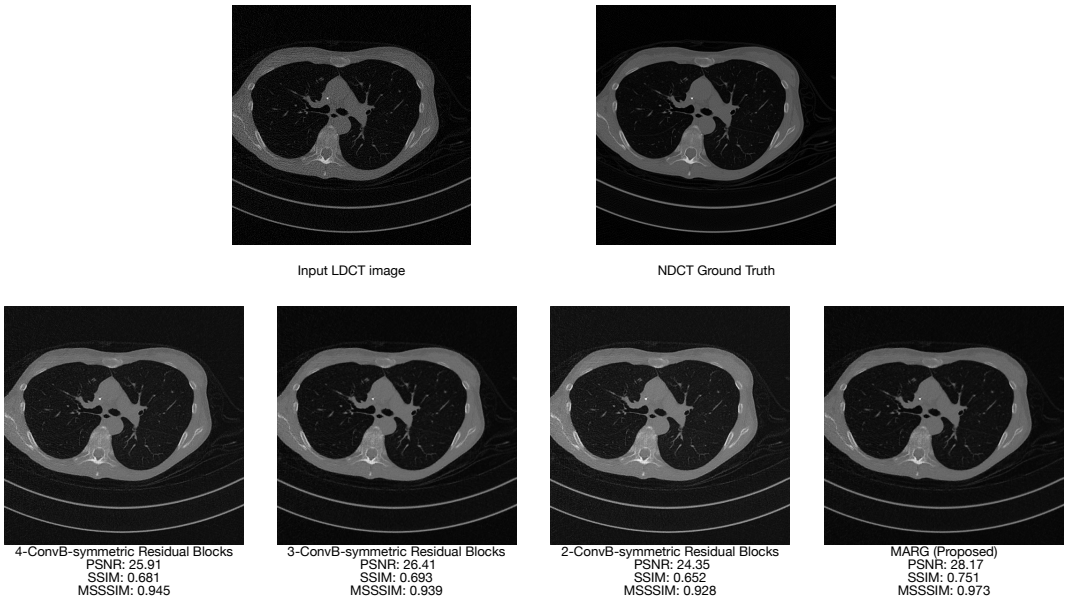


**Figure 3.14: Ablation study between the proposed asymmetric structures in MARG and the symmetric structures with different numbers of convolutional blocks.**

Comparative experiments were also done for the selection of the number of MAR blocks. In our proposed method for LDCT chest image denoising, we discovered that using 4 consecutive MAR blocks gives the best results. Fig 3.15 shows the effect of using 2, 3, and 6 MAR blocks, respectively, on the results. However, this does not mean that using 4 MAR blocks is the optimal solution for all tasks, and it needs to be considered in the context of different actual tasks as well as other parameters.

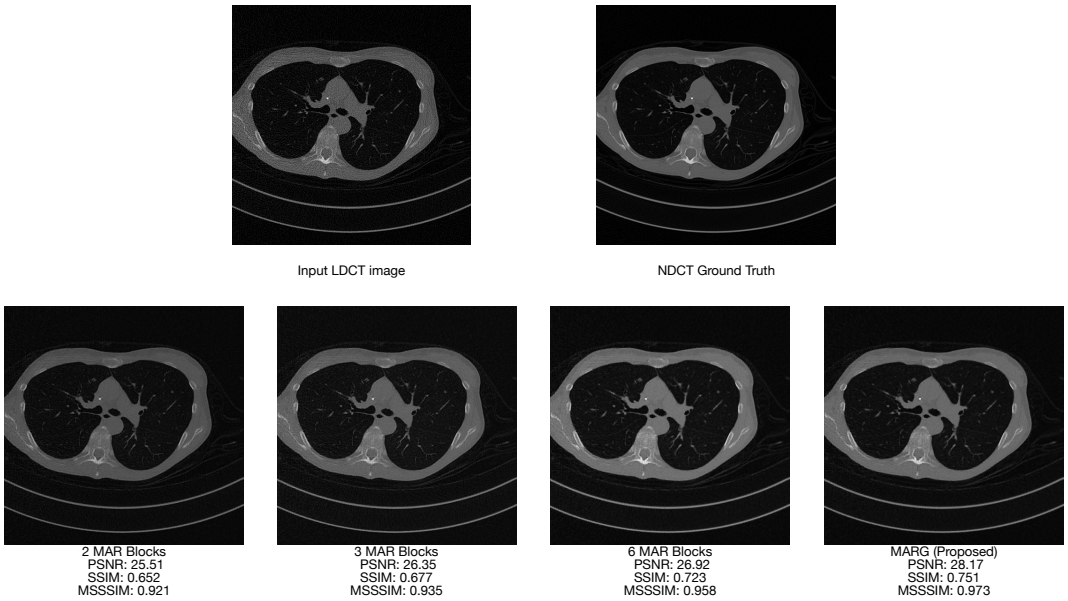


**Figure 3.15: Ablation study of the effect between different number of MAR blocks.**

Table 3.2 summarises the evaluation results of the ablation studies conducted above. To visualise the comparison, the results of their PSNR, SSIM, and MSSSIM are shown in Fig 3.16, Fig 3.17, and Fig 3.18, respectively.

| Ablation Study | PSNR (dB) | SSIM | MSSSIM |
|---|---|---|---|
| Residual Blocks | 26.07 | 0.689 | 0.937 |
| Middle-channel-removed MAR Blocks | 26.97 | 0.685 | 0.941 |
| Right-channel-removed MAR Blocks | 27.51 | 0.715 | 0.952 |
| 4-ConvB-symmetric Residual Blocks | 25.91 | 0.681 | 0.945 |
| 3-ConvB-symmetric Residual Blocks | 26.41 | 0.693 | 0.939 |
| 2-ConvB-symmetric Residual Blocks | 24.35 | 0.652 | 0.928 |
| 2 MAR Blocks | 25.51 | 0.652 | 0.921 |
| 3 MAR Blocks | 26.35 | 0.677 | 0.935 |
| 6 MAR Blocks | 26.92 | 0.723 | 0.958 |
| MARG (Proposed) | **28.17** | **0.751** | **0.973** |

**Table 3.2: Quantitative evaluation results of the ablation studies for LDCT chest image denoising.**

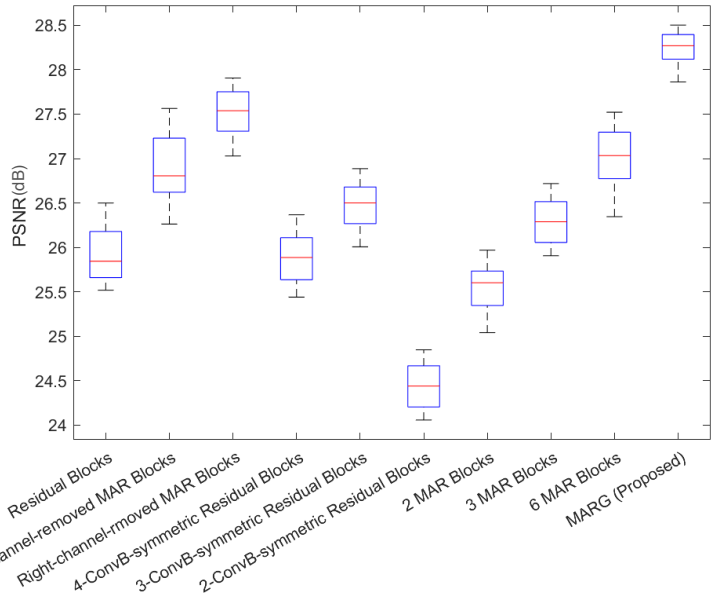


**Figure 3.16: Visual presentation of PSNR for the ablation study of LDCT chest image denoising.**

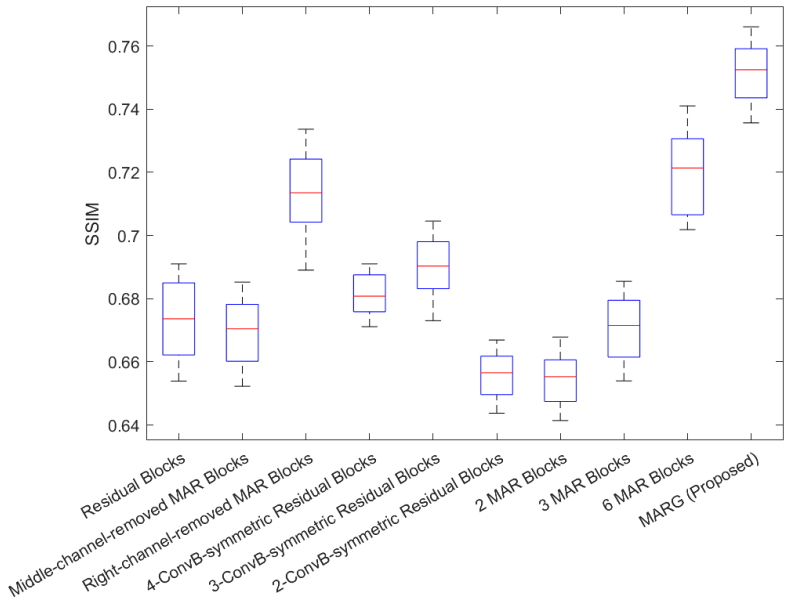


**Figure 3.17: Visual presentation of SSIM for the ablation study of LDCT chest image denoising.**

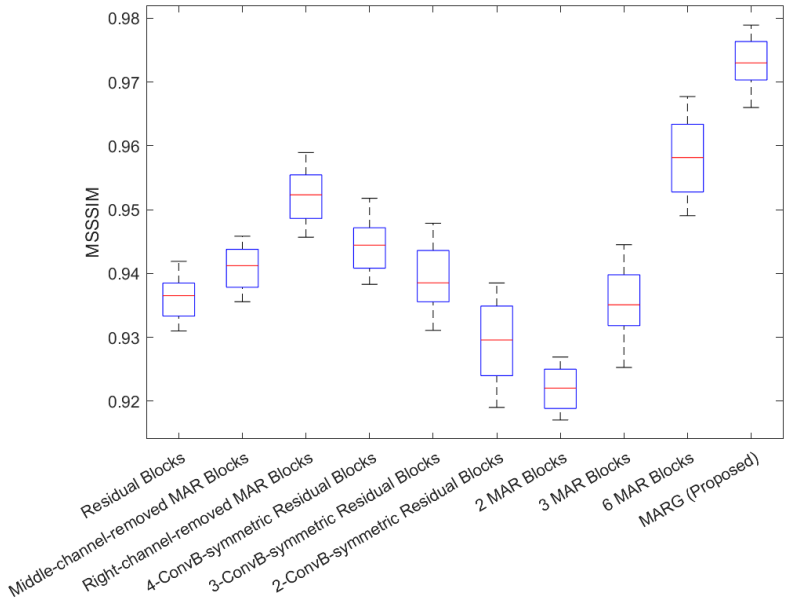


**Figure 3.18: Visual presentation of MSSSIM for the ablation study of LDCT chest image denoising.**

### 3.3.3 Limitation and Discussion

Although the proposed model achieved excellent performance in the task of LDCT chest image denoising, it has to admit that multichannel generators have drawbacks in some aspects, such as the fact that multichannel generators usually contain a large number of parameters, which may lead to an increase in computational complexity. Training and inference about such models may require more computational resources, especially in resource-limited environments. Another aspect is that multi-channel generators are specifically designed to be able to learn more complex image features at more scales; however, it is possible that such a design may be more likely to lead to overfitting, especially when the training data is relatively small or extremely lacking in diversity. Therefore, more effort may need to be spent on tuning the parameters of the model during training compared to a single-channel generator, and more measures need to be taken to mitigate the occurrence of overfitting.

## 3.4 Conclusions

Drawing on the experimental results and analyses presented above, this section summarises the main findings of the proposed unsupervised LDCT denoising approach. In this chapter, an unsupervised denoising framework for low-dose chest CT images was proposed. By introducing the multi-channel asymmetric residual (MAR) block and the corresponding generator architecture, the proposed method effectively suppresses noise while preserving critical anatomical structures. Extensive experimental results demonstrate that the proposed approach achieves state-of-the-art performance compared with existing LDCT denoising methods. More importantly, this chapter establishes a robust unsupervised denoising foundation,

which highlights the potential of learning noise characteristics directly from unpaired data and provides essential insights for more complex medical image processing tasks explored in subsequent chapters.

# Chapter 4

# Cross-Modality Brain MRI Synthesis Using Generative Adversarial Networks With Dual Channel Joint Discriminator

This chapter investigates the problem of unsupervised brain MRI cross-modality synthesis. A generative adversarial network with a dual-channel joint discriminator is proposed to improve the realism and structural fidelity of synthesised MRI images. The effectiveness of the proposed approach is validated through extensive experimental evaluations.

Recent developments in neuroscience have illuminated the utility of multi-modality medical data for examining specific diseases and comprehending human cognition. Despite this, obtaining complete sets of various modalities is hampered by challenges such as lengthy acquisition periods, expensive examination costs, and low patient comfort. Therefore, the synthesis of anatomically meaningful target modality data from source inputs becomes

an urgent need. The primary goal of MRI cross-modality synthesis is to enhance the information contained in MRI, improve image quality, and assist physicians in making more accurate diagnostic and therapeutic decisions. In some instances, MRI may be limited by insufficient contrast, low resolution, or the inability to depict specific tissue structures. Cross-modality synthesis enhances the quality and usability of MRI by extracting and integrating valuable information from other modalities, which enables patients to avoid multiple examination scans in various modalities. Therefore, MRI cross-modality synthesis is a powerful technique that can provide clinicians with additional supporting data and aid in more precise diagnosis and treatment. This is an essential research direction in medical image processing, and its prospective benefits to patient health and medical research are substantial. In this chapter, we propose a novel generative adversarial network with a dual channel discriminator to learn dedicated features from one MRI modality to synthesise another MRI modality. The dual-channel discriminator doubly constrains the generator at the local structure and pixel level, which improves the performance of the generator and produces more realistic results. Extensive experimental results proved the effectiveness and robustness of the proposed method by comparing it with some state-of-the-art methods.

## 4.1 Introduction

This section introduces the background and motivation of MRI cross-modality synthesis, highlights existing challenges in acquiring complete multi-modality data, and reviews related GAN-based synthesis approaches. The limitations of existing methods are discussed, which motivates the proposed dual-channel discriminator framework.

Magnetic Resonance Imaging (MRI) is a non-invasive medical imaging

technique that employs a magnetic field and non-ionising radio waves to generate images of human tissue with high contrast and spatial resolution. MRI is widely used in clinical and research applications for disease diagnosis, treatment planning, and research. Typically, MRI comprises multiple modalities, each of which corresponds to a distinct image parameter and contrast level. Common MRI modalities, for instance, include T1-weighted images, T2-weighted images, T2 FLAIR images, etc., which provide varying tissue contrast and functional information. Different image modalities offer unique advantages for detecting anatomical structures and lesions. Nonetheless, obtaining MRI images in various modalities typically necessitates distinct scan sequences and parameter settings, necessitating additional scanning time and resources. Moreover, acquiring images in multiple modalities may not always be feasible or practical for certain clinical situations and research requirements. Researchers have begun exploring cross-modality synthesis techniques in order to maximise the use of available MRI data and provide more comprehensive information. MRI cross-modality synthesis aims to synthesise an image from one MRI modality into another without requiring additional data collection. This method could provide physicians and researchers with more comprehensive image data to aid in the diagnosis and treatment of disease. By combining images from various modalities, more anatomical and functional information can be obtained to aid in the detection of lesions, the evaluation of their nature and localisation, and the monitoring of disease progression and treatment outcomes. Researchers have relied on machine learning techniques, specifically deep learning techniques, to accomplish cross-modality synthesis for MRI. For cross-modality synthesis tasks, deep learning models such as generative adversarial networks (GAN) [45] and variational auto-encoders (VAE) [73] are extensively employed. These models are capable of learning and establishing non-linear mapping relationships between various modalities to convert a source modality into a target modality. Cross-modality MRI synthesis has

significant potential for medical imaging applications. It can provide clinicians with more comprehensive image data and enhance the precision and reliability of disease diagnosis. It can also provide researchers with additional data resources to aid in the study of disease mechanisms and the development of novel treatments. Despite the success of MRI cross-modality synthesis techniques in a variety of studies and applications, there remain a number of obstacles and limitations. These include the intricate relationships between modalities, the diversity and disparity of the data, as well as the quality and precision of the synthesised images. In order to enhance the performance and dependability of MRI cross-modality synthesis techniques and promote their ubiquitous use in clinical practice and research, additional research and development are required.

Cross-modality synthesis of medical images is the transformation of an image from one modality into an image from another modality. Depending on the modalities that are synthesised, the cross-modality synthesis can be divided into the following categories: First, MRI to CT: due to MRI and CT images have distinct imaging principles and contrast characteristics, MRI to CI synthesis can assist physicians in better understanding MRI images and provide information comparable to CT images. Second, CT to MRI, in contrast to the aforementioned, converts CT images into MRI images. This can be accomplished by transferring the grayscale values of the CT image to the contrast range of the MRI image. Third, PET to CT or MRI, those three different types of images offer distinct information. By aligning and synthesising PET and CT or MRI images, it is possible to combine the functional information of CT or MRI to provide more thorough diagnostic information. Fourth, synthesis of different modality of MRI. As previously stated, MRI images have different modalities, each of which provides different contrast and information by setting different imaging parameters and pulse sequences. The reason for the necessity of acquiring different modalities of MRI is to obtain more comprehensive information that can be used

to help physicians make accurate diagnoses and assessments. Each MRI modality has different sensitivities to specific aspects of the tissue and can provide different contrast and image features to enhance the understanding of tissue structure, function, and pathological changes. Specifically, T1-weighted and T2-weighted images provide excellent contrast of the tissue's anatomical structure. T1-weighted images represent the tissue's fundamental morphology and positional relationships, whereas T2-weighted images depict the tissue's water content and pathological changes such as inflammation and oedema. By simultaneously acquiring images from both modalities, a more comprehensive picture of the anatomy can be obtained. Various pathological processes produce distinct signal characteristics. By scanning multiple modalities, physicians are able to evaluate how a lesion functions in various modalities and determine its characteristics and extent. T1-weighted and T2-weighted images display distinct signal characteristics for tumours, and FLAIR images can help detect oedema surrounding the tumour. Certain MRI techniques can provide information regarding tissue function. For instance, dynamic contrast-enhanced images (DCE-MRI) can assess the blood supply of a tumour, and diffusion-weighted images (DWI) can disclose the diffuse nature of tissue water molecules. This functional information is essential for the diagnosis and assessment of treatment responses.

Although complete multi-modality data can provide obvious benefits, acquiring a full set of paired multi-modality data has significant constraints. For example, many medical institutions are unable to share data because local regulations prohibit the sharing of medical data, although identifying information has been removed to safeguard patient privacy. On the other hand, the patient's movements may significantly misalign the obtained data. Also, collecting paired multi-modality data is quite costly. Therefore, the necessity of using cross-modality synthesis to deal with missing data has become particularly important.

Existing medical image translation methods [67, 74, 121] have proven their potential in the field of medical image processing. Among these methods, the GAN-based method remains the prevailing approach for medical image cross-modality synthesis, especially those supervised GANs [31, 127, 145, 162, 164, 179]. However, supervised learning-based GANs require that the training data be paired, which increases the difficulty of obtaining training data. To avoid the need for paired data, semi-supervised learning-based GANs and unsupervised learning-based GANs become necessary. Thus, some methods [50, 128, 176] have appeared that direct the cross-modality synthesis through the use of high-level tasks without the need to use paired data for training. Huang *et al.* [61, 62] project the unpaired training data into a common space and use the attributed features in that space to assist in synthesising the missing modality data. In addition to this, there are many GAN-based models that have been used to handle a variety of different image modalities. For example, [16, 25, 31, 52, 64, 100, 111, 158, 163] for MRI, [56] for CT, [12, 146] for PET, among others. The implementation of these approaches has significantly enhanced the efficacy of medical image synthesis through the establishment of diverse objectives.

Nevertheless, it is worth noting that the aforementioned GANs-based model may have a potential drawback in that it prioritises the acquisition of a global mapping from the source domain to the target domain. As a result, certain local structural information and pixel-level information are overlooked. Consequently, it becomes challenging to develop a global model that is optimal for each individual sample. Therefore, we propose an appropriate resolution to this issue.

In this chapter, we propose an unsupervised learning-based method that eliminates the need to collect pairs of databases. We construct an innovative dual-channel discriminator structure tailored to the task of medical image cross-modality synthesis that can provide feedback from both local struc-

ture and pixel space to the generator, which enhances the network's performance and the synthesised results. The details are provided in the method section of this chapter. After the network structure has been constructed, the network will be trained to optimise the parameters by minimising the loss function, so that the generator attempts to learn the mapping between the source domain and target domain and produces images that are as close to the target domain as possible. The training process employs an adversarial training strategy, iteratively training the generator and discriminator so that they achieve the goal of enhancing the generative effect and discriminative ability. The model obtained from training is evaluated and optimised, and its efficacy is evaluated using a test dataset. The proposed model achieved state-of-the-art results in an unsupervised brain MRI cross-modality synthesis task.

## 4.2 Method

To formally describe the proposed unsupervised MRI cross-modality synthesis framework, this section introduces the overall network design, problem formulation, and optimisation objectives.

The proposed method in this chapter aims to perform unsupervised MRI cross-modality synthesis, that is, the conversion and synthesis of MRI between different modalities, by learning the mapping from one MRI modality to another without paired data. We propose an innovative dual-channel joint discriminator (DCD) structure for this task.

### 4.2.1 Preliminaries

This subsection briefly reviews the basic principles of generative adversarial networks and CycleGAN-based unsupervised learning, which form the theoretical foundation of the proposed method.

According to GANs, a generator $G$ and a discriminator $D$ interact in a game of zero-sum in order to optimise the learning parameters. Although GANs can generate sharp images, primarily under supervised settings and with slightly erratic performance. CycleGAN [178] is a variant of GANs that models both forward and reverse mapping functions in a closed system, which adds a cycle-consistency constraint to promote more realistic image style transformation.

### 4.2.2 Problem Formulation

Building upon the GAN framework introduced above, this subsection formulates the unsupervised MRI cross-modality synthesis problem addressed in this chapter, including the definition of source and target domains and the bidirectional mapping strategy.

The aim of the proposed method is to learn two mappings between different modalities of MRI. The two different modalities have been defined as domain $T$ and domain $P$. The two mappings are: $G : T \rightarrow P$ and $F : P \rightarrow T$. We adapt the idea of cyclic strategies for purpose of unsupervised learning, using cycle consistency loss to restrict between the two generators $G$ and $F$. In contrast, given two discriminators $D_P$ and $D_T$ to learn and train against the generators. In this work, the MRI cross-modality synthesis can be divide into two stages:

(1) $G : T \rightarrow P$ is constructed to synthesise the domain $P$ from the domain

$T$.

(2) $F : P \rightarrow T$ is used to restore the domain $T$ from the domain $P$ and thus detect the stability of the generator.

For two sets of unpaired images $\{I_i^T\}_{i=1}^M \in T$ and $\{I_j^P\}_{j=1}^N \in P$, we start with a synthesis from domain $T$ to domain $P$. The synthesis process is to learn a generator $G$ that maps an input image $I_i^T$ to another modality's corresponding image $I_j^P$. Formally, the process can be expressed as:

$$(I_i^T)_{synthesised} = G(I_j^P) \approx I_i^T, \tag{4.1}$$

where $(I_i^T)_{synthesised}$ denotes the synthesised modality image. Generally, cross-modality synthesis can be seen as image translation. Hence, the use of GAN-based techniques can enhance the visual quality of the synthesised image. In contrast to typical GANs that utilise noise vectors to generate images, our cross-modality synthesis model exclusively relies on one modality of MRI as input. This study employed the CycleGAN [178] as the backbone to demonstrate the performance of the innovative Dual-Channel joint Discriminator (DCD). The use of CycleGAN as a backbone network is due to the fact that it is based on unsupervised learning and excels at cross-modal transformations. CycleGAN has been widely used and maturely implemented in image translation.

Typically, the discriminator of the GAN-based cross-modality synthesis approaches downsamples the input to a scalar value progressively. This will lead the discriminator to easily disregard previous samples due to the distribution of synthesised image changes as the generator adjusts continuously during the training process, preventing it from maintaining a robust data representation to describe the local and pixel-wise image variations. To deal with this problem, we update the standard classification discriminator in GANs with a dual-channel joint discriminator that can combine the local

and pixel information of the input image to improve the performance of the discriminator. The detailed discriminator structure is shown in Fig 4.3 and will be introduced in the next section.

### 4.2.3 Loss Function

Following the problem formulation, this subsection defines the loss functions used to optimise the proposed synthesis model, including adversarial loss, cycle-consistency loss, and pixel-level constraints.

The DCD contains two channels, we use $D_{Patch}$ to denote the patch discriminator channel, and $D_{Pixel}$ represents the pixel discriminator channel. The traditional discriminator only distinguishes the input as real or fake based on the encoder's output. Compared to the traditional discriminator, the DCD loss is computed based on the outputs of both $D_{Patch}$ and $D_{Pixel}$, which can provide pixel-level and local structural information to the generator. In the proposed method, we apply least-squares GANs [99] instead of conventional GANs [45] to enhance the stability of the training process while improving the visual quality of the synthesised image. Formally, the discriminator loss for the discriminator $D_P$ can be written as Eq. 4.2:

$$\begin{aligned}\mathcal{L}_{D_P} =& \mathbb{E}_{\boldsymbol{I}^T} \parallel D_{Patch}(\boldsymbol{I}^T) - 1 \parallel_2^2 + \mathbb{E}_{\boldsymbol{I}^P} \parallel D_{Patch}(\boldsymbol{I}^T_{synthesised}) \parallel_2^2 \\ &+ \mathbb{E}_{\boldsymbol{I}^T} \parallel D_{Pixel}(\boldsymbol{I}^T) - 1 \parallel_2^2 + \mathbb{E}_{\boldsymbol{I}^P} \parallel D_{Pixel}(\boldsymbol{I}^T_{synthesised}) \parallel_2^2,\end{aligned} \tag{4.2}$$

where $\boldsymbol{I}^T$ is the image from domain $T$, $\boldsymbol{I}^P$ is the image from domain $P$, $\boldsymbol{I}^T_{synthesised}$ is the synthesised image of domain $T$, $\boldsymbol{I}^P_{synthesised}$ is the synthesised image of domain $P$, and 1 is the least-squares GANs decision boundary. The first two terms represent the loss of the patch discriminator channel, and the last two terms represent the loss of the pixel discriminator

channel. The sum of the four terms is the total loss of the discriminator $D_P$. Correspondingly, the loss of the discriminator $D_T$ can be written as Eq (4.3):

$$\begin{aligned}\mathcal{L}_{D_T} =& \mathbb{E}_{\boldsymbol{I}^P} \parallel D_{Patch}(\boldsymbol{I}^P) - 1 \parallel_2^2 + \mathbb{E}_{\boldsymbol{I}^T} \parallel D_{Patch}(\boldsymbol{I}^P_{synthesised}) \parallel_2^2 \\ &+ \mathbb{E}_{\boldsymbol{I}^P} \parallel D_{Pixel}(\boldsymbol{I}^P) - 1 \parallel_2^2 + \mathbb{E}_{\boldsymbol{I}^T} \parallel D_{Pixel}(\boldsymbol{I}^P_{synthesised}) \parallel_2^2 .\end{aligned} \tag{4.3}$$

Therefore, the DCD loss can be defined as:

$$\mathcal{L}_D = \mathcal{L}_{D_P} + \mathcal{L}_{D_T}. \tag{4.4}$$

In the proposed method, we adopt the sum of the two adversarial losses to both mapping functions as the total adversarial loss. The objective can be expressed as:

$$\begin{aligned}\mathcal{L}_{adv} =& \mathbb{E}_{\boldsymbol{I}^P \sim p_{\text{data}}(\boldsymbol{I}^P)}[\log D_G(\boldsymbol{I}^P)] + \mathbb{E}_{\boldsymbol{I}^T \sim p_{\text{data}}(\boldsymbol{I}^T)}[\log(1 - D_G(G(\boldsymbol{I}^T)))] + \\ & \mathbb{E}_{\boldsymbol{I}^T \sim p_{\text{data}}(\boldsymbol{I}^T)}[\log D_F(\boldsymbol{I}^T)] + \mathbb{E}_{\boldsymbol{I}^P \sim p_{\text{data}}(\boldsymbol{I}^P)}[\log(1 - D_F(F(\boldsymbol{I}^P)))].\end{aligned} \tag{4.5}$$

Meanwhile, we use the cycle consistency loss to restrict the image from remaining consistent as it transitions from one modality to another, which can be written as:

$$\mathcal{L}_{cyc} = \mathbb{E}_{\boldsymbol{I}^T \sim p_{\text{data}}(\boldsymbol{I}^T)} \parallel \boldsymbol{I}^T - F(G(\boldsymbol{I}^T)) \parallel_1 + \mathbb{E}_{\boldsymbol{I}^P \sim p_{\text{data}}(\boldsymbol{I}^P)} \parallel \boldsymbol{I}^P - G(F(\boldsymbol{I}^P)) \parallel_1 . \tag{4.6}$$

In conjunction with the innovative dual-channel discriminator proposed

in this method, in order to encourage the generator output of the cross-modality synthesised image to be more closely matched with the original modality image at the pixel level, we employ the pixel-wise loss between the original image and synthesised image, which is defined as follows:

$$\mathcal{L}_{pixel} = \mathbb{E}_{(\boldsymbol{I}^T,\boldsymbol{I}^P)} \| \boldsymbol{I}^T - \boldsymbol{I}^P_{synthesised} \|_F^2 + \mathbb{E}_{(\boldsymbol{I}^T,\boldsymbol{I}^P)} \| \boldsymbol{I}^T_{synthesised} - \boldsymbol{I}^P \|_F^2, \tag{4.7}$$

where $\| \|_F$ refers to the Frobenius norm.

In combination with the above, our final loss function used to optimise the generators can be represented as:

$$\mathcal{L}_{G,F} = \lambda_{adv}\mathcal{L}_{adv} + \lambda_{cyc}\mathcal{L}_{cyc} + \lambda_{pixel}\mathcal{L}_{pixel}, \tag{4.8}$$

where $\lambda_{adv}, \lambda_{cyc}, \lambda_{pixle}$ represent the weights of $\mathcal{L}_{adv}, \mathcal{L}_{cyc}, \mathcal{L}_{pixel}$ respectively.

## 4.3 Experiments

To evaluate the effectiveness and robustness of the proposed method, this section presents the experimental setup, network architectures, and quantitative and qualitative analyses for MRI cross-modality synthesis.

### 4.3.1 Network Structures

This subsection describes the architecture of the proposed generators and the dual-channel joint discriminator, detailing how local structural and pixel-level information are jointly exploited.

The structure of the CMS network is shown in Fig 4.1, where the generator $G$ is responsible for generating the image in domain $P$ from domain $T$, and the generator $F$ is responsible for generating the image in domain $T$ from domain $P$. The Dual-Channel Discriminator $D_P$ and $D_T$ are used to judge the authenticity of the generated image, respectively. The specific structure of the generators and discriminators is illustrated below.

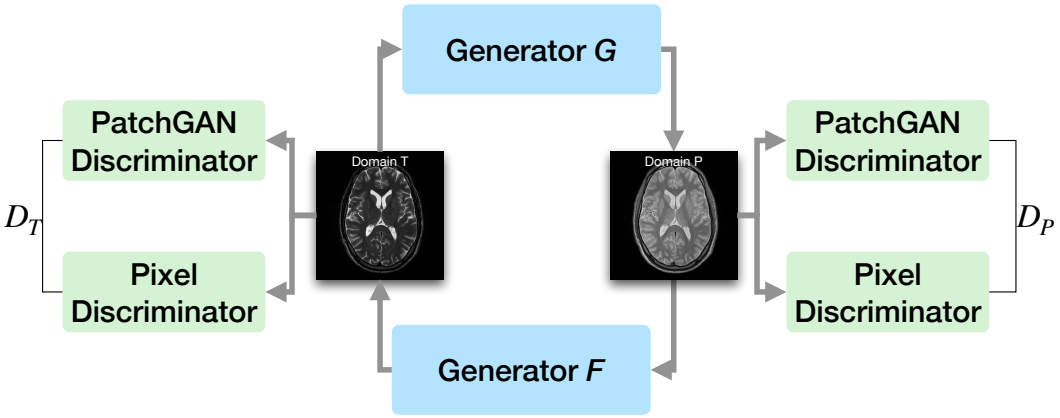


**Figure 4.1: The network structure of the proposed Dual-Channel Discriminator GANs**

The architecture of the generative networks is inspired by Johnson *et al.* [69], which demonstrated remarkable results for image style transfer. The generator is comprised of an encoder module followed by a decoder module. The encoder module includes an initial block of layers, downsampling blocks, and residual blocks. The initial block is composed of an input layer, a 2D convolutional layer with a stride of (1, 1), a batch normalisation layer, and an activation layer. In the downsampling block, the structure is organised as 2DConv-BatchNorm-ReLU. The input is downsampled by a factor of 2 according to the number of downsampling blocks using the encoder module. The decoder module contains the upsampling block and

final block. The upsampling layer upsamples the data by a factor of 2 according to the number of upsampling blocks. The final block is constructed of a 2D convolutional layer with a stride of (1, 1) and an activation layer, which has been set as a Tanh layer. The structure of the generator is shown in Fig 4.2.

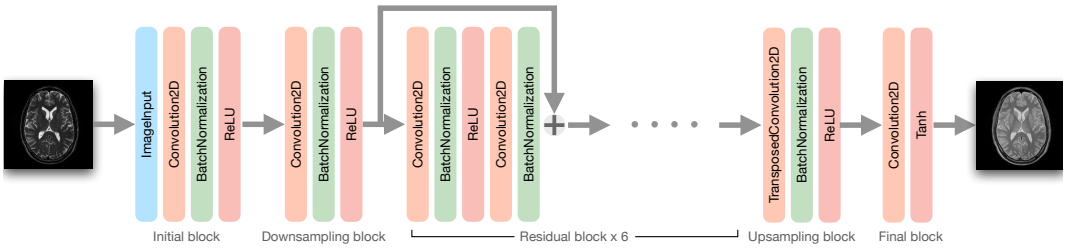


**Figure 4.2: The generator structure of DCD GANs**

For the discriminator, we create an innovative Dual-Channel Discriminator (DCD) structure that provides feedback to the generator from both local structure and pixel space, respectively. The feedback from the local structure promotes the generator to produce more realistic images, and the feedback from the pixel space enhances the edge information and reduces artifacts. This structure combines a patchGAN discriminator and a pixel discriminator. The structure of the DCD is shown in Fig 4.3 and the details of both channels in the DCD are described below.

(1) The first channel of DCD is the patchGAN discriminator channel, which consists of an encoder module that includes an initial block, downsampling blocks, and final block. The initial block contains the image input

layer, a 2D convolutional layer with a stride of (2, 2), and a leaky ReLU layer as an activation layer. The downsampling block consists of a 2D convolutional layer with a stride of (2, 2), a batch normalisation layer, and also a Leaky ReLU layer. The final block is built with a 2D convolution layer with a stride of (1, 1), a batch normalisation layer, a Leaky ReLU layer, and a second 2D convolutional layer with a stride of (1, 1) and 1 output channel. The patchGAN discriminator channel analysis is shown in Fig 4.4.

(2) The second channel of DCD is the pixel discriminator, which does not perform downsampling, so it only has an initial block and a final block. The initial block is built up with an image input layer, a 2D convolutional layer with a stride of (1, 1), and a leaky ReLU layer. The final block contains a 2D convolutional layer with a stride of (1, 1), a batch normalisation layer, a Leaky ReLU layer, and another 2D convolutional layer with a stride of (1, 1), and the output channel has been set to 1. The pixel discriminator channel analysis is shown in Fig 4.5.

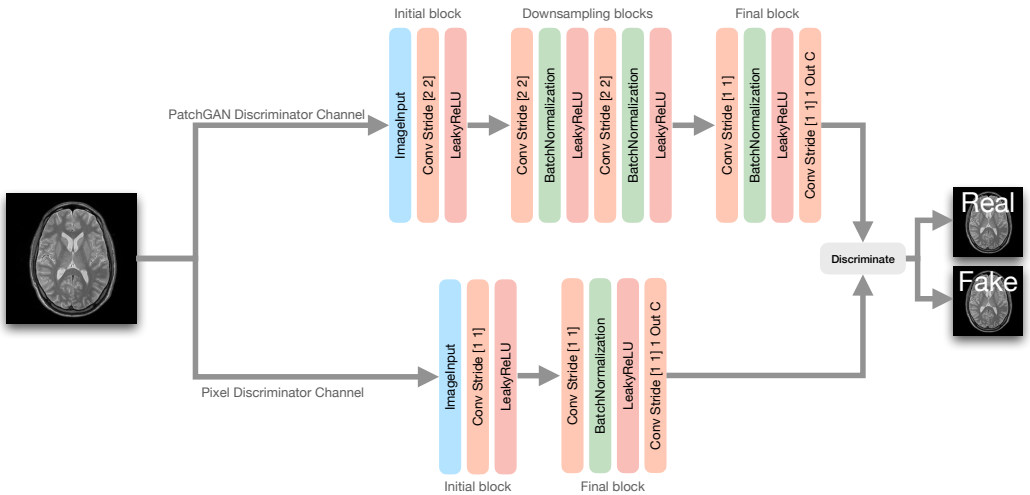


**Figure 4.3: The structure of Dual-Channel-Discriminator(DCD)**

| | Name | Type | Activations | Learnable Properties | States |
|---|---|---|---|---|---|
| 1 | input_top<br>128×128×1 images | Image Input | 128(S) × 128(S) × 1(C) × 1(B) | - | - |
| 2 | conv2d_top<br>64 4×4×1 convolutions with stride [2 2] and padding [1 1 1 1] | 2-D Convolution | 64(S) × 64(S) × 64(C) × 1(B) | Weights 4 × 4 × 1 × 64<br>Bias 1 × 1 × 64 | - |
| 3 | act_top<br>Leaky ReLU with scale 0.2 | Leaky ReLU | 64(S) × 64(S) × 64(C) × 1(B) | - | - |
| 4 | conv2d_mid_1<br>128 4×4×64 convolutions with stride [2 2] and padding [1 1 1 1] | 2-D Convolution | 32(S) × 32(S) × 128(C) × 1(B) | Weights 4 × 4 × 64 × 128<br>Bias 1 × 1 × 128 | - |
| 5 | norm2d_mid_1<br>Batch normalization with 128 channels | Batch Normalization | 32(S) × 32(S) × 128(C) × 1(B) | Offset 1 × 1 × 128<br>Scale 1 × 1 × 128 | TrainedMean 1 × 1 × 128<br>TrainedVariance 1 × 1 × 128 |
| 6 | act_mid_1<br>Leaky ReLU with scale 0.2 | Leaky ReLU | 32(S) × 32(S) × 128(C) × 1(B) | - | - |
| 7 | conv2d_mid_2<br>256 4×4×128 convolutions with stride [2 2] and padding [1 1 1 1] | 2-D Convolution | 16(S) × 16(S) × 256(C) × 1(B) | Weights 4 × 4 × 128 × 256<br>Bias 1 × 1 × 256 | - |
| 8 | norm2d_mid_2<br>Batch normalization with 256 channels | Batch Normalization | 16(S) × 16(S) × 256(C) × 1(B) | Offset 1 × 1 × 256<br>Scale 1 × 1 × 256 | TrainedMean 1 × 1 × 256<br>TrainedVariance 1 × 1 × 256 |
| 9 | act_mid_2<br>Leaky ReLU with scale 0.2 | Leaky ReLU | 16(S) × 16(S) × 256(C) × 1(B) | - | - |
| 10 | conv2d_tail<br>512 4×4×256 convolutions with stride [1 1] and padding [1 1 1 1] | 2-D Convolution | 15(S) × 15(S) × 512(C) × 1(B) | Weights 4 × 4 × 256 × 512<br>Bias 1 × 1 × 512 | - |
| 11 | norm2d_tail<br>Batch normalization with 512 channels | Batch Normalization | 15(S) × 15(S) × 512(C) × 1(B) | Offset 1 × 1 × 512<br>Scale 1 × 1 × 512 | TrainedMean 1 × 1 × 512<br>TrainedVariance 1 × 1 × 512 |
| 12 | act_tail<br>Leaky ReLU with scale 0.2 | Leaky ReLU | 15(S) × 15(S) × 512(C) × 1(B) | - | - |
| 13 | conv2d_final<br>1 4×4×512 convolutions with stride [1 1] and padding [1 1 1 1] | 2-D Convolution | 14(S) × 14(S) × 1(C) × 1(B) | Weights 4 × 4 × 512 × 1<br>Bias 1 × 1 × 1 | - |

**Figure 4.4: PatchGAN Discriminator Channel Analysis**

| | Name | Type | Activations | Learnable Properties | States |
|---|---|---|---|---|---|
| 1 | input_top<br>128×128×1 images | Image Input | 128(S) × 128(S) × 1(C) × 1(B) | - | - |
| 2 | conv2d_top<br>64 1×1×1 convolutions with stride [1 1] and padding [0 0 0 0] | 2-D Convolution | 128(S) × 128(S) × 64(C) × 1(B) | Weights 1 × 1 × 1 × 64<br>Bias 1 × 1 × 64 | - |
| 3 | act_top<br>Leaky ReLU with scale 0.2 | Leaky ReLU | 128(S) × 128(S) × 64(C) × 1(B) | - | - |
| 4 | conv2d_mid_1<br>128 1×1×64 convolutions with stride [1 1] and padding [0 0 0 0] | 2-D Convolution | 128(S) × 128(S) × 128(C) × 1(B) | Weights 1 × 1 × 64 × 128<br>Bias 1 × 1 × 128 | - |
| 5 | norm2d_mid_1<br>Batch normalization with 128 channels | Batch Normalization | 128(S) × 128(S) × 128(C) × 1(B) | Offset 1 × 1 × 128<br>Scale 1 × 1 × 128 | TrainedMean 1 × 1 × 128<br>TrainedVariance 1 × 1 × 128 |
| 6 | act_mid_1<br>Leaky ReLU with scale 0.2 | Leaky ReLU | 128(S) × 128(S) × 128(C) × 1(B) | - | - |
| 7 | conv2d_final<br>1 1×1×128 convolutions with stride [1 1] and padding [0 0 0 0] | 2-D Convolution | 128(S) × 128(S) × 1(C) × 1(B) | Weights 1 × 1 × 128 × 1<br>Bias 1 × 1 × 1 | - |

**Figure 4.5: Pixel Discriminator Channel Analysis**

### 4.3.2 Experimental Setup

In accordance with the described network architecture, this subsection details the datasets, training configuration, and evaluation protocols used in the experiments.

#### Datasets

Information extraction from the Images (IXI) [123] dataset is used in the chapter, which consists of 578 MR images from normal and healthy subjects. The protocol for acquiring MR images for each subject contains T1-weighted images, T2-weighted images, proton density (PD)-weighted images, Magnetic Resonance Angiography (MRA) images, and 15 directions of Diffusion-weighted images. Fig 4.6 shows T2-weighted and PD-weighted images collected from the IXI dataset.

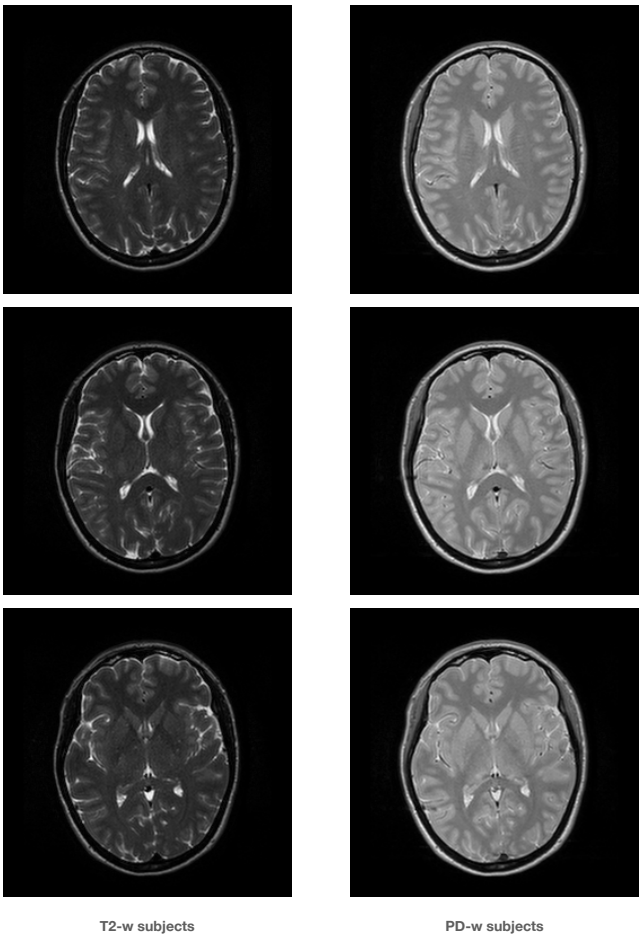


**Figure 4.6: Illustration of the example from the IXI dataset. The first column is the T2-weighted image, and the second column shows the corresponding PD-w image.**

The images were collected by different systems in three different hospitals in London. The Philips 3T system has been used in Hammersmith Hospital; the scanner parameters are shown in Table 4.1.

Philips Medical Systems Intera 3T

| | T1 parameters | T2 parameters | PD parameters | MRA parameters | DTI parameters |
|---|---|---|---|---|---|
| Repetition Time | 9.60000038146972 | 5725.79052734375 | 5725.79052734375 | 16.7210998535156 | 11894.4384765625 |
| Echo time | 4.60269975662231 | 100.0 | 8.0 | 5.75335741043090 | 51.0 |
| Number of Phase Encoding Steps | 208 | 187 | 187 | 286 | 110 |
| Echo Train Length | 208 | 16 | 16 | 0 | 0 |
| Reconstruction Diameter | 240.0 | 240.0 | 240.0 | 240.0 | 224.0 |
| Acquisition Matrix | 208 x 208 | 192 x 187 | 192x187 | 288 x 286 | 112 x 110 |
| Flip Angle | 8.0 | 90.0 | 90.0 | 16.0 | 90.0 |

**Table 4.1: Philips Medical Systems Intera 3T scanning parameters**

Philips 1.5T system has been used in Guy's Hospital, Table 4.2 shows the scanner parameters. Institute of Psychiatry using a GE 1.5T system but the scanner parameters are not available. All collected data has been stored in NIFTI format.

Philips Medical Systems Gyroscan Intera 1.5T

| | T1 parameters | T2 parameters | PD parameters | MRA parameters | DTI parameters |
|---|---|---|---|---|---|
| Repetition Time | 9.813 | 8178.34 | 8178.34 | 20 | 9054.01 |
| Echo time | 4.603 | 100.0 | 8 | 6.9052 | 80 |
| Number of Phase Encoding Steps | 192 | 187 | 187 | 286 | 94 |
| Echo Train Length | 0 | 16 | 16 | 0 | 0 |
| Reconstruction Diameter | 240 | 240 | 240 | 240 | 224 |
| Flip Angle | 8.0 | 90 | 90 | 25 | 90 |

**Table 4.2: Philips Medical Systems Gyroscan Intera 1.5T scanning parameters**

### Training Details

In the training process, the model is trained to update the generators and discriminators in alternating steps for each batch. For the setting of hyper-parameters, the initial learning rate is set to 0.0002. Utilise the stochastic gradient descent (SGD) with mini-batch size 16. We adopt the Adam solver [72] for optimisation. The gradient decay factor is 0.5, and the squared gradient decay factor is 0.999. Data augmentation involves rotating, stretching, and reflecting the training data. 8 patches with a size of $128 \times 128$ are randomly cropped for each training image. Before each training and validation epoch, shuffle the data. Following [88, 178], to manage the association between the objectives, we set $\lambda_{adv} = 1, \lambda_{cyc} = 10$, and $\lambda_{pixel} = 2$. A total of 100 epochs were trained.

### Experiments

The proposed method has been evaluated in two scenarios: (1) synthesising PD-w images from T2-w images. (2) generating T2-w images from PD-w images. Since the proposed approach is an unsupervised method, the preprocessing of the dataset needs to be done in a way that ensures it is unpaired. The IXI dataset has 578 paired subjects, with only T2-w data and the PD-w data involved in the experiment. For quantitative evaluation, we conduct a 5-fold cross-validation to evaluate our method. Remove the last 78 subjects from the IXI dataset and separate the remaining 500 subjects into 5 groups. Selecting 2 unpaired groups from T2-w and PD-w for training and the remaining 1 paired group data as testing data. Doing so ensures that there is no intersection of the training data between domain $T$ and domain $P$, i.e., there will be no pairs of T2-w and PD-w images in both domains. From the perspective of the xy-plane, the image size is $256 \times 256$. During the training process, in order to ensure that there is no overlap in the training set and to augment the number of training samples, the training data is randomly cropped into a $128 \times 128$ patch, and 8 patches are randomly cropped for each training image. This allows for a relative 8-fold increase in the size of the training data and better avoids overfitting. Before starting training at each epoch, the order of the training data is reshuffled, which is used to increase the diversity of the samples and thus improve the generalisation ability of the model.

### Baselines

As a comparison of the proposed method, several state-of-the-art cross-modality synthesis approaches were used as baseline methods, which include GAN [110], CycleGAN [178], pix2pix [65], cGAN [158], and pGAN [31]. Specifically, pix2pix and pGAN have been trained on paired data,

GAN, CycleGAN, and cGAN, and our method is unsupervised. The implementations are derived from the authors' available source code. We loaded and preprocessed the data in the same way as the method aforementioned. The parameters of these models are also set to maintain the same conditions as those of the authors. We compare the experiment results conducted on the same dataset. However, it is not uncommon that the results may differ slightly from those of the authors' due to hardware, randomness, or other minor factors.

### Qualitative Evaluations

The following figures show the results of cross-modality synthesis of brain MRI performed with the baseline methods and our method on the IXI dataset. Fig 4.7 shows the synthesised results from T2-w to PD-w. Fig 4.8 shows the synthesised results from PD-w to T2-w.

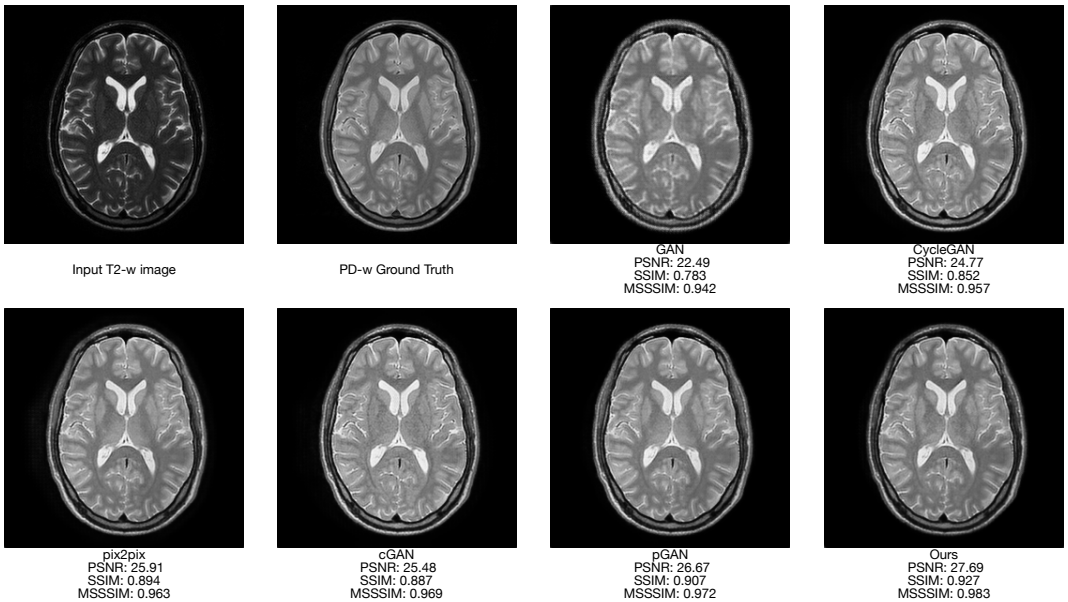


**Figure 4.7: Visual comparisons of GAN, CycleGAN, pix2pix, cGAN, pGAN and Ours for T2-w → PD-w.**

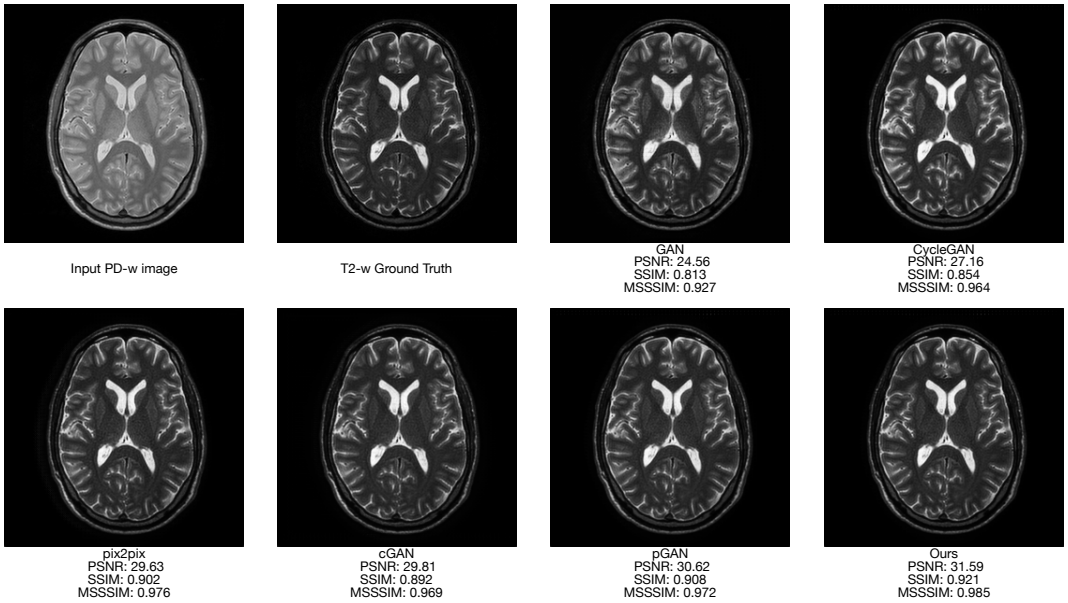


**Figure 4.8: Visual comparisons of GAN, CycleGAN, pix2pix, cGAN, pGAN and Ours for PD-w → T2-w.**

### Quantitative Results

We use PSNR, SSIM, and MSSSIM as evaluation criteria to objectively assess the quality of the synthesised results. The results of the proposed method and the five baseline methods are compared in Table 4.3 and Table 4.4. Notably, the proposed method achieves superior visual and quantitative performance than the other five baseline methods in all cases. The average performance of our method on the IXI dataset is (PSNR: 27.69dB, SSIM: 0.927, MSSSIM: 0.983) for synthesis from T2-w to PD-w and (PSNR: 31.59dB, SSIM: 0.921, MSSSIM: 0.985) for synthesis from PD-w to T2-w. Compared to baseline methods, our method has a significant performance improvement.

IXI: T2-w → PD-w

| Metric (avg.) | GAN | CycleGAN | pix2pix | cGAN | pGAN | Ours |
|---|---|---|---|---|---|---|
| PSNR (dB) | 22.49 | 24.77 | 25.91 | 25.48 | 26.67 | **27.69** |
| SSIM | 0.783 | 0.852 | 0.894 | 0.887 | 0.907 | **0.927** |
| MSSSIM | 0.942 | 0.957 | 0.963 | 0.969 | 0.972 | **0.983** |

**Table 4.3: Quantitative evaluation(PSNR(dB), SSIM, and MSSSIM): GAN, CycleGAN, pix2pix, cGAN, pGAN and Ours for T2-w → PD-w.**

IXI: PD-w → T2-w

| Metric (avg.) | GAN | CycleGAN | pix2pix | cGAN | pGAN | Ours |
|---|---|---|---|---|---|---|
| PSNR (dB) | 24.56 | 27.16 | 29.63 | 29.81 | 30.62 | **31.59** |
| SSIM | 0.813 | 0.854 | 0.902 | 0.892 | 0.908 | **0.921** |
| MSSSIM | 0.927 | 0.964 | 0.976 | 0.969 | 0.972 | **0.985** |

**Table 4.4: Quantitative evaluation(PSNR(dB), SSIM, and MSSSIM): GAN, CycleGAN, pix2pix, cGAN, pGAN and Ours for PD-w → T2-w.**

In order to provide a clearer presentation of the evaluation's findings, we present the results of the data visualisations in the following figures.

The stability of the overall performance of the proposed model can be seen in these plots. Specifically, it can be noticed that there are some baseline methods whose upper limit of performance exceeds the mean value of our model. As stated before, one of the main reasons for this situation is that the test data is a complete set of brain MRIs consisting of multiple slices. This also leads to the fact that the structural information contained in the topmost or bottommost slices is very limited, and most of them have a value of 0. This leads to some fluctuations in the performance metrics of each model in the visualisations displayed. Actually, these are acceptable, and we ultimately judge the performance of the models mainly by the average evaluation metric values.

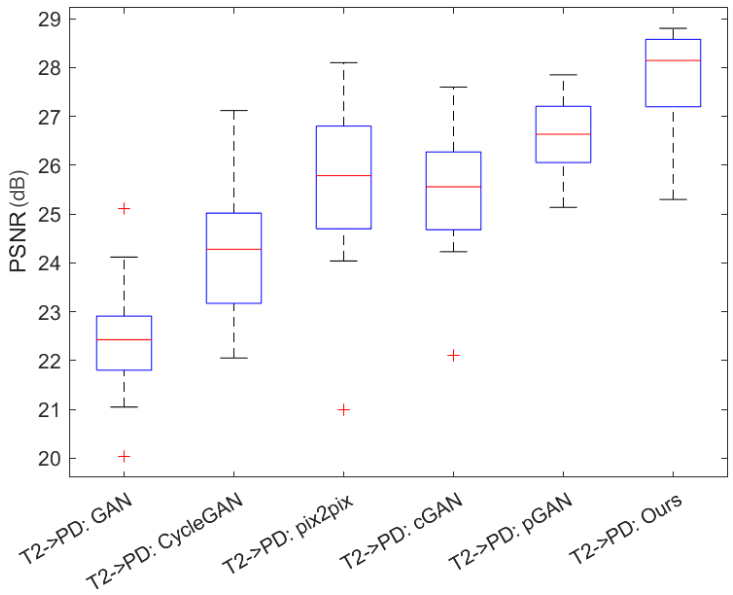


**Figure 4.9: Visual presentation of PSNR for the baseline methods and our method of T2-w → PD-w.**

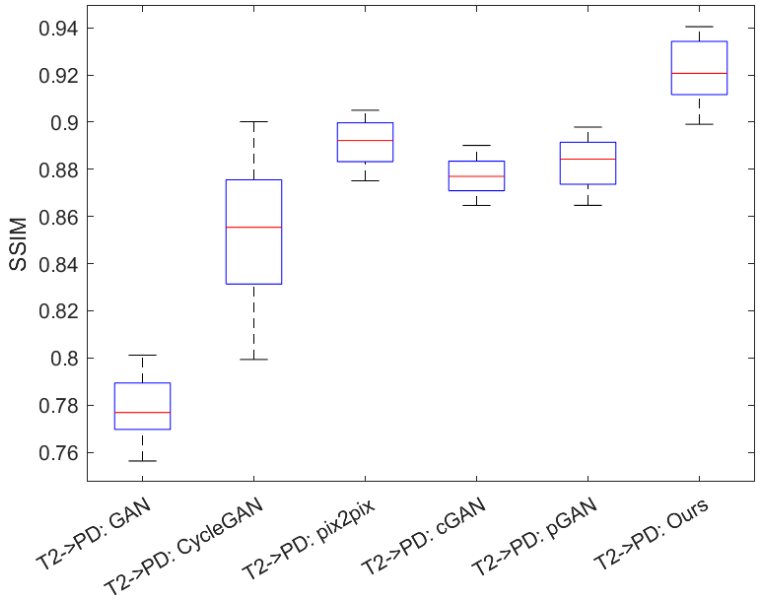


**Figure 4.10: Visual presentation of SSIM for the baseline methods and our method of T2-w → PD-w.**

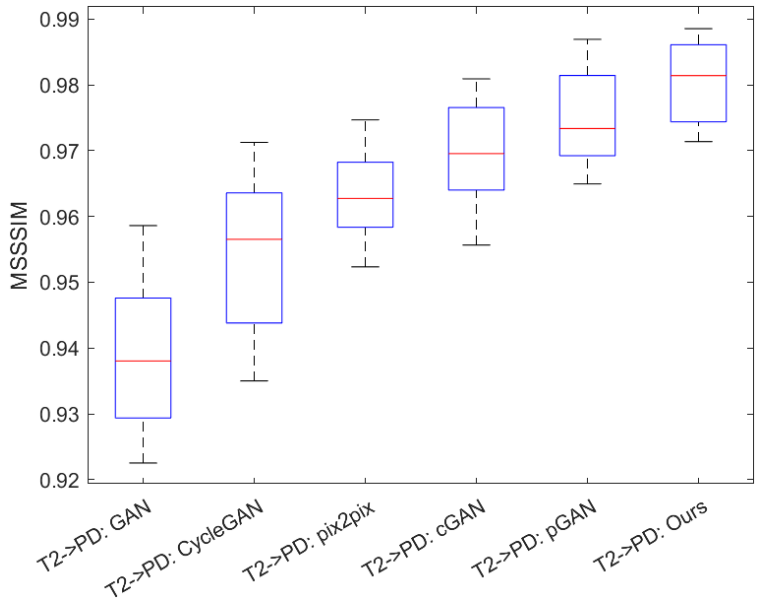


**Figure 4.11: Visual presentation of MSSSIM for the baseline methods and our method of T2-w → PD-w.**

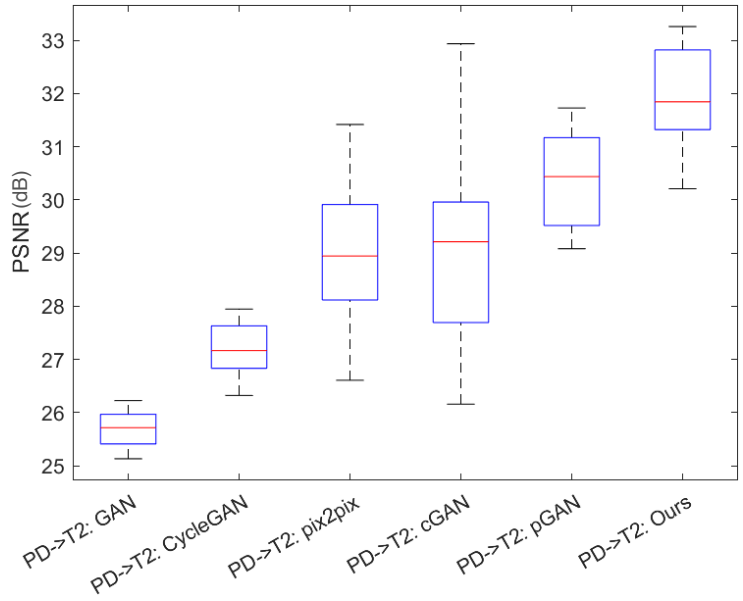


**Figure 4.12: Visual presentation of PSNR for the baseline methods and our method of PD-w → T2-w.**

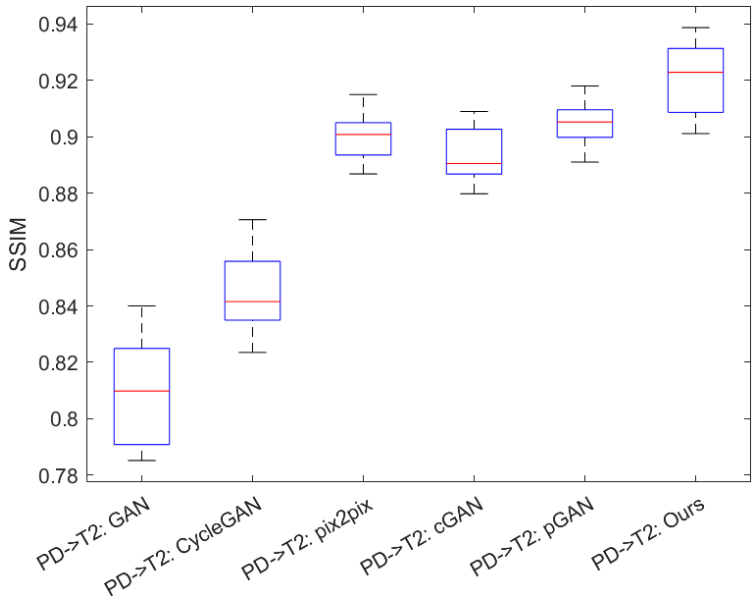


**Figure 4.13: Visual presentation of SSIM for the baseline methods and our method of PD-w → T2-w.**

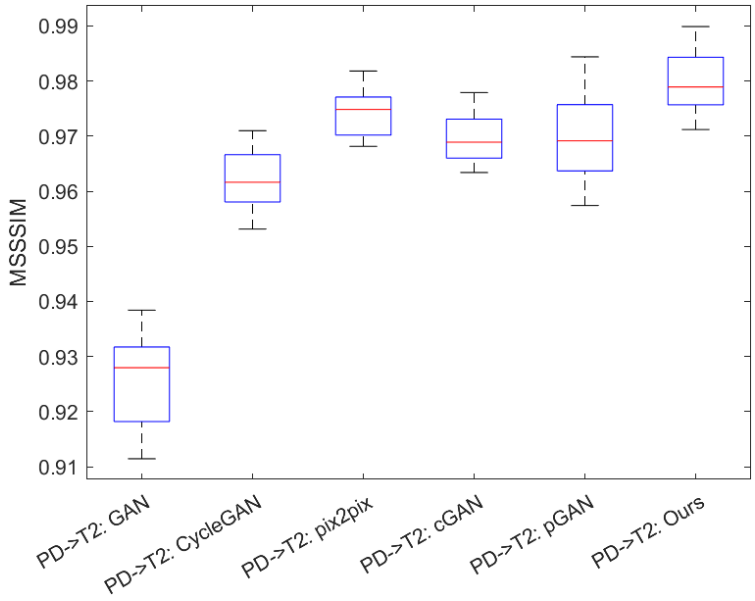


**Figure 4.14: Visual presentation of MSSSIM for the baseline methods and our method of PD-w → T2-w.**

### Ablation Studies

In this section, we performed an ablation study of the proposed method to better understand the reasons why our method achieves state-of-the-art performance. Since the proposed method contains multiple components, the ablation study will be performed separately from different perspectives. Firstly, we start with the structure of the discriminator. For our method, one of the most important innovations comes from the Dual-Channel Discriminator (DCD). To demonstrate the effectiveness of DCD, we will replace the DCD by substitution with a pixel discriminator, a patch discriminator, and a global discriminator, respectively. The same experiments as above will be performed on the IXI dataset, which are in two scenarios: (1) Cross-modality synthesis of the T2-w image from the PD-w image. (2) Cross-modality synthesis of the PD-w image from the T2-w image. For a fair comparison, the same patch size as in the original experiments was used in the experiments that replaced DCD with a pixel discriminator and patch

discriminator separately. As for the experiments of replacing DCD with a global discriminator, we use the global image size of $256 \times 256$ as the input size. Fig 4.15 shows the synthesised results from T2-w to PD-w of different discriminator structures, and the PD-w to T2-w results are shown in Fig 4.16. The evaluation results are shown in Table 4.5, which illustrates the effectiveness of the DCD structure. In order to better present the evaluation results visually, the PSNR, SSIM, and MSSSIM of the two scenarios are shown in Fig 4.17, Fig 4.18, and Fig 4.19 respectively.

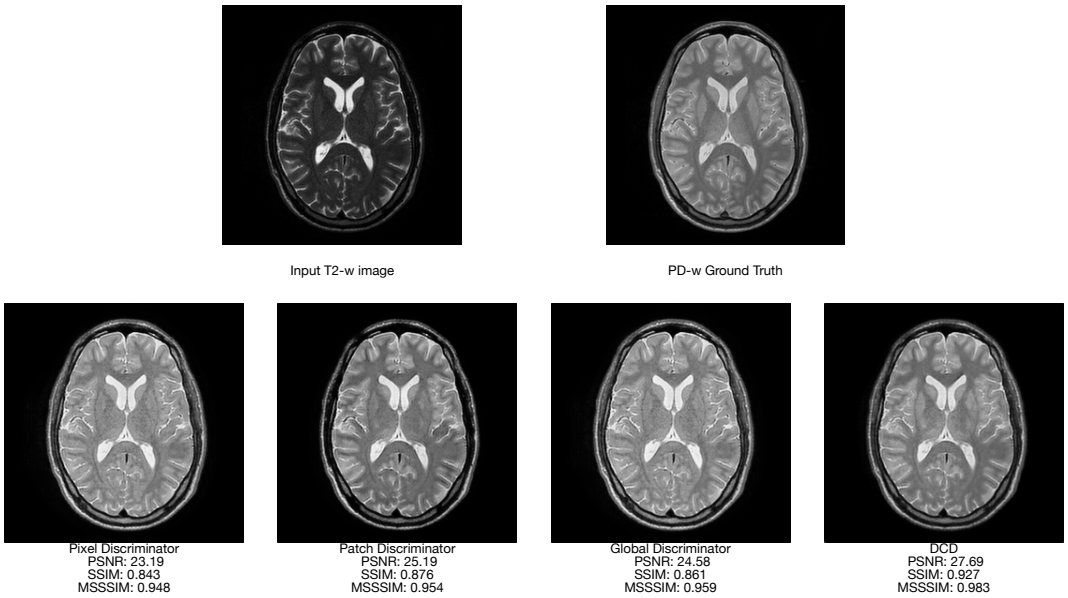


**Figure 4.15: Ablation study of different discriminators results from T2-w → PD-w.**

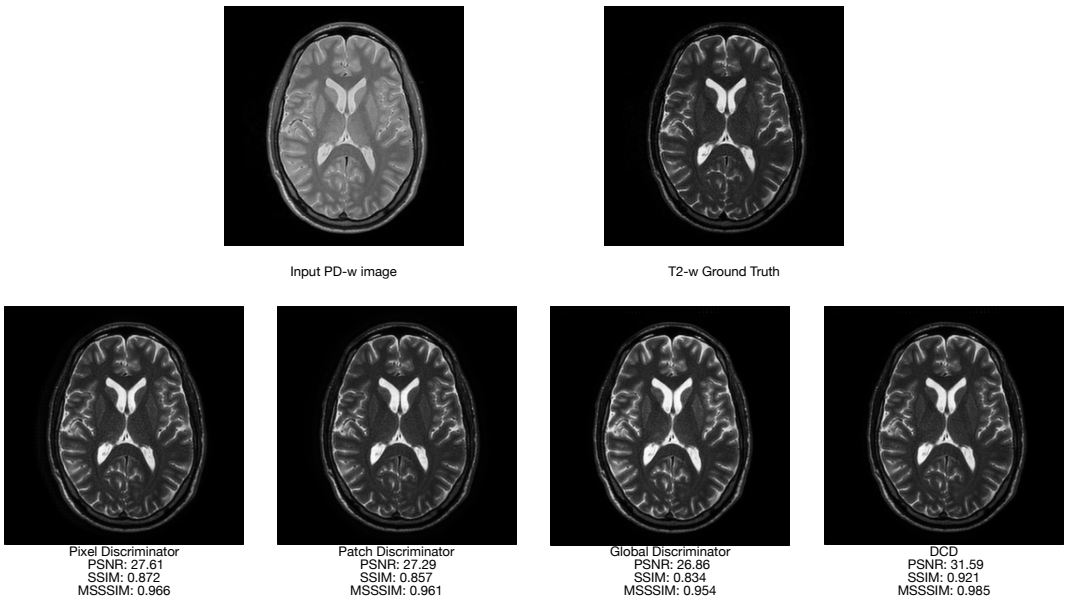


**Figure 4.16: Ablation study of different discriminators results from PD-w → T2-w.**

| Metric (avg.) | Pixel Discriminator | Patch Discriminator | Global Discriminator | DCD(Proposed) |
|---|---|---|---|---|
| **IXI: T2-w —> PD-w** | | | | |
| PSNR (dB) | 23.19 | 25.19 | 24.58 | **27.69** |
| SSIM | 0.843 | 0.876 | 0.861 | **0.927** |
| MSSSIM | 0.948 | 0.954 | 0.959 | **0.983** |
| **IXI: PD-w —> T2-w** | | | | |
| PSNR (dB) | 27.61 | 27.29 | 26.86 | **31.59** |
| SSIM | 0.872 | 0.857 | 0.834 | **0.921** |
| MSSSIM | 0.966 | 0.961 | 0.954 | **0.985** |

**Table 4.5: Ablation study of different discriminators evaluation results of the two scenarios. The best results are marked in bold.**

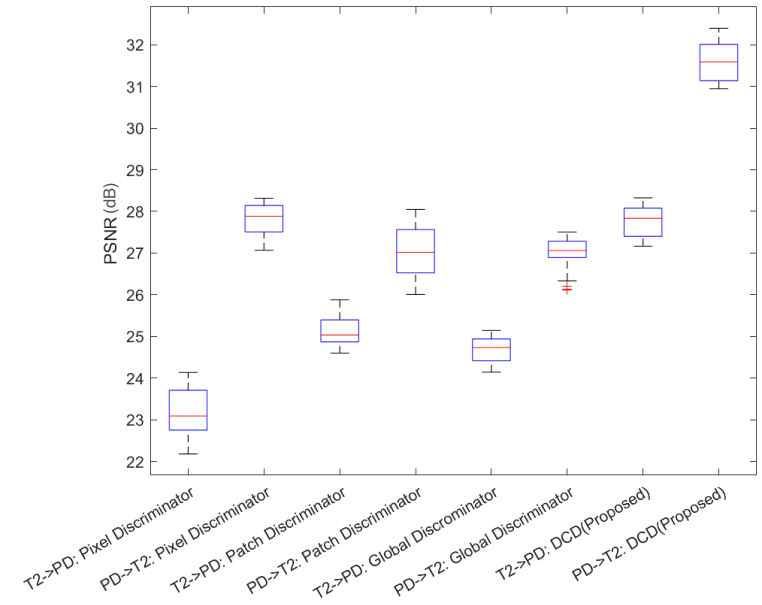


**Figure 4.17: Ablation study of different discriminators PSNR visual presentation of the two scenarios.**

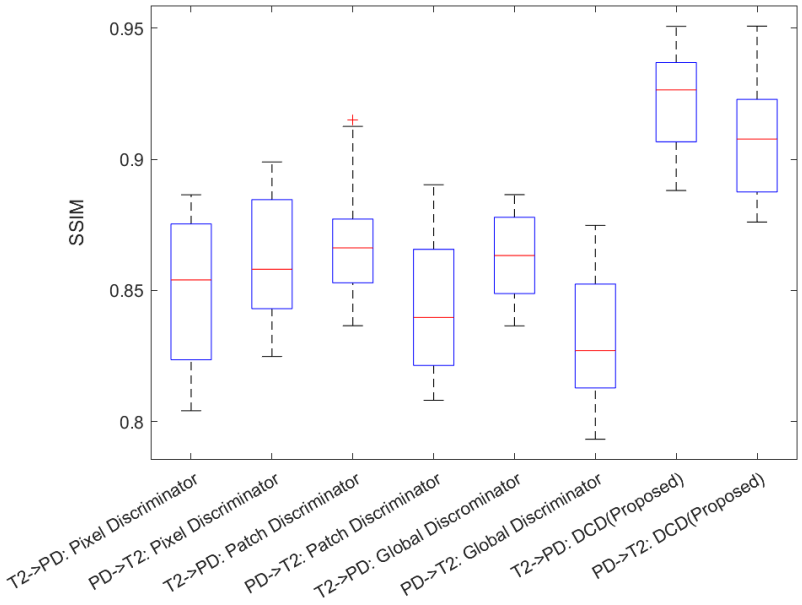


**Figure 4.18: Ablation study of different discriminators SSIM visual presentation of the two scenarios.**

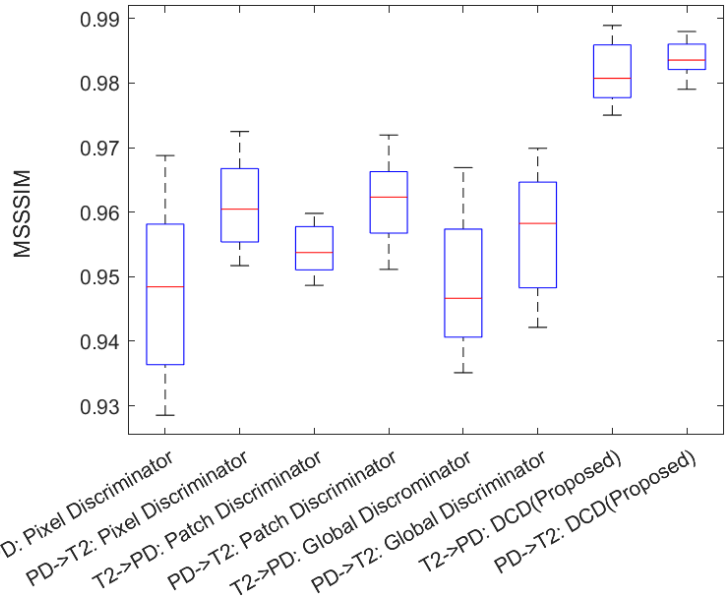


**Figure 4.19: Ablation study of different discriminators MSSSIM visual presentation of the two scenarios.**

Another important factor is the patch size of the training image. Smaller patch sizes are suitable for capturing local details and features in an image, such as edges, textures, etc. A medium-sized image patch provides a balance between capturing local detail and overall contextual information, which can better adapt to the characteristics of different scales, from smaller details to larger structures. A larger patch size helps to capture global contextual information such as the overall scene, object position, etc. For our model, we have done ablation studies for different patch sizes, which have been performed in the two scenarios as above. (1) Cross-modality synthesis of the T2-w image from the PD-w image. (2) Cross-modality synthesis of the PD-w image from the T2-w image. In the two scenarios, we trained our model with a patch size of 70×70, 128×128, and 256×256, respectively. The T2-w to PD-w synthesised results of different patch sizes are shown in Fig 4.20, and the PD-w to T2-w results are shown in Fig 4.21. The evaluation results of the different patch sizes are shown in Table 4.6, which justified the size of $128 \times 128$ to achieve better performance. To better visualise the

evaluation results, Fig 4.22, Fig 4.23, and Fig 4.24 show the PSNR, SSIM, and MSSSIM separately.

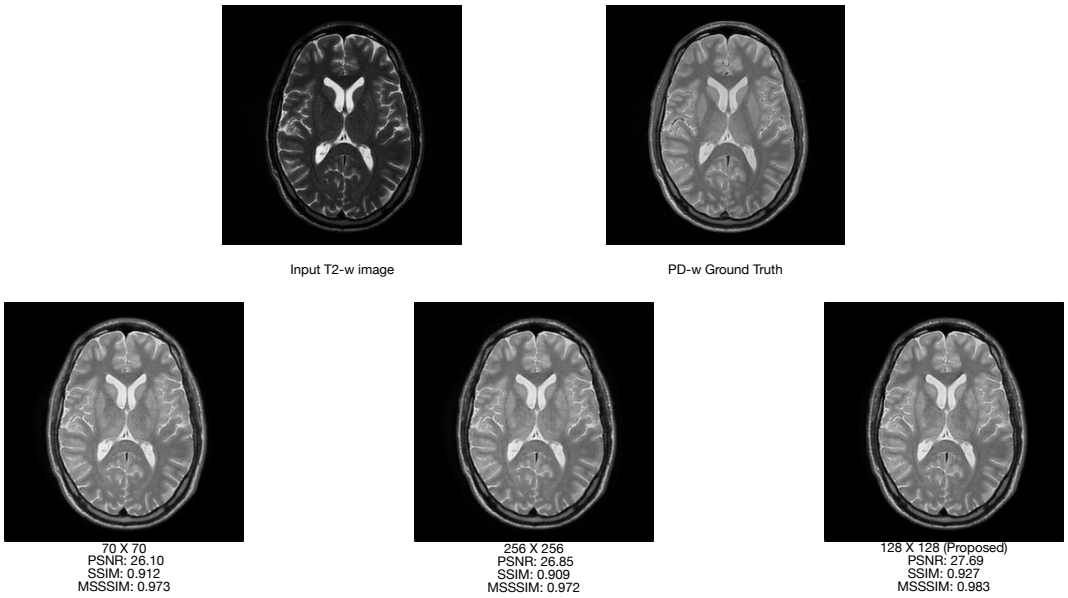


**Figure 4.20: Ablation study of different patch sizes results from T2-w → PD-w.**

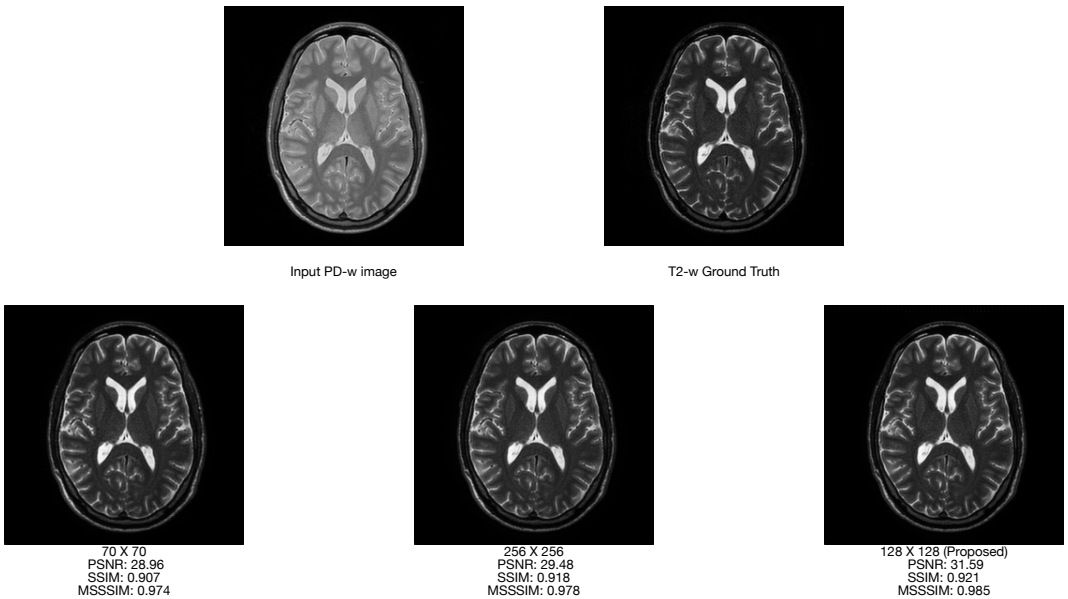


**Figure 4.21: Ablation study of different patch sizes results from PD-w → T2-w.**

| Metric (avg.) | 70 × 70 | 256 × 256 | 128 × 128 (Proposed) |
|---|---|---|---|
| **IXI: T2-w —> PD-w** | | | |
| PSNR (dB) | 26.10 | 26.85 | **27.69** |
| SSIM | 0.912 | 0.909 | **0.927** |
| MSSSIM | 0.973 | 0.972 | **0.983** |
| **IXI: PD-w —> T2-w** | | | |
| PSNR (dB) | 28.96 | 29.48 | **31.59** |
| SSIM | 0.907 | 0.918 | **0.921** |
| MSSSIM | 0.974 | 0.978 | **0.985** |

**Table 4.6: Ablation study of different patch sizes evaluation results of the two scenarios. The best results are marked in bold.**

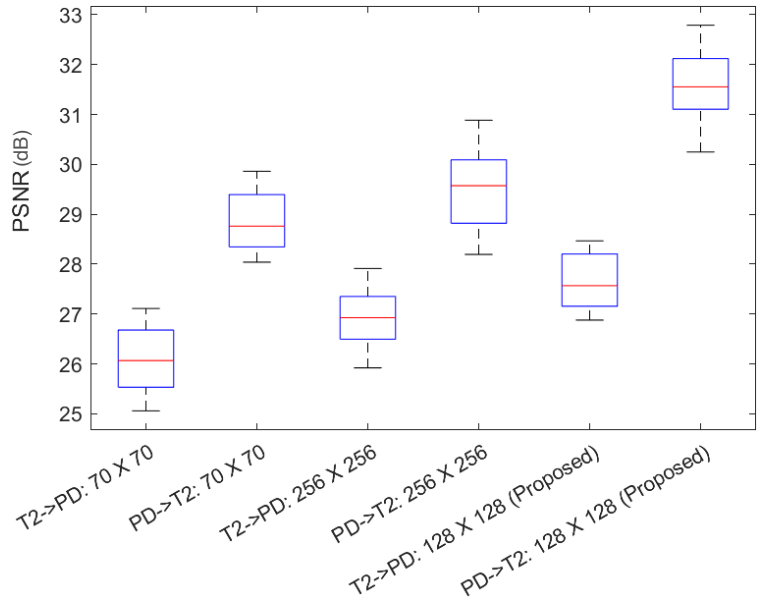


**Figure 4.22: Ablation study of different patch sizes PSNR visual presentation of the two scenarios.**

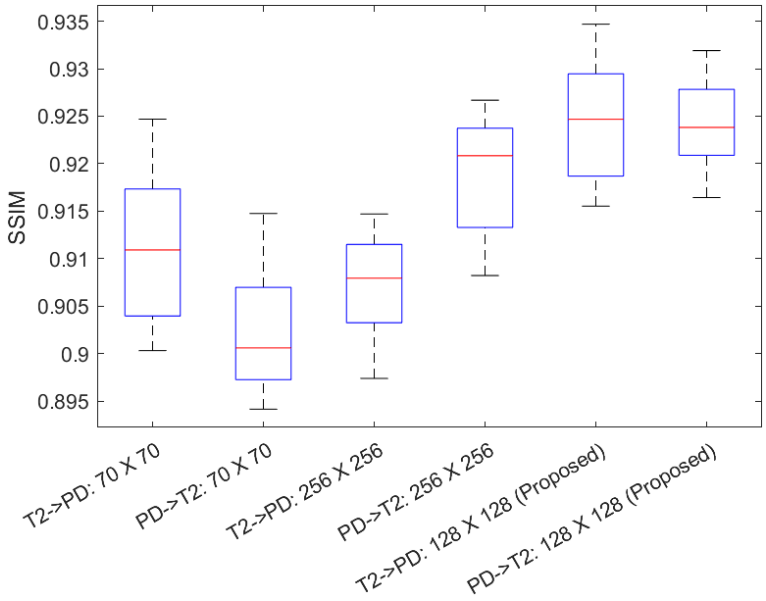


**Figure 4.23: Ablation study of different patch sizes SSIM visual presentation of the two scenarios.**

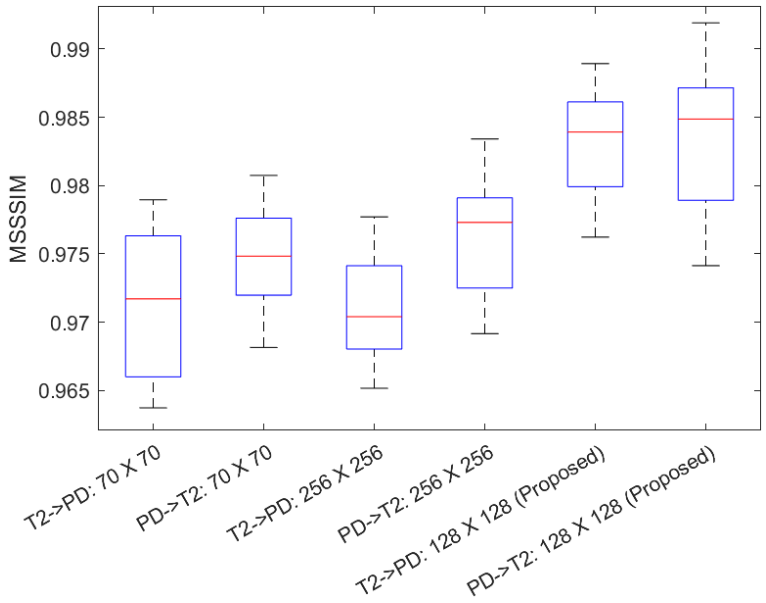


**Figure 4.24: Ablation study of different patch sizes MSSSIM visual presentation of the two scenarios.**

## 4.4 Conclusions

Drawing on the experimental results and analyses presented above, this section summarises the main findings and contributions of the proposed unsupervised MRI cross-modality synthesis approach. This chapter investigated unsupervised brain MRI cross-modality synthesis and proposed a generative adversarial framework equipped with a dual-channel joint discriminator. By jointly exploiting local structural information and pixel-level constraints, the proposed discriminator enhances the quality and fidelity of the synthesised images. Experimental results on benchmark datasets demonstrate that the proposed method achieves competitive or superior performance compared with existing state-of-the-art approaches. This study further illustrates that carefully designed adversarial supervision can effectively capture modality-specific representations, laying the groundwork for extending cross-modality synthesis beyond isolated tasks.

# Chapter 5

# Simultaneous Cross-Modality Synthesis and Denoising for Brain MRI

This chapter investigates the feasibility of jointly performing medical image denoising and cross-modality synthesis within a unified learning framework. By integrating the strengths of the proposed denoising and synthesis models, a dual-task network is developed to address both noise degradation and modality transformation simultaneously. Extensive experiments are conducted to demonstrate the effectiveness of the proposed approach.

MRI cross-modality synthesis and denoising are two distinct responsibilities in medical image processing. The process of removing noise from MRI is known as MRI denoising. The MRI scanning process can be affected by a variety of factors, of which noise is a common problem. MRI noise can come from several sources, including the instrument, environmental disturbances, and inhomogeneities in biological tissues, among others. Noises can affect the quality and readability of the MRI. Denoising techniques aim to

maintain structural information while minimising or removing noise. MRI cross-modality synthesis is the synthesis of data from one modality into another. In medical image processing, cross-modality synthesis is advantageous because some image modalities provide specific information while others provide different information. In practice, scanning costs, scanning time, and patient comfort present obstacles that limit the amount of collected data in clinical or research investigations. By synthesising MRI into multiple modalities, it is possible to obtain more comprehensive and detailed information. The objective of simultaneous MRI cross-modality synthesis and denoising is to improve the quality and information density of medical images. These two tasks can work in collaboration to provide more accurate and comprehensive information for medical image analysis and diagnosis. Simultaneous cross-modality synthesis and denoising can also contribute to automation and intelligence in medical image processing. Inspired by adversarial learning frameworks, we propose an unsupervised model that can deal with the two tasks simultaneously, which combines the multi-channel asymmetric residual generator and dual-channel joint discriminator. With task utility and visually high-fidelity synthesis both considered, the experimental evaluations on multiple cross-modality synthesis and denoising tasks show that the proposed CMS-DN GAN not only produces impressive visual results but also outperforms the related state-of-the-art methods. In general, simultaneous MRI cross-modality synthesis and denoising is a pioneering study direction in the field of medical image processing that is expected to result in substantial improvements in medical image quality and diagnostic accuracy.

## 5.1 Introduction

This section revisits the limitations of treating medical image denoising and cross-modality synthesis as independent tasks and motivates the need for a unified framework. The challenges of simultaneous learning are discussed, followed by an overview of the proposed dual-task strategy. Building upon the unsupervised denoising framework developed in Chapter 3 and the cross-modality synthesis approach introduced in Chapter 4, this chapter further explores a unified learning strategy that simultaneously addresses both tasks within a single network architecture.

Medical imaging is comprised of a variety of imaging modalities, such as computed tomography (CT), positron-emission tomography (PET), and magnetic resonance imaging (MRI). These modalities provide the opportunity to acquire knowledge regarding the features of various tissues through the utilisation of various physical acquisition principles or parameters. The concurrent availability of multi-modal imaging has contributed to an extensive variety of brain image analysis tasks. Despite the fact that MRI is capable of producing detailed images with high contrast of soft tissue, the long acquisition times remain unsolved. In particular, when scanning patients with Alzheimer's and Parkinson's, the scans are required to obtain the results in a shorter time, which leads to very poor-quality images. In order to solve these problems, researchers have made a great deal of effort. Among them, medical image cross-modality synthesis methods play an important role. Especially after the rise of GANs [45], more and more GAN-based methods are working well for the task of medical image cross-modality synthesis. In the previous review, we introduced a variety of GAN-based methods for medical image cross-modality synthesis. In addition to that, there are some other methods with remarkable contributions that are also worth mentioning. Yang *et al.* [157] proposed an unified hyper-GAN model that

encodes the source contrast image into a shared feature space and then decodes it again to the target contrast image. Using this method, it is possible to achieve the synthesis of multiple modalities. CycleGAN [178] has been widely used for image translation, and many medical image cross-modality synthesis models [56,68,84,149,154,156] have validated the performance of CycleGAN. In previous studies, CycleGAN outperformed other methods in medical image cross-modality synthesis tasks using unpaired training data. However, all of these studies only considered the problem of image cross-modality synthesis, and even when synthesising multiple different modalities, they still could not address the noise problem during image acquisition.

As previously stated, during the acquisition of medical images, the images are frequently accompanied by varying levels of noise. For instance, the thermal noise caused by the thermal motion of electronic components can introduce random noise during image acquisition and transmission. Thermal noise can significantly affect image quality at low signal levels. Electronic components in medical imaging equipment may generate electronic noise that is amplified during signal amplification, digitisation, and transmission. Electromagnetic interference from external environments and fluctuations in the power supply may affect the efficacy of the medical equipment, resulting in image noise. The movement of the patient during image acquisition may result in blurred or indistinct images, known as motion artifacts. Noise and artefacts may be introduced into medical images by motion artifacts. Variations in radiation dose may affect the signal intensity and image quality of radiological imaging techniques (e.g., X-ray imaging and nuclear medicine imaging). When the acquired signal strength is weak, noise can have a significant impact on the image quality, particularly in areas with low contrast. Inappropriate parameter settings for scanning, such as selecting weaker magnetic field strengths or shorter scanning durations, can result in an increase in noise. The use of obsolete or inferior

acquisition hardware may increase noise levels. During image transmission and storage, compression algorithms may introduce compression noise, notably at high compression ratios. The organisational structure of an organism may induce inhomogeneities in various regions, resulting in signal and noise variations. The effects of noise on medical images can be extensive and significant. The quality of medical images is critical for proper diagnosis and treatment decisions, and noise can lead to distortion of image information, illegible structures, and inaccurate analysis results. Noise blurs the image's structures and details, diminishing the image's clarity. Physicians may struggle to differentiate between essential anatomical structures and lesions. Noise reduces an image's signal-to-noise ratio, thereby diminishing the contrast between various tissues. This may make it challenging for the physician to differentiate between various tissues or lesions. Image signal and noise can interact to produce artifacts. This could make it difficult to identify anatomical structures in the image. Noise can obscure or conceal lesions in an image, which can lead to misdiagnosis or underdiagnosis. In quantitative analyses, noise can contribute to inaccuracies in measurements, which can affect volume, area, and intensity measurements. In computer-aided analysis and diagnosis, noise can make it challenging for algorithms to detect and localise lesions precisely. Image processing tasks such as image alignment, segmentation, and feature extraction may be hindered by noise, resulting in inaccurate results. Therefore, a significant amount of research has been devoted to medical image denoising. Combined with the methods mentioned in Chapters 2 and 3, it can be summarised into the following categories: filtering-based and frequency domain-based methods [4, 10, 27, 39, 51, 177], deep learning-based methods [36, 38, 44, 76, 87, 118, 132, 171], GAN-based methods [6, 22, 104, 144, 159], and transformer-based methods [19, 161, 174]. From the previous presentation, it is clear that either the absence of medical image modalities or the acquisition of images contaminated with noise can have a negative impact

on medical diagnosis. However, these two problems are common in the process of acquiring medical images due to various influencing factors. For this reason, we proposed a model that can perform cross-modality synthesis and denoising simultaneously, thus avoiding the image quality degradation that occurs when dealing with these two problems separately. Cross-modality synthesis generates missing modality information and improves the functional and anatomical structure of the medical image; denoising reduces image noise and restores image detail and contrast. By performing both tasks simultaneously, image quality can be further enhanced.

## 5.2 Method

To formally describe the proposed simultaneous denoising and cross-modality synthesis framework, this section introduces the overall network design, problem formulation, and optimisation objectives for joint learning.

In this chapter, we propose an unsupervised learning network architecture for the simultaneous processing of cross-modality synthesis and denoising tasks for medical images. In chapter 3, we demonstrated the use of the innovative MARG for denoising medical images. In chapter 4, we utilised the novel DCD to perform cross-modality synthesis of medical images. Since these two tasks are completely different but essential in the preprocessing of medical images, it is promising for practical use if these two problems can be solved simultaneously with minimal loss of image quality. So we creatively combined the MARG and DCD together, as well as the addition of the loss function for performing these two tasks simultaneously. The remarkable results have been obtained after a significant number of experiments and modifications to the network structure. Comparison with a newly designed validation process shows the effectiveness and robustness of this innovative structure in handling the two tasks simultaneously.

### 5.2.1 Preliminaries

This subsection briefly reviews the key concepts underlying the proposed framework, including GAN-based image translation, unsupervised learning strategies, and dual-task optimisation.

Based on the importance of GANs in generative tasks, UNIT [86], MUNIT [59], and CycleGAN [178] play an important role in image translation tasks, enabling high-quality image translation between different domains. Based on these important GAN variants and using their network construction ideas, we have also introduced a network structure with a dual-generator and dual-discriminator structure that utilises cycle consistency loss to constrain the image translation process, thus achieving the purpose of simultaneously handling the dual tasks of cross-modality synthesis and denoising of medical images.

### 5.2.2 Problem Formulation

Building upon the individual formulations of denoising and cross-modality synthesis introduced in previous chapters, this subsection formulates the simultaneous dual-task learning problem, defining the input domains, output targets, and bidirectional mappings.

The aim of our proposed method is to solve the tasks of medical image cross-modality synthesis and denoising while minimising the loss of image quality and structure information. The tasks can be seen as an image-to-image translation. Unlike dealing with a single cross-modality synthesis task or a single denoising task, dealing with both tasks simultaneously is equivalent to spatially spanning a greater distance to deal with the translation of two more disparate modalities. In terms of the formula, $\boldsymbol{I}_N^T \in \mathbb{R}^{i \times j \times k}$

expresses the image in the first modality with noise, $\boldsymbol{I}^T \in \mathbb{R}^{i \times j \times k}$ represents the image in the first modality without noise, $\boldsymbol{I}_N^P \in \mathbb{R}^{i \times j \times k}$ denotes that the image is in the second modality with noise, and $\boldsymbol{I}^P \in \mathbb{R}^{i \times j \times k}$ indicates that the image is in the second modality without noise. Where $N$ means the added noise, $T$ refers to the first modality, $P$ means the second modality, $i$ and $j$ are the dimensions of the volumetric's axial view, and $k$ represents the dimensions along the z-axis. $\boldsymbol{I}_N^T$, $\boldsymbol{I}^T$, $\boldsymbol{I}_N^P$, and $\boldsymbol{I}^P$ can be considered as different modalities. The generator for the single denoising task can be expressed as $G_D$, and the generator for single cross-modality synthesis can be written as $G_C$. Due to the use of dual generator and dual discriminator in our model, we write the generator as $G_{C\&D}$, and the reconstruction generator as $F_{C\&D}$. The corresponding discriminators are $D_{G_{C\&D}}$ and $D_{F_{C\&D}}$, respectively. Therefore, for a single denoising task, the process can be expressed as:

$$\boldsymbol{I}^T = G_D(\boldsymbol{I}_N^T), \tag{5.1}$$

where $G_D$ stands for a trained denoising generator. The procedure for a single cross-modality task can be expressed as:

$$\boldsymbol{I}^P = G_C(\boldsymbol{I}^T), \tag{5.2}$$

where $G_C$ means a trained cross-modality generator. Therefore, the cross-modality synthesis and denoising task using conventional methods can be represented as:

$$\boldsymbol{I}^P = G_D(G_C(\boldsymbol{I}_N^T)), \tag{5.3}$$

or

$$\boldsymbol{I}^P = G_C(G_D(\boldsymbol{I}_N^T)), \tag{5.4}$$

depending on the sequence of processing tasks. The simultaneous cross-modality synthesis and denoising procedure can be written as:

$$\boldsymbol{I}^P = G_{C\&D}(\boldsymbol{I}_N^T). \tag{5.5}$$

### 5.2.3 Loss Function

Following the problem formulation, this subsection defines the combined loss functions used to optimise the proposed joint model, including adversarial loss, cycle-consistency constraints, denoising-related losses, and task-balancing terms.

In accordance with the preceding procedure, adversarial loss is applied to our model, which is founded on a generator adversarial network. In conjunction with the aforementioned illustration, the total adversarial loss within this framework can be denoted as:

$$\begin{aligned}\mathcal{L}_{adv} =& \mathbb{E}_{\boldsymbol{I}^P \sim p_{\text{data}}(\boldsymbol{I}^P)}[\log D_{G_{C\&D}}(\boldsymbol{I}^P)] + \mathbb{E}_{\boldsymbol{I}_N^T \sim p_{\text{data}}(\boldsymbol{I}_N^T)}[\log(1 - D_{G_{C\&D}}(G_{C\&D}(\boldsymbol{I}_N^T)))] + \\ & \mathbb{E}_{\boldsymbol{I}^T \sim p_{\text{data}}(\boldsymbol{I}^T)}[\log D_{F_{C\&D}}(\boldsymbol{I}^T)] + \mathbb{E}_{\boldsymbol{I}_N^P \sim p_{\text{data}}(\boldsymbol{I}_N^P)}[\log(1 - D_{F_{C\&D}}(F_{C\&D}(\boldsymbol{I}_N^P)))].\end{aligned} \tag{5.6}$$

In the meantime, the cycle consistency loss is implemented to ensure the image maintains its consistency during the procedure, which can be expressed as:

$$\mathcal{L}_{cyc} = \mathbb{E}_{\boldsymbol{I}^T \sim p_{\text{data}}(\boldsymbol{I}^T)} \| \boldsymbol{I}^T - F_{C\&D}(G_{C\&D}(\boldsymbol{I}_N^P)) \|_1 + \mathbb{E}_{\boldsymbol{I}^P \sim p_{\text{data}}(\boldsymbol{I}^P)} \| \boldsymbol{I}^P - G_{C\&D}(F_{C\&D}(\boldsymbol{I}_N^T)) \|_1 . \tag{5.7}$$

Since we perform the tasks of cross-modality synthesis and denoising simultaneously, we have to ensure both that the structural information of the medical image is not lost when performing the denoising and that there is an accurate representation between pixels when performing cross-modality synthesis. So we need to limit the structural differences between the denoised image and the noisy image, as well as the difference between the pixel representations after cross-modality synthesis and the original modality image. Therefore, we use the following loss function to constrain the structural similarity of the noise image and denoised image:

$$\mathcal{L}_s = \mathbb{E}_{(\boldsymbol{I}_N^T, \boldsymbol{I}^T)} \| \boldsymbol{I}_N^T - \boldsymbol{I}_{synthesised}^T \|_2^2 + \mathbb{E}_{(\boldsymbol{I}_N^P, \boldsymbol{I}^P)} \| \boldsymbol{I}_N^P - \boldsymbol{I}_{synthesised}^P \|_2^2, \tag{5.8}$$

and the following loss function is utilised to constrain the pixel representation between the cross-modality synthesised image and its original modality image:

$$\mathcal{L}_p = \mathbb{E}_{(\boldsymbol{I}_N^T, \boldsymbol{I}_N^P)} \| \boldsymbol{I}_N^T - \boldsymbol{I}_N^P \|_F^2 + \mathbb{E}_{(\boldsymbol{I}^T, \boldsymbol{I}^P)} \| \boldsymbol{I}_{synthesised}^T - \boldsymbol{I}_{synthesised}^P \|_F^2, \tag{5.9}$$

where $\| \|_F$ means the Frobenius norm.

Therefore, the final objective of the proposed model is:

$$\mathcal{L} = \lambda_{adv}\mathcal{L}_{adv} + \lambda_{cyc}\mathcal{L}_{cyc} + \lambda_p\mathcal{L}_p + \lambda_s\mathcal{L}_s, \tag{5.10}$$

where $\lambda_{adv}, \lambda_{cyc}, \lambda_p, \lambda_s$ represent the weights of $\mathcal{L}_{adv}, \mathcal{L}_{cyc}, \mathcal{L}_p, \mathcal{L}_s$ respectively.

## 5.3 Experiments

To evaluate the effectiveness of the proposed simultaneous learning framework, this section presents the experimental setup, network architectures, and comprehensive analyses for the dual-task scenario.

### 5.3.1 Network Structure

This subsection describes the architecture of the proposed joint model, detailing how the generator and discriminator components are shared or integrated to support both denoising and cross-modality synthesis.

As demonstrated in chapters 3 and 4, the MARG structure has achieved excellent performance in the task of medical image denoising, and the DCD structure has shown promising results in the medical image cross-modality synthesis task. Therefore, when performing simultaneous cross-modality synthesis and denoising tasks, we have combined the advantages of the two structures and created a new network architecture. As shown in Fig 5.1, firstly, the T2-w image with noise was taken as input, and the cross-modality synthesised and denoised PD-w image was generated by a CMS-DN MARG that was modified in order to perform simultaneous cross-modality synthesis and denoising tasks. At this point, the noise-free PD-w ground truth and the generated denoised PD-w image are input to CMS-DN DCD that has been modified to perform the cross-modality synthesis and denoising tasks simultaneously. The resulting discriminative results are used to give feedback to the generator to improve the quality of the generated results in the next iteration. Up to now, we have completely described the generative part of the model, and the obtained cross-modality synthesised and denoised results are the objective. In order to achieve the unsupervised learning property, we proceed to the next stage, where the previously gener-

ated denoised PD-w image is used as input and another CMS-DN MARG is used to reconstruct the original input. The reconstructed T2-w image with noise is obtained. After that, another CMS-DN DCD is used to judge the truthfulness of the reconstructed image in conjunction with the T2-w with noise ground truth, thus providing feedback to improve the performance of the generator used for reconstruction. By utilising the idea of a cyclic network, the cycle consistency loss is used to constrain the distance between the reconstructed image and the original image so that the reconstructed image is as close as possible to the original input.

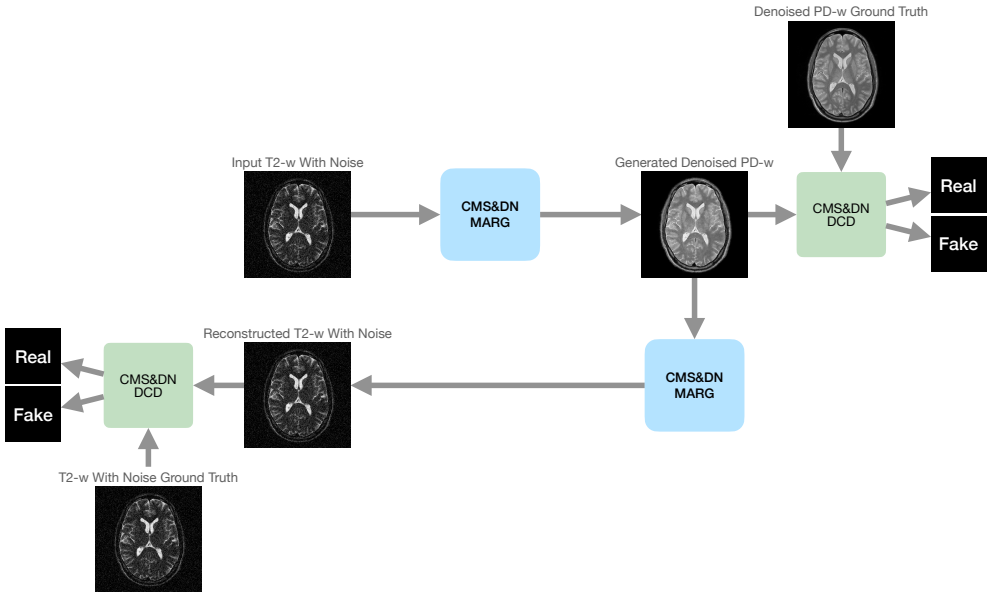


**Figure 5.1: The structure of the CMS-DN network.**

The CMS-DN MARG and CMS-DN DCD will be described in detail next. Firstly, for the generator, we modified the structure of the MARG accordingly to suit the tasks of simultaneous medical image cross-modality synthesis and denoising. Fig 5.2 shows the structure of CMS-DN MARG, the initial block stack 3 convolutional structures. The first convolution layer has $7 \times 7$ kernels with strides 1. The second convolution layer has $5 \times 5$ kernels with strides 2. The third convolution layer uses $3 \times 3$ kernels with

strides 2. By stacking three convolutional layers with different perception fields, it allows the network to learn a wide range of different local features that can be stacked together to build more complex feature representations, which helps to improve the performance of simultaneous multitasking models. The convolutional structure has been formed as Convolution-Batchnormalisation-ReLU. Then 4 CMS-DN MAR blocks stack linearly. The subsequent part follows the MARG in Chapter 3.

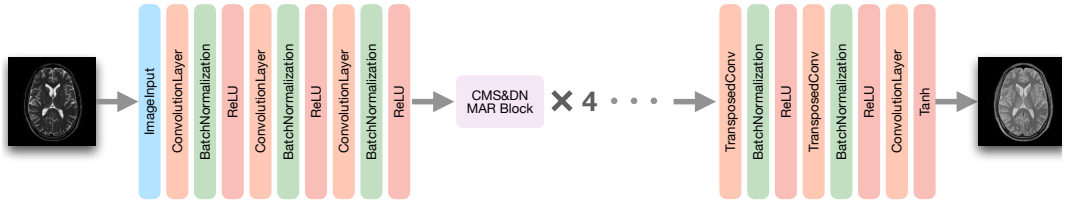


**Figure 5.2: The generator structure of CMS-DN network.**

Fig 5.3 shows the CMS-DN MAR blocks. Unlike the MAR Blocks, the CMS-DN MAR Blocks reduced a convolutional block for each channel. The proposed network combines both MARG and DCD; the disadvantage is that it increases the complexity of the network and the training time. In order to reduce the weight of the network, through a great deal of experimentation and evaluation, we made appropriate changes without reducing the performance of the model.

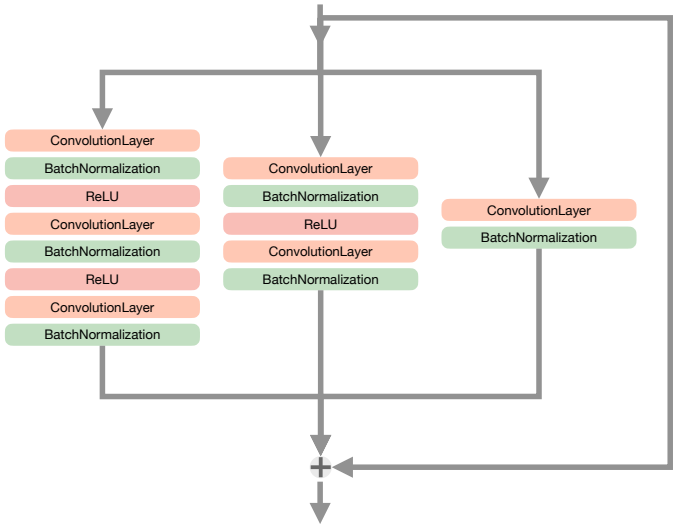


**Figure 5.3: The structure of CMS-DN MAR Block**

The discriminator utilised the dual-channel discriminator structure with some improvements for simultaneous medical image cross-modality synthesis and denoising tasks (CMS-DN DCD). As shown in Fig 5.4, the PatchGAN Discriminator Channel provides local structure feedback to the generator, and the Pixel Discriminator Channel provides pixel information feedback to the generator. The generator is encouraged to generate more realistic images through the feedback from the patchGAN discriminator, while the enhancement of edge information and reduction of artefacts are achieved through the feedback from the pixel discriminator. In the proposed model, the generator has a more complex structure compared to the generator for single medical image cross-modality synthesis task. In other words, the generator is enhanced, but the discriminator is not sufficient to give stricter judgement during iterative training. The experimental results show that if the same discriminator structure continues to be used, it will result in the edges of the image appearing blurred. Therefore, we enhance the performance of the pixel discriminator by adding an intermediate layer, which consists of a convolution layer with a stride of (1, 1), a batch normalisation

layer, and a leaky ReLU layer. Experimental results confirm the effectiveness of the structure, which will be shown in the results section.

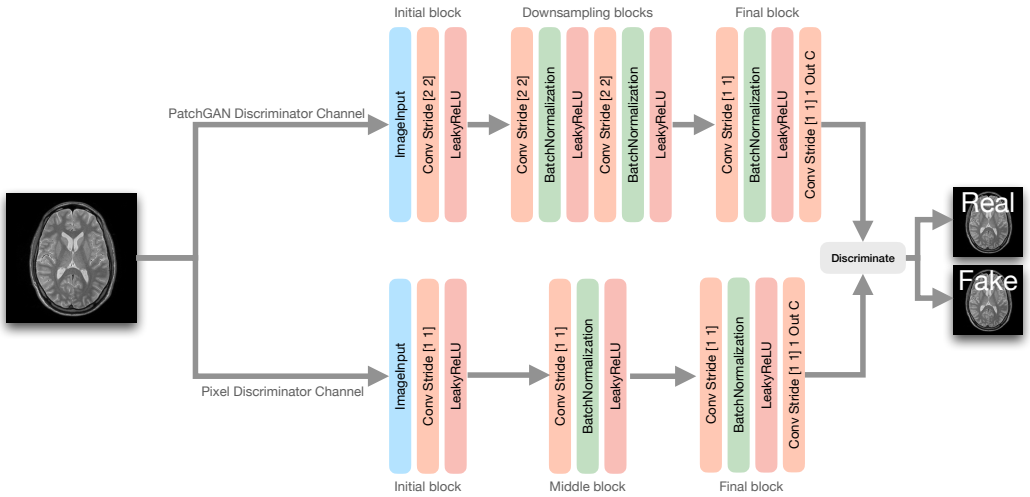


**Figure 5.4: The structure of CMS-DN DCD**

### 5.3.2 Experimental Setup

In accordance with the described network architecture, this subsection details the datasets, training configuration, and evaluation protocols used in the experiments.

#### Experimental Flow Design

Since our method is the first to propose cross-modality synthesis and denoising simultaneously for medical images based on unsupervised learning, there is no existing baseline method to compare it with, so we designed a new experimental process for the comparison.

Cross-modality synthesis and denoising of medical images are two dif-

ferent tasks. As shown in Fig 5.5, we set up our experiments into three different routes when designing the experimental flow, which are A → B → D, A → C → D, and A → D. Among them, route A → B → D is to perform the denoising task from A → B first, and then conduct the cross-modality synthesis task from B → D. Route A → C → D is to perform the cross-modality synthesis task from A → C followed by the denoising task from C → D. Route A → D then performs simultaneous cross-modality synthesis and denoising with our proposed model.

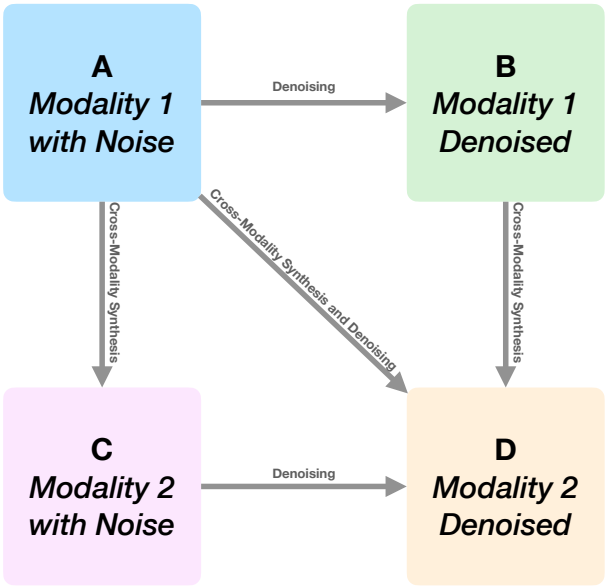


**Figure 5.5: The Designed Experimental Flow**

For the denoising task from A → B and from C → D, we use the baseline method used in Chapter 3 and select the model with the best results for training. The difference is that the inputs from A → B are Modality 1, while the inputs from C → D are Modality 2. For the cross-modality synthesis tasks from A → C and from B → D, we select the model that yields the best results from the baseline methods mentioned in Chapter 4 for training. Where the inputs of A → C are noised images and the inputs of B → D are noiseless images.

After all the models have been trained, the testing process is that for flow A $\rightarrow$ B $\rightarrow$ D, use modality 1 with noise image as input into the trained A $\rightarrow$ B denoising model, and then the denoised result is obtained. After that, the result is used as the input to the cross-modality synthesis model B $\rightarrow$ D, and finally, the result is obtained with denoising first and then cross-modality synthesis. For the testing of the flow A $\rightarrow$ C $\rightarrow$ D, the image of Modality 1 with noise is input into the trained A $\rightarrow$ C cross-modality synthesis model, then the cross-modality synthesised image is obtained, which is then used as the input to the denoising model C $\rightarrow$ D, and finally the result is obtained where cross-modality synthesis is performed first and denoising is carried out later. The results obtained from these two processes are then compared with the results obtained from our proposed method.

**Training Details**

The initial learning rate is set to $1e-4$, and the batch size is set to 8. The Adam solver [72] is used as an optimizer. The gradient decay factor is set to 0.9, and the squared gradient decay factor is set to 0.99. To avoid division by zero, the denominator offset is $1e-8$. We augment the training data through rotation, stretching, and reflection. Crop 16 patches from each training image by size $128 \times 128$. Also, shuffle the training data before each training epoch, and shuffle the validation data before each validation. The parameters for the final objective are set to $\lambda_{adv} = 1, \lambda_{cyc} = 10, \lambda_p = 2$, and $\lambda_s = 2$. Train 200 epochs to obtain the final model.

**Qualitative Evaluations**

The following figure shows the results of cross-modality synthesis and denoising of brain MRI on the IXI dataset using the experimental procedure designed above. According to the different noise variances, we divided the

experiments into four groups with variance 0.01, variance 0.0075, variance 0.005, and variance 0.0025. Analyse them in two scenarios. The experimental results demonstrate that our proposed method outperforms state-of-the-art methods that perform cross-modality synthesis and denoising separately for different variances as well as for different modality transitions.

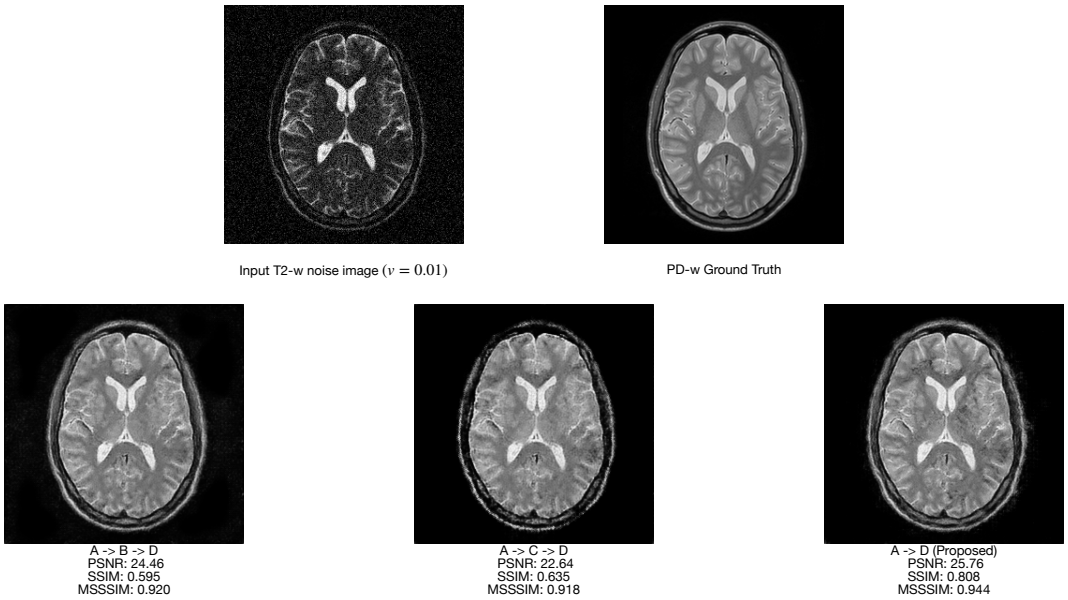


**Figure 5.6: Visual comparisons of denoising first then cross-modality synthesis (A $\rightarrow$ B $\rightarrow$ D), cross-modality first then denoising (A $\rightarrow$ C $\rightarrow$ D), and simultaneous cross-modality synthesis and denoising (A $\rightarrow$ D) for T2-w $\rightarrow$ PD-w, with the variance** ($v = 0.01$)

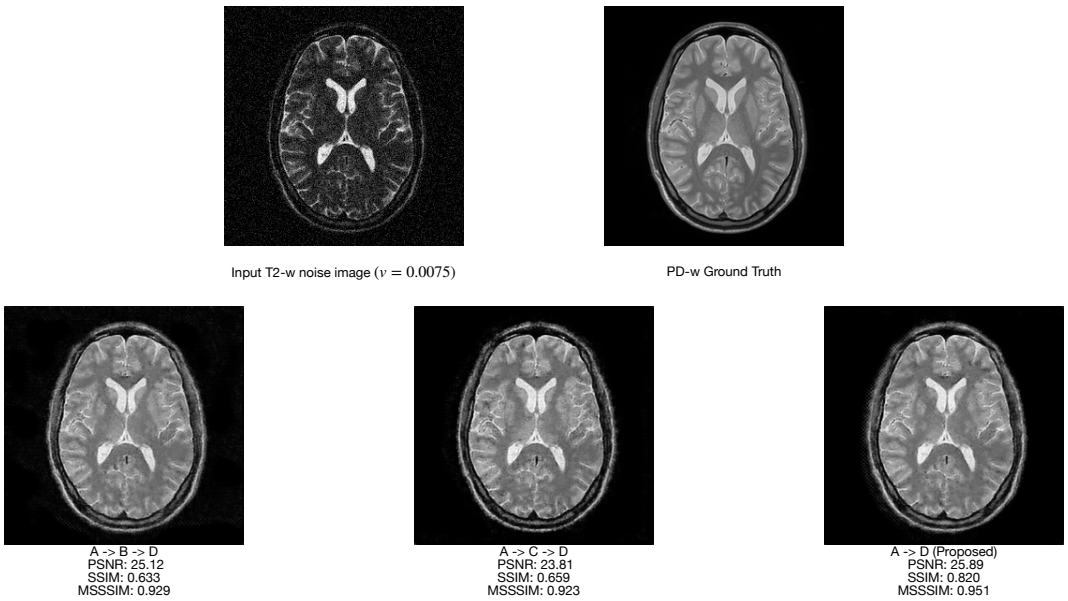


**Figure 5.7: Visual comparisons of denoising first then cross-modality synthesis (A → B → D), cross-modality first then denoising (A → C → D), and simultaneous cross-modality synthesis and denoising (A → D) for T2-w → PD-w, with the variance** ($v = 0.0075$)

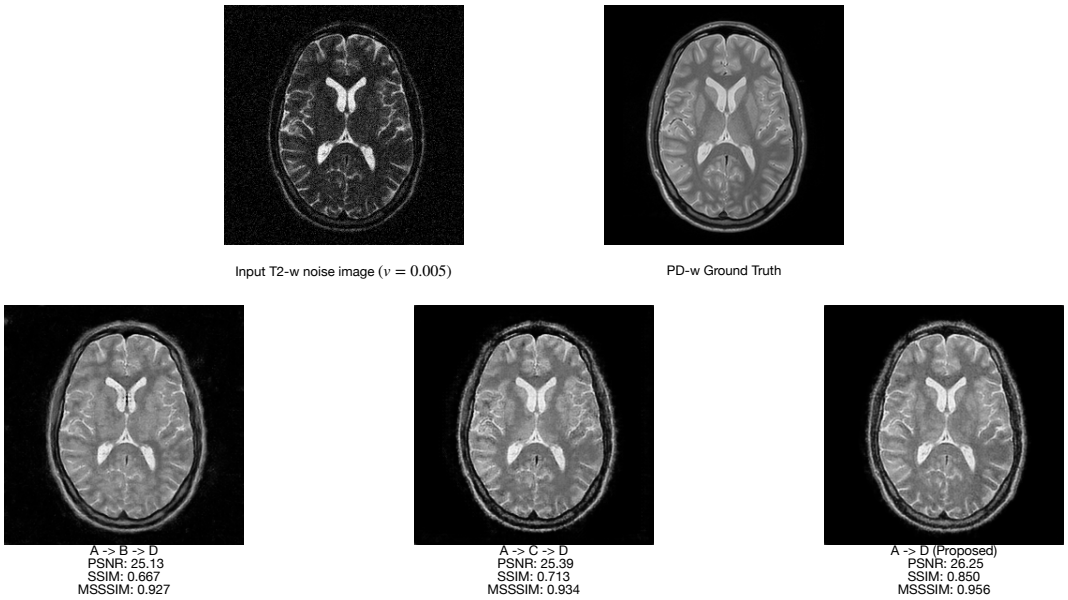


**Figure 5.8: Visual comparisons of denoising first then cross-modality synthesis (A → B → D), cross-modality first then denoising (A → C → D), and simultaneous cross-modality synthesis and denoising (A → D) for T2-w → PD-w, with the variance** ($v = 0.005$)

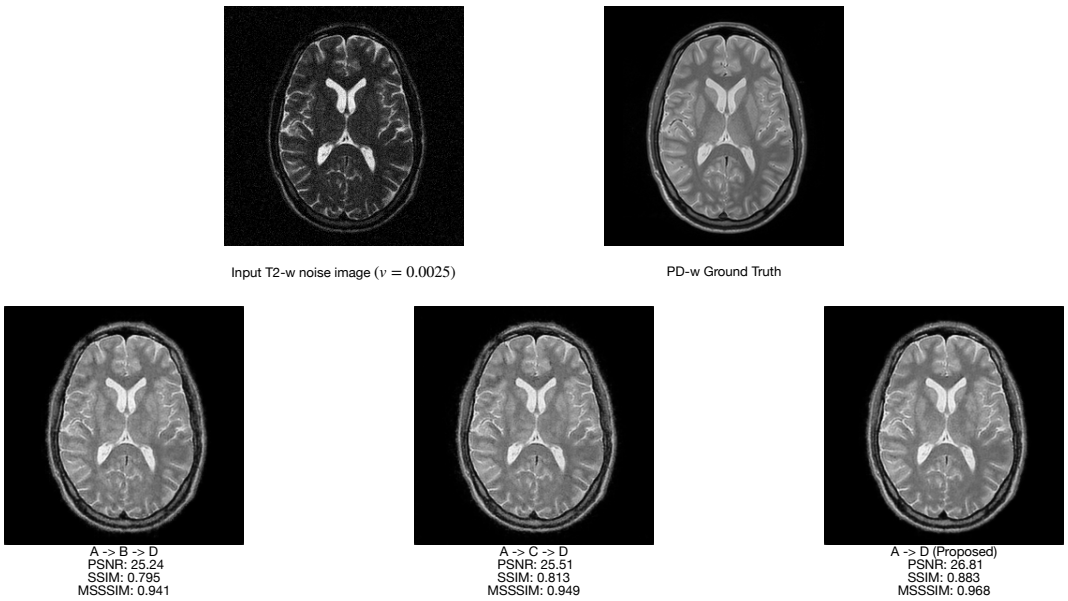


**Figure 5.9: Visual comparisons of denoising first then cross-modality synthesis (A → B → D), cross-modality first then denoising (A → C → D), and simultaneous cross-modality synthesis and denoising (A → D) for T2-w → PD-w, with the variance** $(v = 0.0025)$

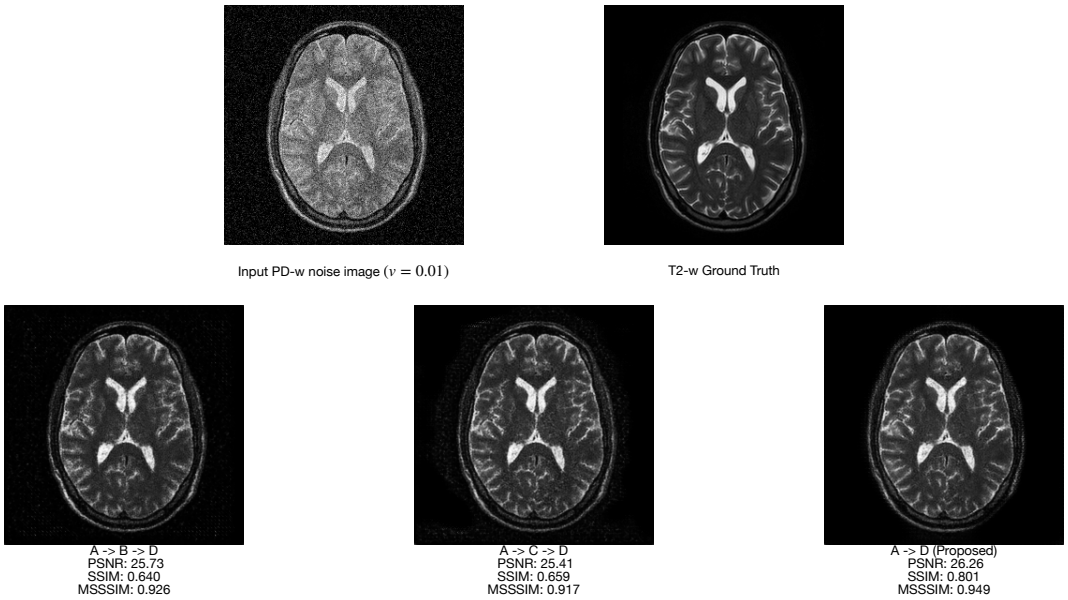


**Figure 5.10: Visual comparisons of denoising first then cross-modality synthesis (A → B → D), cross-modality first then denoising (A → C → D), and simultaneous cross-modality synthesis and denoising (A → D) for PD-w → T2-w, with the variance** $(v = 0.01)$

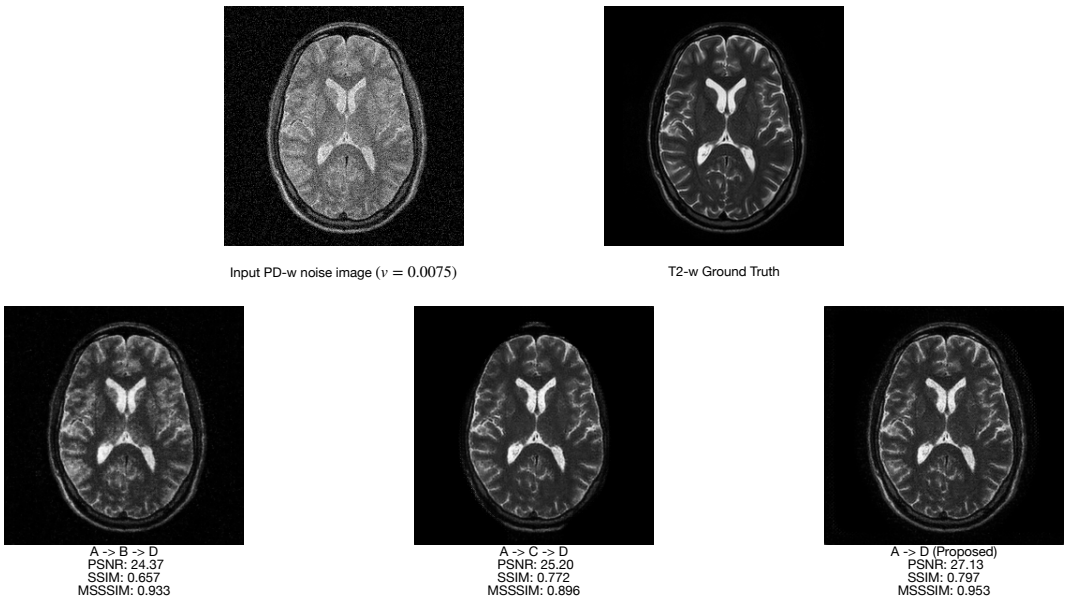


**Figure 5.11: Visual comparisons of denoising first then cross-modality synthesis (A → B → D), cross-modality first then denoising (A → C → D), and simultaneous cross-modality synthesis and denoising (A → D) for PD-w → T2-w, with the variance** ($v = 0.0075$)

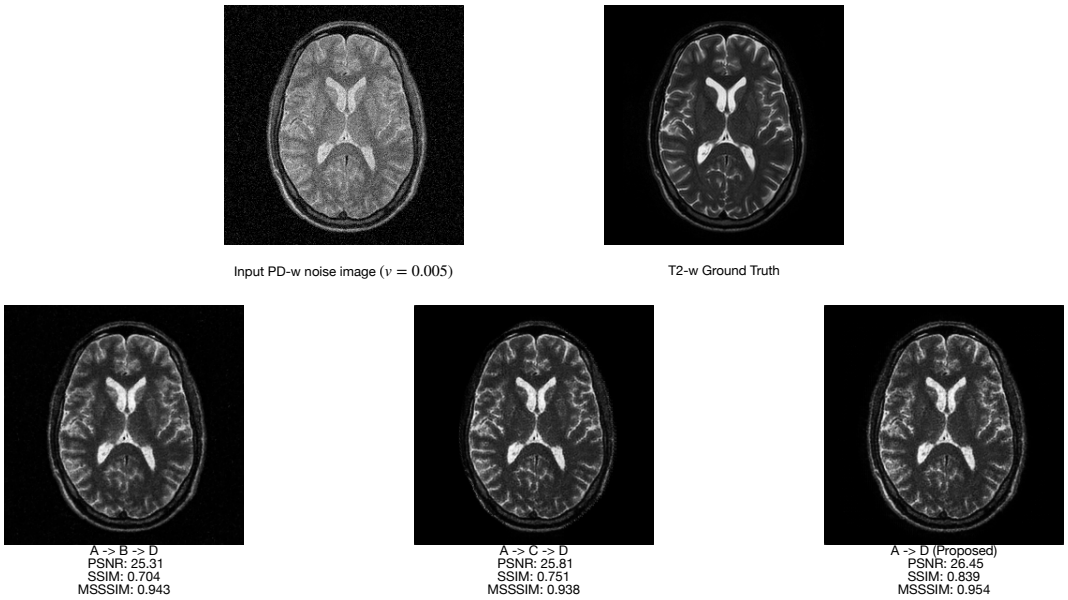


**Figure 5.12: Visual comparisons of denoising first then cross-modality synthesis (A → B → D), cross-modality first then denoising (A → C → D), and simultaneous cross-modality synthesis and denoising (A → D) for PD-w → T2-w, with the variance** ($v = 0.005$)

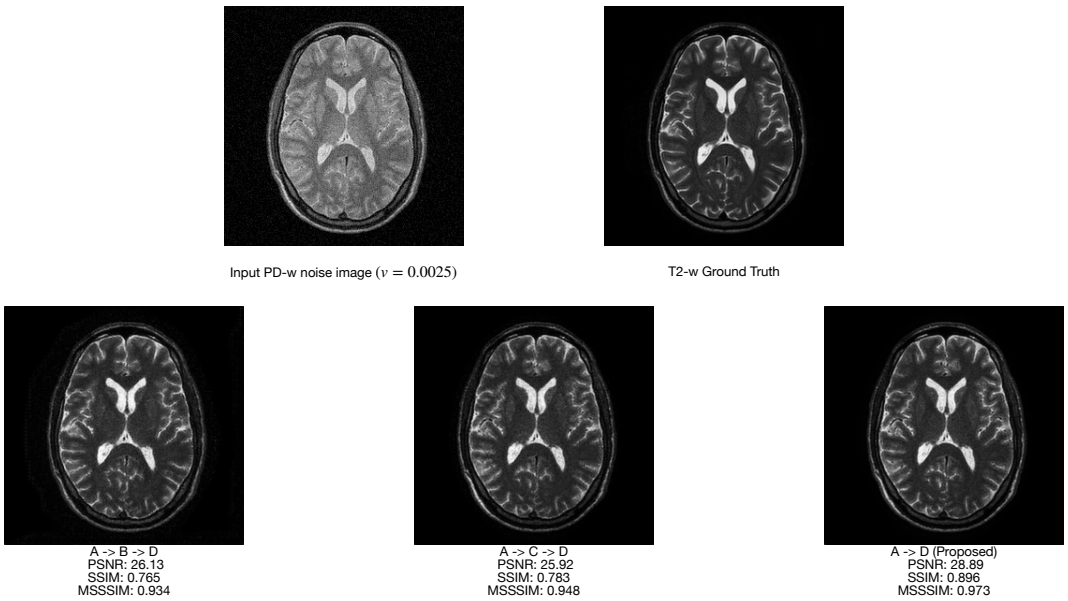


**Figure 5.13: Visual comparisons of denoising first then cross-modality synthesis (A $\rightarrow$ B $\rightarrow$ D), cross-modality first then denoising (A $\rightarrow$ C $\rightarrow$ D), and simultaneous cross-modality synthesis and denoising (A $\rightarrow$ D) for PD-w $\rightarrow$ T2-w, with the variance** ($v = 0.0025$)

### Quantitative Evaluations

We use PSNR, SSIM, and MSSSIM as evaluation criteria to objectively assess the quality of the synthesis results. Table 5.1 shows the results of synthesis from T2-w to PD-w with different variances, and Table 5.2 shows the results of synthesis from PD-w to T2-w with different variances, respectively. In terms of evaluation criteria, the proposed method of simultaneous cross-modality synthesis and denoising obtained better results, both in terms of different noise variance levels and transitions between modalities. When the cross-modality synthesis and denoising tasks are performed in steps, more losses are incurred, leading to worse results. Whether cross-modality synthesis is followed by denoising or denoising is followed by cross-modality synthesis, even with state-of-the-art models, our proposed simultaneous cross-modality synthesis and denoising model remains the

| Metric (avg.) | A -> B -> D | A -> C -> D | A -> D (Proposed) |
|---|---|---|---|
| IXI: T2-w —> PD-w ($\mathcal{V}$ = 0.01) | | | |
| PSNR (dB) | 24.46 | 22.64 | **25.76** |
| SSIM | 0.595 | 0.635 | **0.808** |
| MSSSIM | 0.920 | 0.918 | **0.944** |
| IXI: T2-w —> PD-w ($\mathcal{V}$ = 0.0075) | | | |
| PSNR (dB) | 25.12 | 23.81 | **25.89** |
| SSIM | 0.633 | 0.659 | **0.820** |
| MSSSIM | 0.929 | 0.923 | **0.951** |
| IXI: T2-w —> PD-w ($\mathcal{V}$ = 0.005) | | | |
| PSNR (dB) | 25.13 | 25.39 | **26.25** |
| SSIM | 0.667 | 0.713 | **0.850** |
| MSSSIM | 0.927 | 0.934 | **0.956** |
| IXI: T2-w —> PD-w ($\mathcal{V}$ = 0.0025) | | | |
| PSNR (dB) | 25.24 | 25.51 | **26.81** |
| SSIM | 0.795 | 0.813 | **0.883** |
| MSSSIM | 0.941 | 0.949 | **0.968** |

**Table 5.1: Quantitative evaluations of the process A → B → D (denoising first then cross-modality synthesis), A → C → D (cross-modality first then denoising), and A → D (simultaneous cross-modality synthesis and denoising) for T2-w → PD-w with 4 different variance groups.**

highest performance.

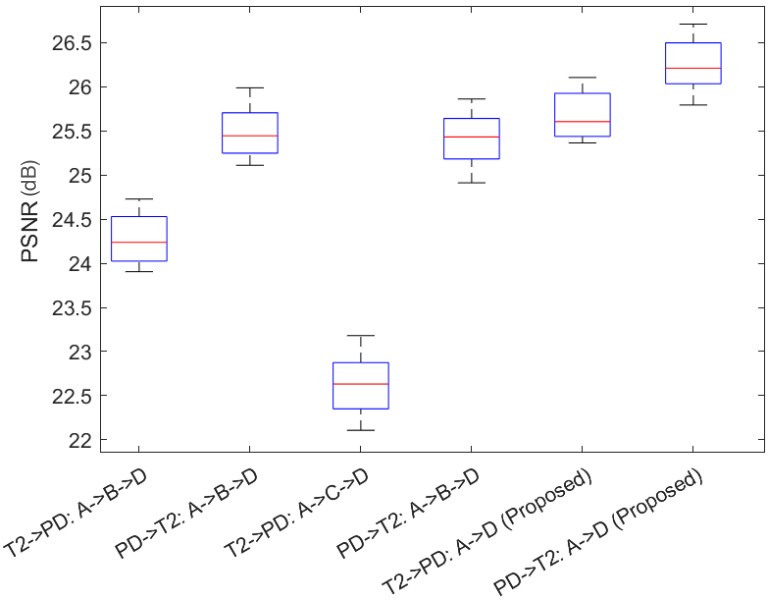


**Figure 5.14: Visual presentation of PSNR for the process A → B → D , A → C → D, and A → D for both T2-w → PD-w and PD-w → T2-w.** ($v = 0.01$)

| Metric (avg.) | A -> B -> D | A -> C -> D | A -> D (Proposed) |
|---|---|---|---|
| IXI: PD-w —> T2-w ($\nu$ = 0.01) | | | |
| PSNR (dB) | 25.73 | 25.41 | **26.26** |
| SSIM | 0.640 | 0.659 | **0.801** |
| MSSSIM | 0.926 | 0.917 | **0.949** |
| IXI: PD-w —> T2-w ($\nu$ = 0.0075) | | | |
| PSNR (dB) | 24.37 | 25.20 | **27.13** |
| SSIM | 0.657 | 0.772 | **0.797** |
| MSSSIM | 0.933 | 0.896 | **0.953** |
| IXI: PD-w —> T2-w ($\nu$ = 0.005) | | | |
| PSNR (dB) | 25.31 | 25.81 | **26.45** |
| SSIM | 0.704 | 0.751 | **0.839** |
| MSSSIM | 0.943 | 0.938 | **0.954** |
| IXI: PD-w —> T2-w ($\nu$ = 0.0025) | | | |
| PSNR (dB) | 26.13 | 25.92 | **28.89** |
| SSIM | 0.765 | 0.783 | **0.896** |
| MSSSIM | 0.934 | 0.948 | **0.973** |

**Table 5.2: Quantitative evaluations of the process A → B → D (denoising first then cross-modality synthesis), A → C → D (cross-modality first then denoising), and A → D (simultaneous cross-modality synthesis and denoising) for PD-w → T2-w with 4 different variance groups.**

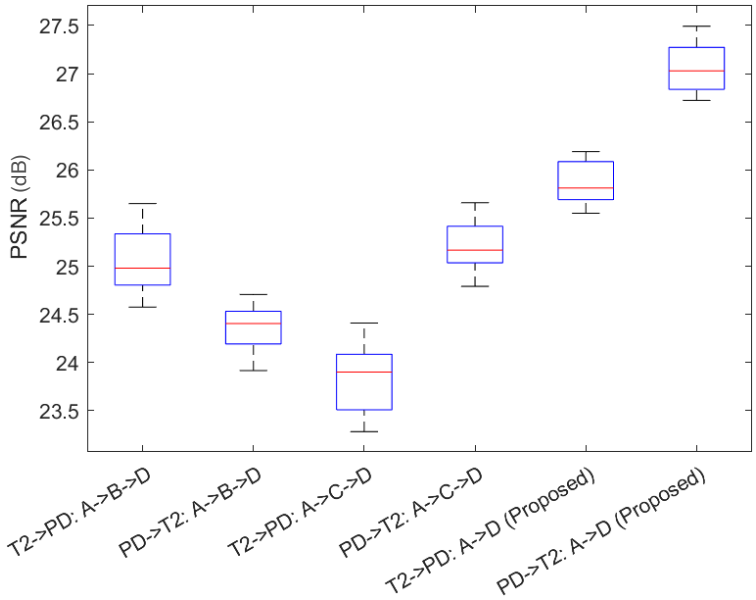


**Figure 5.15: Visual presentation of PSNR for the process A → B → D, A → C → D, and A → D for both T2-w → PD-w and PD-w → T2-w.** ($v = 0.0075$)

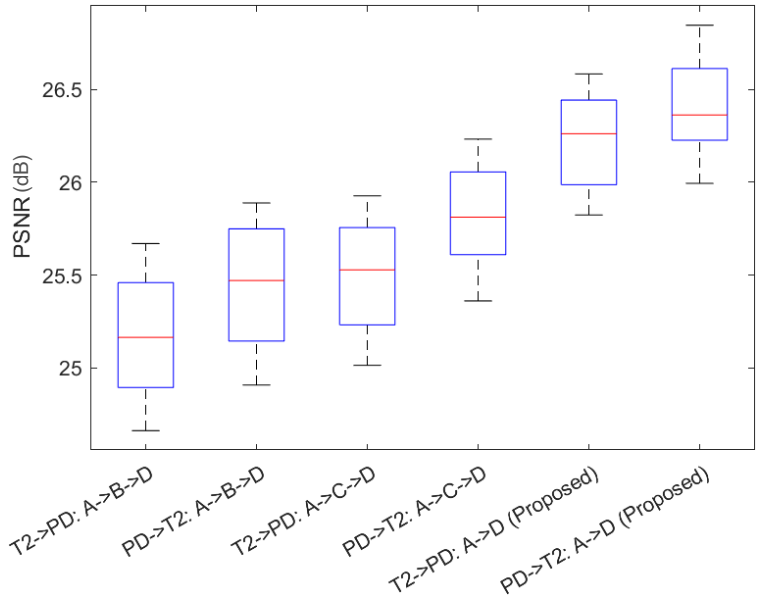


**Figure 5.16: Visual presentation of PSNR for the process A → B → D, A → C → D, and A → D for both T2-w → PD-w and PD-w → T2-w.** ($v = 0.005$)

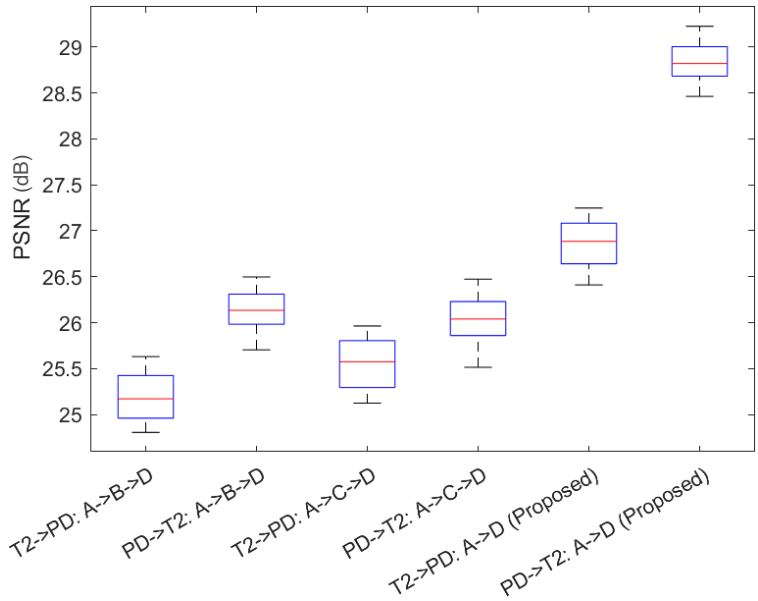


**Figure 5.17: Visual presentation of PSNR for the process A → B → D, A → C → D, and A → D for both T2-w → PD-w and PD-w → T2-w.** ($v = 0.0025$)

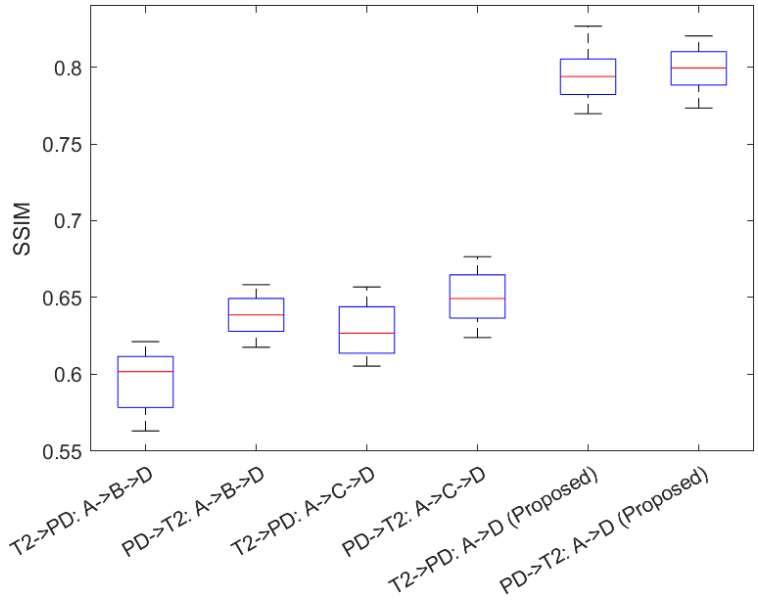


**Figure 5.18: Visual presentation of SSIM for the process A → B → D, A → C → D, and A → D for both T2-w → PD-w and PD-w → T2-w.** ($v = 0.01$)

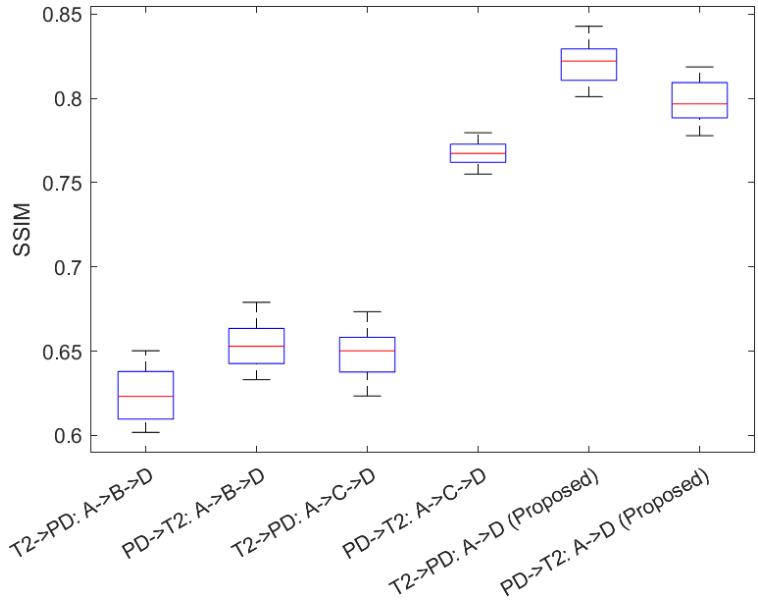


**Figure 5.19: Visual presentation of SSIM for the process A → B → D, A → C → D, and A → D for both T2-w → PD-w and PD-w → T2-w.** ($v = 0.0075$)

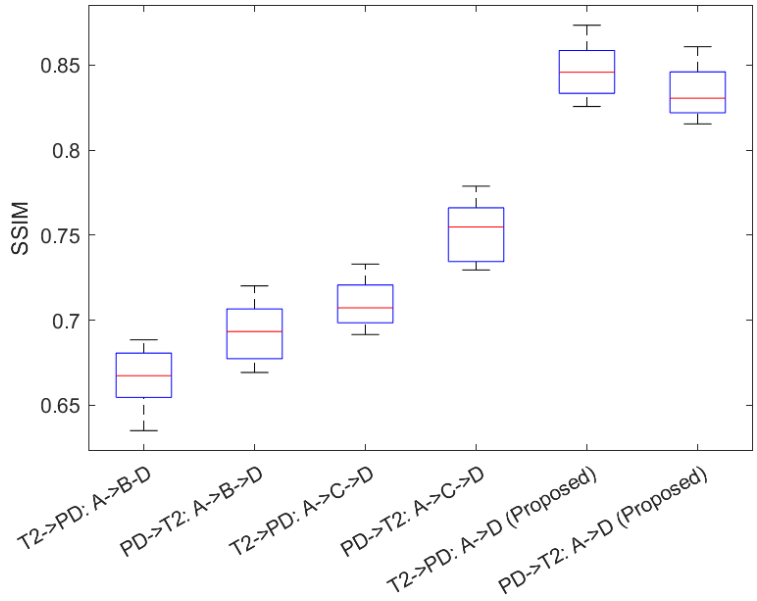


**Figure 5.20: Visual presentation of SSIM for the process A → B → D, A → C → D, and A → D for both T2-w → PD-w and PD-w → T2-w.** ($v = 0.005$)

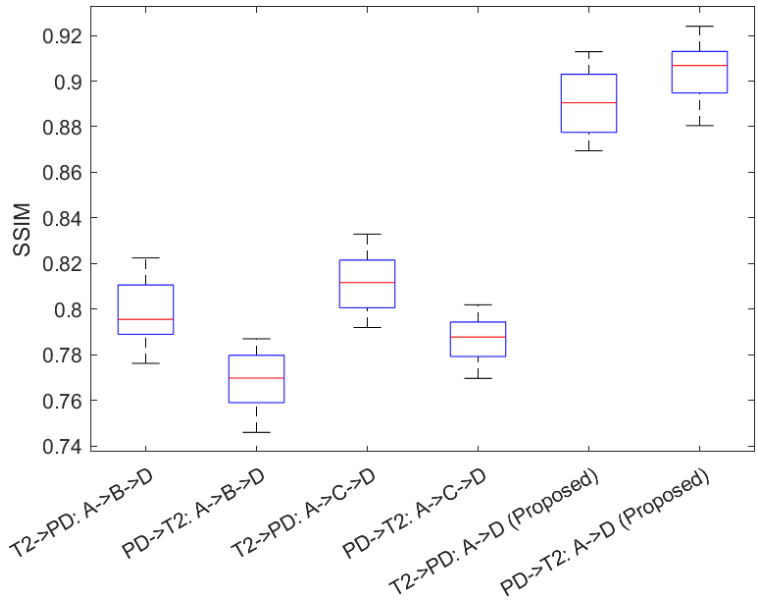


**Figure 5.21: Visual presentation of SSIM for the process A → B → D, A → C → D, and A → D for both T2-w → PD-w and PD-w → T2-w.** ($v = 0.0025$)

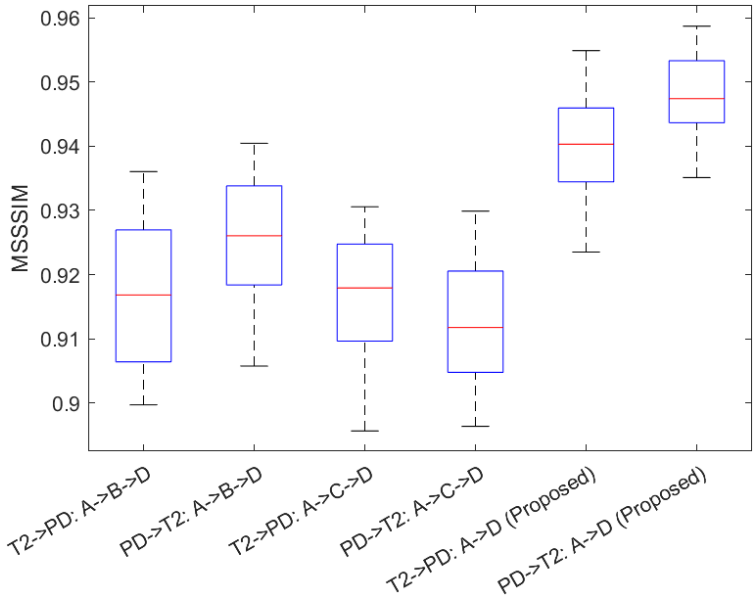


**Figure 5.22: Visual presentation of MSSSIM for the process A → B → D, A → C → D, and A → D for both T2-w → PD-w and PD-w → T2-w.** ($v = 0.01$)

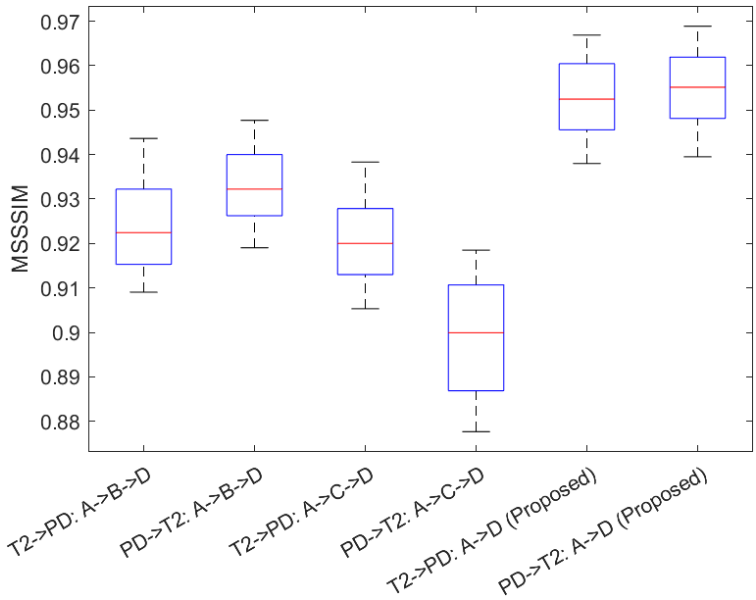


**Figure 5.23: Visual presentation of MSSSIM for the process A → B → D, A → C → D, and A → D for both T2-w → PD-w and PD-w → T2-w.** ($v = 0.0075$)

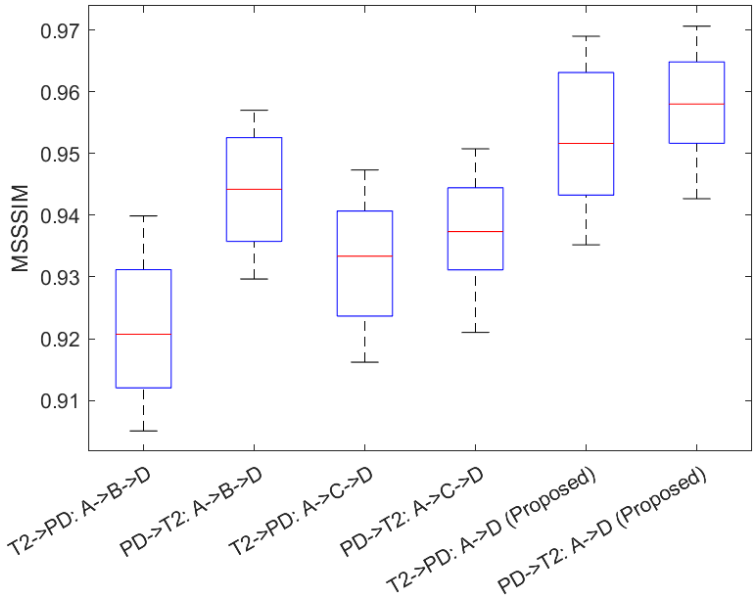


**Figure 5.24: Visual presentation of MSSSIM for the process A → B → D, A → C → D, and A → D for both T2-w → PD-w and PD-w → T2-w.** ($v = 0.005$)

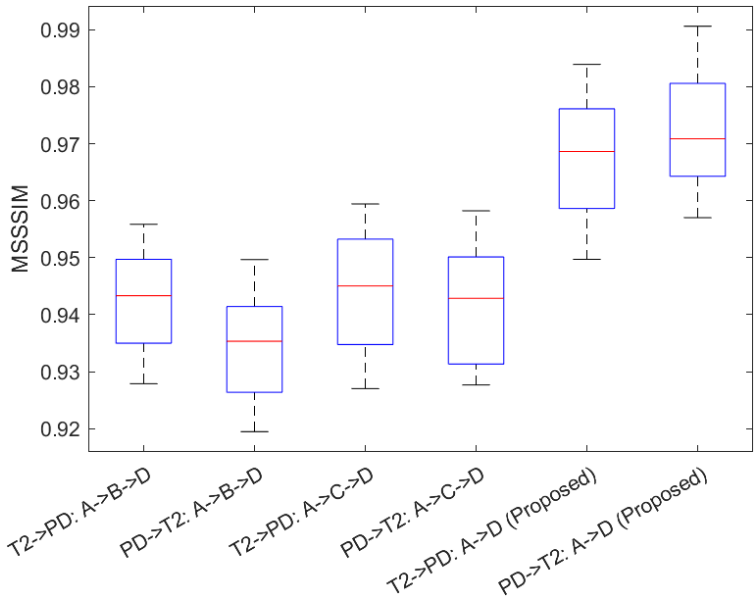


**Figure 5.25: Visual presentation of MSSSIM for the process A → B → D, A → C → D, and A → D for both T2-w → PD-w and PD-w → T2-w.** ($v = 0.0025$)

**Ablation Studies**

To further analyse the contribution of individual components and design choices in the proposed joint framework, ablation studies are conducted in this subsection.

In this section, we will confirm the effectiveness of the individual structures of the proposed model when dealing with simultaneous medical image cross-modality synthesis and denoising tasks through ablation studies. Meanwhile, confirm the validity of improvements made to the model. The proposed model is composed of multiple modules; in order to prove the necessity of each module, we will perform ablation studies based on each module.

Firstly, since the proposed MARG has achieved good performance in the LDCT image denoising task and the proposed DCD network structure has obtained great results in MRI cross-modality synthesis, it is necessary to study the performance of these two network structures in performing simultaneous cross-modality synthesis and denoising tasks. For this reason, we have done a large number of experiments, and even with different parameter settings, the obtained results lose some structural information and cannot deal with the noise while synthesising the other modality, which is insufficient for clinical purposes. Fig 5.26 shows some of the results, which demonstrate that the MARG network or DCD network alone is incapable of handling such multitasking work.

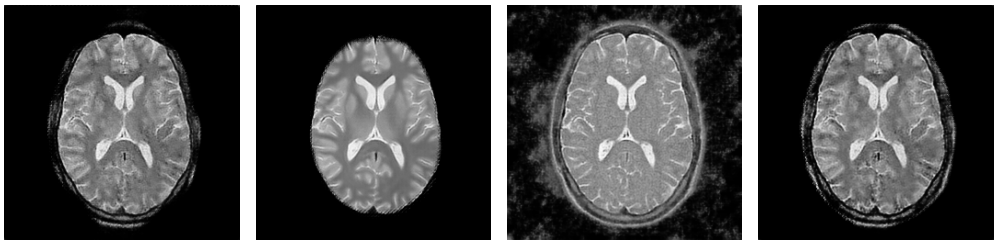

**Figure 5.26: Results of using single MARG or DCD networks for simultaneous cross-modality synthesis and denoising task**

Secondly, to demonstrate the effectiveness of our improved network structure, we compare the advantages of the modified CMS-DN MARG and CMS-DN DCD network structures against the direct use of the original MARG and DCD network structures. We use the group with a noise variance of 0.01 for the comparison, which will compare the two scenarios: T2-w to PD-w and PD-w to T2-w, respectively. Fig 5.27(a) shows the obtained result from T2-w to PD-w using the original MARG and DCD models, and Fig 5.27(c) is the result of the modified CMS-DN MARG and CMS-DN DCD models. Fig 5.28(a) presents the PD-w to T2-w results of the original MARG and DCD models, and Fig 5.28(c) is the corresponding result of using the CMS-DN MARG and CMS-DN DCD models. As can be seen from the results, the original MARG and DCD networks produce lacking structural information images, which have a significantly lower contrast and are visually inferior to the results (c) obtained from the proposed modified CMS-DN MARG and CMS-DN DCD models. On the other hand, the reduced weight of the network leads to a significant reduction in training time. All

modules are trained on a GTX 3090Ti GPU; the training time is shown in Table 5.3. In order to facilitate a visual comparison, Fig 5.32 is used to visualise the differences. The original MARG and DCD models cost 495 seconds per epoch, but the modified CMS-DN MARG and CMS-DN DCD models only took 290 seconds per epoch. Therefore, it is confirmed by this ablation study that the improved model outperforms the direct use of the two models superimposed on each other both in terms of performance and time consumption when dealing with multitasking.

Third, the generator design was executed with the intention of optimising the generator module's receptive field. By employing a multi-scale receptive field, the performance of multitasking is enhanced. The effectiveness of the improvement is demonstrated in the experiment. Similarly, use the group with a noise variance of 0.01 for the experiment and compare in two scenarios: from T2-w to PD-w and from PD-w to T2-w. Fig 5.27(b) shows the result from T2-w to PD-w of using the original receptive field, and Fig 5.27(c) is the result of the modified CMS-DN MARG and CMS-DN DCD models with the multi-scale receptive field. Fig 5.28(b) is the PD-w to T2-w results of the original receptive field, and Fig 5.28(c) presents the corresponding result of using the CMS-DN MARG and CMS-DN DCD models with the multi-scale receptive field. As can be seen from the results, the optimised multi-scale receptive field obtained better results. However, the training time is improved by about 10 seconds per epoch compared to the network structure using the original receptive field, which is shown in Table 5.3 and Fig 5.32. Since the difference is relatively small, we consider the time loss to be worthwhile relative to the performance gain.

Through the ablation studies, we have gained a deeper understanding of the proposed model's performance and have experimentally demonstrated the effectiveness of the improved model. It also further explains why our model achieves state-of-the-art results when performing the task of process-

ing simultaneous medical image cross-modality synthesis and denoising.

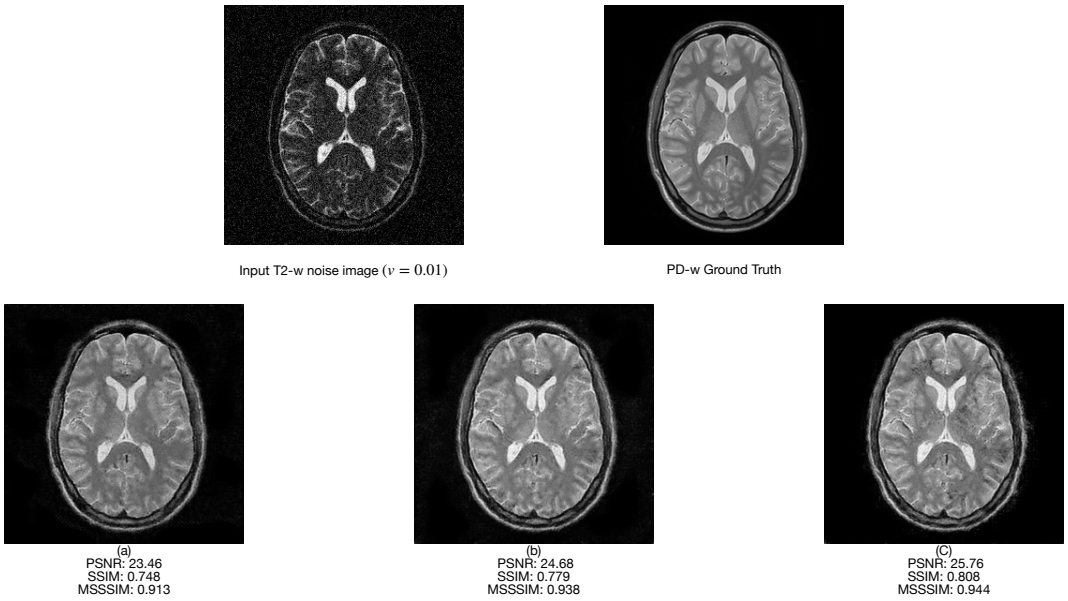


**Figure 5.27: Quantitative evaluation results of the ablation study for CMS-DN GAN. (a) Original MARG and DCD from T2-w to PD-w; (b) Unchanged receptive field from T2-w to PD-w; (c) Proposed method.**

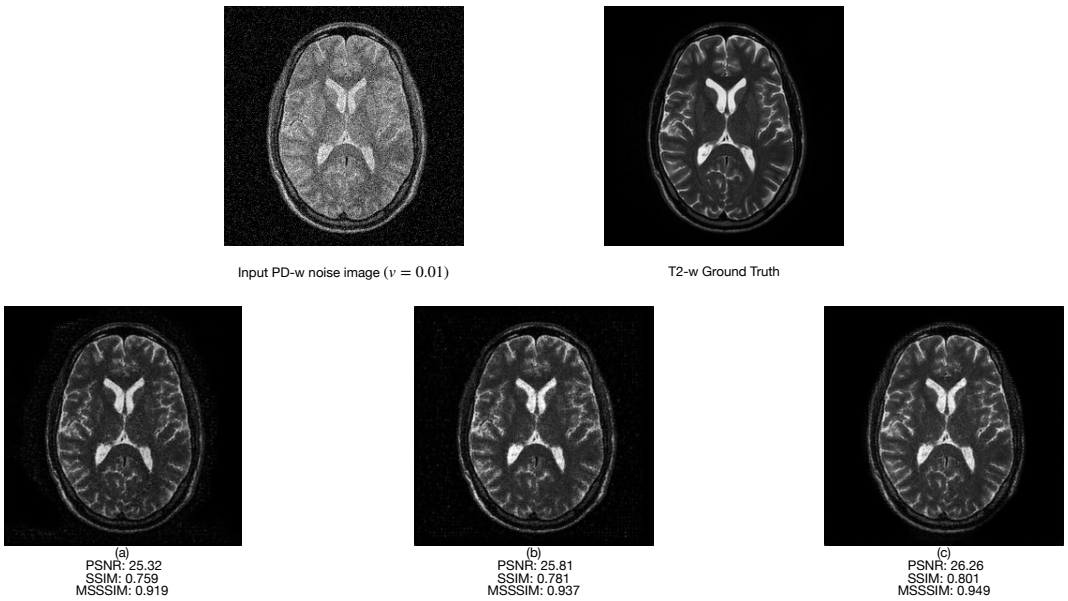


**Figure 5.28: Quantitative evaluation results of the ablation study for CMS-DN GAN. (a) Original MARG and DCD from PD-w to T2-w; (b) Unchanged receptive field from PD-w to T2-w; (c) Proposed method.**

| Metric (avg.) | (a) | (b) | (c) (Proposed) |
|---|---|---|---|
| IXI: T2-w —> PD-w | | | |
| PSNR (dB) | 23.46 | 24.68 | **25.76** |
| SSIM | 0.748 | 0.779 | **0.808** |
| MSSSIM | 0.913 | 0.938 | **0.944** |
| Training time (Sec/Epoch) | 495 | **280** | 290 |
| IXI: PD-w —> T2-w | | | |
| PSNR (dB) | 25.32 | 25.81 | **26.26** |
| SSIM | 0.759 | 0.781 | **0.801** |
| MSSSIM | 0.919 | 0.937 | **0.949** |
| Training time (Sec/Epoch) | 491 | **278** | 292 |

**Table 5.3: Quantitative evaluation(PSNR(dB), SSIM, MSSSIM, and Training time(epoch/sec)): (a) Original MARG and DCD; (b) Unchanged receptive field; (c) Proposed method. The best results are marked in bold.**

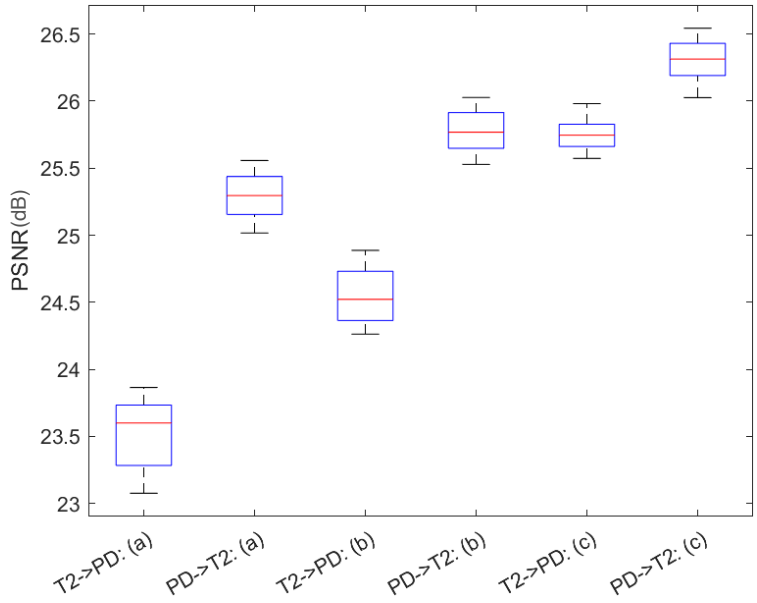


**Figure 5.29: Visual presentation of PSNR for the ablation study of CMS-DN GAN. (a) Original MARG and DCD; (b) Unchanged receptive field; (c) Proposed method.**

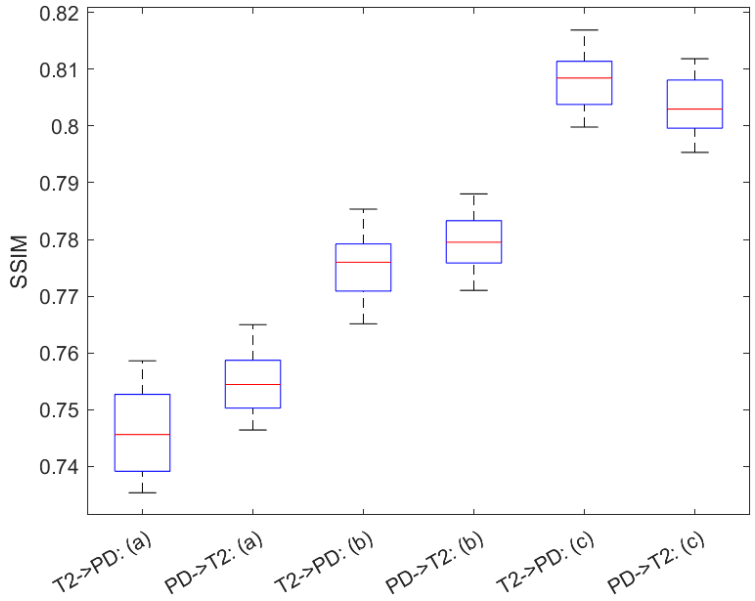


**Figure 5.30: Visual presentation of SSIM for the ablation study of CMS-DN GAN. (a) Original MARG and DCD; (b) Unchanged receptive field; (c) Proposed method.**

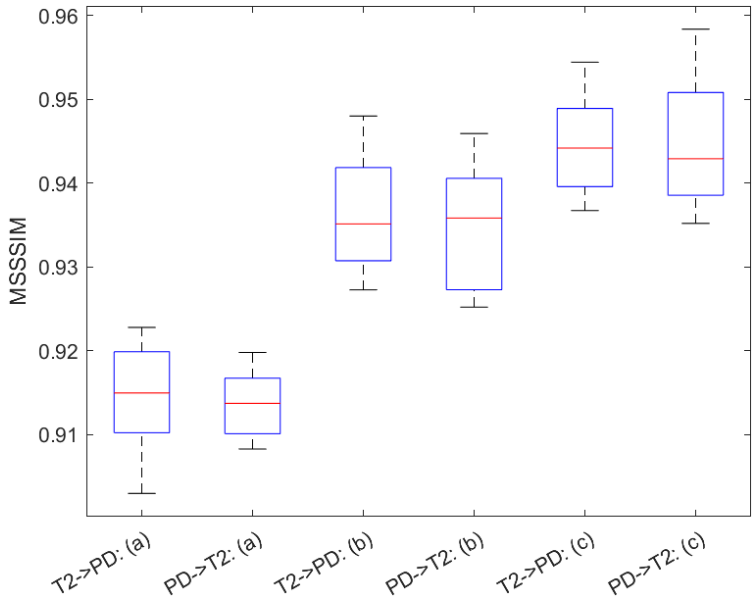


**Figure 5.31: Visual presentation of MSSSIM for the ablation study of CMS-DN GAN. (a) Original MARG and DCD; (b) Unchanged receptive field; (c) Proposed method.**

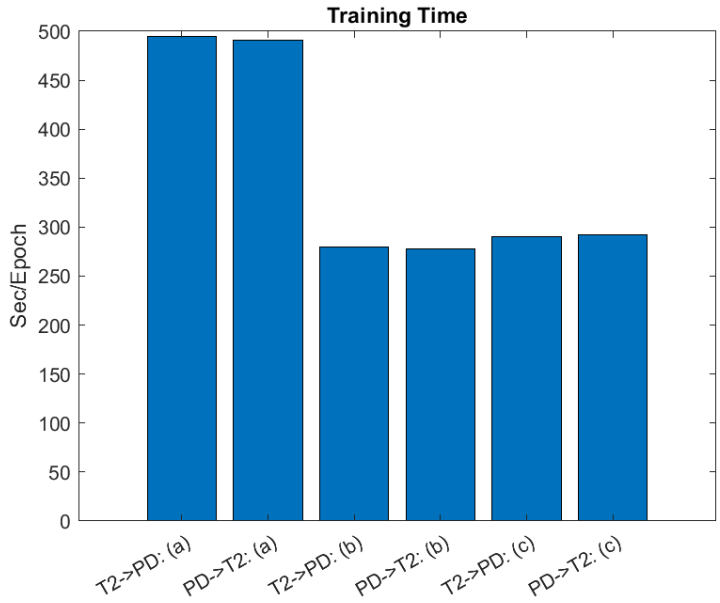


**Figure 5.32: Visual presentation of training time for the ablation study of CMS-DN GAN. (a) Original MARG and DCD; (b) Unchanged receptive field; (c) Proposed method.**

## 5.4 Conclusions

Drawing on the experimental results and analyses presented above, this section summarises the main findings of the proposed simultaneous medical image denoising and cross-modality synthesis approach and discusses its effectiveness within the scope of this chapter.

In this chapter, a unified framework was proposed to simultaneously address medical image denoising and cross-modality synthesis within a single network architecture. By integrating the unsupervised denoising strategy developed in Chapter 3 with the cross-modality synthesis mechanism introduced in Chapter 4, the proposed CMS-DN GAN enables joint optimisation of both tasks under a consistent learning objective.

Extensive experimental results on the IXI dataset demonstrate that the

proposed approach consistently outperforms sequential processing strategies, including denoising followed by synthesis and synthesis followed by denoising, across multiple noise levels and modality directions. For brain MRI cross-modality synthesis and denoising from T2-w to PD-w, the proposed method achieves improvements of up to 1.3 dB PSNR, 35.8% SSIM, and 2.6% MSSSIM compared with denoising-first strategies, and up to 3.12 dB PSNR, 27.2% SSIM, and 2.8% MSSSIM compared with synthesis-first strategies under high-noise conditions. Similar performance gains are observed for the reverse PD-w to T2-w task, confirming the robustness of the proposed model across different modality transformations.

These results indicate that simultaneous learning can better preserve structural consistency and visual realism than sequential pipelines, particularly under noisy and incomplete modality conditions. By jointly modelling denoising and cross-modality synthesis, the proposed framework reflects more realistic clinical imaging scenarios and provides a practical solution for integrated medical image processing tasks.

# Chapter 6

# Conclusion and Future Works

## 6.1 Conclusions

This thesis has systematically investigated the problems of medical image denoising and cross-modality synthesis under unsupervised learning settings, with a particular focus on realistic clinical scenarios characterised by noisy observations and incomplete modality acquisition. Three interconnected research objectives were addressed through the development of novel network architectures and learning strategies. In Chapter 3, an unsupervised low-dose CT (LDCT) image denoising framework was proposed to address severe noise degradation caused by reduced radiation dose. A Multi-Channel Asymmetric Residual Generator (MARG) was designed to capture noise characteristics and anatomical features at different scales. The asymmetric and multi-channel design was shown to enhance feature representation capability and robustness, enabling effective denoising without paired training data. Experimental results demonstrated that the proposed approach achieved state-of-the-art performance compared with existing denoising methods, validating the effectiveness of the proposed generator ar-

chitecture. In Chapter 4, an unsupervised brain MRI cross-modality synthesis framework was developed to mitigate the difficulty and cost of acquiring paired multi-modality data. A Dual-Channel Joint Discriminator (DCD) was introduced to jointly model local structural information and pixel-level fidelity. By providing complementary adversarial feedback from both channels, the discriminator strengthened the generator's ability to preserve anatomical structures and fine-grained details. Extensive experiments on brain MRI datasets demonstrated that the proposed method outperformed existing state-of-the-art cross-modality synthesis approaches. Building upon the individual denoising and cross-modality synthesis models, Chapter 5 pioneered a unified unsupervised framework for simultaneous medical image denoising and cross-modality synthesis. By jointly optimising both tasks within a single adversarial learning architecture, the proposed CMS-DN model effectively avoided error accumulation inherent in conventional sequential pipelines. Experimental evaluations confirmed that simultaneous learning not only preserved the performance of individual tasks but also improved overall image consistency and realism under varying noise levels and modality-missing conditions. Overall, the proposed multi-channel asymmetric residual generator and dual-channel joint discriminator constitute key methodological contributions to unsupervised medical image processing. The results presented in this thesis demonstrate that integrating architectural design with task-aware learning strategies can significantly improve performance in denoising, cross-modality synthesis, and their joint optimisation. These contributions provide a solid foundation for more advanced multi-task medical image processing frameworks and support the practical applicability of unsupervised learning methods in clinical imaging environments.

## 6.2 Future Works

Although this thesis demonstrates the effectiveness of unsupervised medical image denoising, cross-modality synthesis, and their joint optimisation within a unified adversarial learning framework, several important research directions remain open for further investigation. In realistic clinical imaging scenarios, medical images are often affected by multiple degradations simultaneously, including noise corruption, missing modalities, and limited spatial resolution. Therefore, a natural extension of this work is the development of a fully unified multi-task learning framework that can simultaneously address medical image denoising, cross-modality synthesis, and super-resolution reconstruction within a single network architecture. Building upon the proposed Multi-Channel Asymmetric Residual Generator (MARG) and Dual-Channel Joint Discriminator (DCD), future research will explore how these components can be further extended to support triple-task or more complex task configurations while maintaining training stability and image fidelity.

Another important direction for future work lies in improving the generalisation ability of the proposed models across different scanners, imaging protocols, and anatomical regions. Although the methods presented in this thesis have been validated on representative CT and MRI datasets, medical images acquired from different vendors and clinical centres often exhibit substantial variability in appearance, noise characteristics, and contrast distributions. Future studies will therefore investigate strategies such as domain adaptation, modality-invariant feature learning, and adaptive normalisation to enhance robustness and ensure consistent performance across heterogeneous clinical datasets.

In addition, while the use of multi-channel generators and dual-channel discriminators has demonstrated clear advantages in representation learn-

ing and image quality, these designs inevitably increase computational complexity. Future research will focus on reducing model complexity through techniques such as parameter sharing, channel pruning, and lightweight discriminator design, with the aim of achieving a better balance between performance and computational efficiency. Such efforts are particularly important for enabling practical deployment in time-sensitive or resource-constrained clinical environments, including real-time image guidance and intraoperative applications.

From an evaluation perspective, future work will also consider extending beyond conventional image quality metrics such as PSNR, SSIM, and MSSSIM. Although these metrics provide useful quantitative comparisons, they do not fully capture the clinical relevance of reconstructed or synthesised images. Incorporating task-driven evaluation strategies, such as assessing the impact of denoising and synthesis on downstream tasks including segmentation accuracy, lesion detection, and diagnostic consistency, may provide more meaningful insights into the practical utility of the proposed methods.

Although this thesis focuses on fully unsupervised learning, future research may explore semi-supervised or weakly supervised extensions in scenarios where limited paired or annotated data are available. Hybrid learning strategies that combine unsupervised learning with sparse supervision could further improve performance while preserving robustness in situations where large-scale paired datasets remain difficult to obtain. Expanding the training datasets to include a broader range of pathological cases and rare disease conditions may also enhance the clinical applicability of the proposed framework.

In summary, future work will aim to develop a comprehensive and scalable medical image processing framework capable of jointly addressing multiple imaging challenges within a unified learning paradigm. By reducing

reliance on multiple acquisition protocols and multi-stage post-processing pipelines, such a framework has the potential to improve clinical workflow efficiency, reduce patient burden, and provide more reliable inputs for downstream clinical analysis and decision-making.

# Bibliography

[1] K. Z. Abd-Elmoniem, A.-B. Youssef, and Y. M. Kadah, "Real-time speckle reduction and coherence enhancement in ultrasound imaging via nonlinear anisotropic diffusion," *IEEE Transactions on Biomedical Engineering*, vol. 49, no. 9, pp. 997–1014, 2002.

[2] M. Aharon, M. Elad, and A. Bruckstein, "K-svd: An algorithm for designing of overcomplete dictionaries for sparse representation technion—israel inst. of technology, 2005," *Tech. Ref*.

[3] S. Aja-Fernández, C. Alberola-López, and C.-F. Westin, "Noise and signal estimation in magnitude mri and rician distributed images: a lmmse approach," *IEEE transactions on image processing*, vol. 17, no. 8, pp. 1383–1398, 2008.

[4] G. Andria, F. Attivissimo, G. Cavone, N. Giaquinto, and A. Lanzolla, "Linear filtering of 2-d wavelet coefficients for denoising ultrasound medical images," *Measurement*, vol. 45, no. 7, pp. 1792–1800, 2012.

[5] M. Arjovsky, S. Chintala, and L. Bottou, "Wasserstein generative adversarial networks," in *International conference on machine learning*. PMLR, 2017, pp. 214–223.

[6] K. Armanious, C. Jiang, M. Fischer, T. Küstner, T. Hepp, K. Nikolaou, S. Gatidis, and B. Yang, "Medgan: Medical image translation using

gans," *Computerized medical imaging and graphics*, vol. 79, p. 101684, 2020.

[7] J. J. J. Babu and G. F. Sudha, "Adaptive speckle reduction in ultrasound images using fuzzy logic on coefficient of variation," *Biomedical Signal Processing and Control*, vol. 23, pp. 93–103, 2016.

[8] F. Baselice, G. Ferraioli, V. Pascazio, and A. Sorriso, "Denoising of mr images using kolmogorov-smirnov distance in a non local framework," *Magnetic resonance imaging*, vol. 57, pp. 176–193, 2019.

[9] U. Bergmann, N. Jetchev, and R. Vollgraf, "Learning texture manifolds with the periodic spatial gan," *arXiv preprint arXiv:1705.06566*, 2017.

[10] D. Bhonsle, V. Chandra, and G. Sinha, "Medical image denoising using bilateral filter," *International Journal of Image, Graphics and Signal Processing*, vol. 4, no. 6, p. 36, 2012.

[11] M. I. H. Bhuiyan, M. O. Ahmad, and M. Swamy, "Spatially adaptive thresholding in wavelet domain for despeckling of ultrasound images," *IET Image processing*, vol. 3, no. 3, pp. 147–162, 2009.

[12] L. Bi, J. Kim, A. Kumar, D. Feng, and M. Fulham, "Synthesis of positron emission tomography (pet) images via multi-channel generative adversarial networks (gans)," in *Molecular Imaging, Reconstruction and Analysis of Moving Body Organs, and Stroke Imaging and Treatment: Fifth International Workshop, CMMI 2017, Second International Workshop, RAMBO 2017, and First International Workshop, SWITCH 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 14, 2017, Proceedings 5*. Springer, 2017, pp. 43–51.

[13] A. Borsdorf, R. Raupach, T. Flohr, and J. Hornegger, "Wavelet based noise reduction in ct-images using correlation analysis," *IEEE transactions on medical imaging*, vol. 27, no. 12, pp. 1685–1703, 2008.

[14] A. Brock, J. Donahue, and K. Simonyan, "Large scale gan training for high fidelity natural image synthesis," *arXiv preprint arXiv:1809.11096*, 2018.

[15] A. Buades, B. Coll, and J.-M. Morel, "A non-local algorithm for image denoising," in *2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05)*, vol. 2. Ieee, 2005, pp. 60–65.

[16] F. Calimeri, A. Marzullo, C. Stamile, and G. Terracina, "Biomedical data augmentation using generative adversarial neural networks," in *International conference on artificial neural networks*. Springer, 2017, pp. 626–634.

[17] T. Che, Y. Li, A. P. Jacob, Y. Bengio, and W. Li, "Mode regularized generative adversarial networks," *arXiv preprint arXiv:1612.02136*, 2016.

[18] C. Chen, Q. Dou, H. Chen, and P.-A. Heng, "Semantic-aware generative adversarial nets for unsupervised domain adaptation in chest x-ray segmentation," in *Machine Learning in Medical Imaging: 9th International Workshop, MLMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 9*. Springer, 2018, pp. 143–151.

[19] H. Chen, G. Yang, and H. Zhang, "Hider: A hyperspectral image denoising transformer with spatial–spectral constraints for hybrid noise removal," *IEEE Transactions on Neural Networks and Learning Systems*, 2022.

[20] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose ct with a residual encoder-decoder convolutional

neural network," *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.

[21] ——, "Low-dose ct with a residual encoder-decoder convolutional neural network," *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.

[22] J. Chen, J. Chen, H. Chao, and M. Yang, "Image blind denoising with generative adversarial network based noise modeling," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 3155–3164.

[23] X. Chen, Y. Duan, R. Houthooft, J. Schulman, I. Sutskever, and P. Abbeel, "Infogan: Interpretable representation learning by information maximizing generative adversarial nets," *Advances in neural information processing systems*, vol. 29, 2016.

[24] Y. Chen, L. Shi, Q. Feng, J. Yang, H. Shu, L. Luo, J.-L. Coatrieux, and W. Chen, "Artifact suppressed dictionary learning for low-dose ct image processing," *IEEE transactions on medical imaging*, vol. 33, no. 12, pp. 2271–2292, 2014.

[25] H. Choi and D. S. Lee, "Generation of structural mr images from amyloid pet: application to mr-less quantification," *Journal of Nuclear Medicine*, vol. 59, no. 7, pp. 1111–1117, 2018.

[26] Y. Choi, M. Choi, M. Kim, J.-W. Ha, S. Kim, and J. 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*, 2018, pp. 8789–8797.

[27] T. R. Corle and G. S. Kino, "Chapter 1: Introduction," *Library Technology Reports*, vol. 48, no. 4, pp. 6–9, 2012.

[28] P. Coupé, P. Hellier, S. Prima, C. Kervrann, and C. Barillot, "3d wavelet subbands mixing for image denoising," *International Journal of Biomedical Imaging*, vol. 2008, 2008.

[29] P. Coupé, J. V. Manjón, M. Robles, and D. L. Collins, "Adaptive multiresolution non-local means filter for three-dimensional magnetic resonance image denoising," *IET image Processing*, vol. 6, no. 5, pp. 558–568, 2012.

[30] O. Dalmaz, M. Yurt, and T. Çukur, "Resvit: Residual vision transformers for multimodal medical image synthesis," *IEEE Transactions on Medical Imaging*, vol. 41, no. 10, pp. 2598–2614, 2022.

[31] S. U. Dar, M. Yurt, L. Karacan, A. Erdem, E. Erdem, and T. Cukur, "Image synthesis in multi-contrast mri with conditional generative adversarial networks," *IEEE transactions on medical imaging*, vol. 38, no. 10, pp. 2375–2388, 2019.

[32] Q. Ding, Y. Long, X. Zhang, and J. A. Fessler, "Statistical image reconstruction using mixed poisson-gaussian noise model for x-ray ct," *arXiv preprint arXiv:1801.09533*, 2018.

[33] M. Diwakar and M. Kumar, "Edge preservation based ct image denoising using wiener filtering and thresholding in wavelet domain," in *2016 Fourth International Conference on Parallel, Distributed and Grid Computing (PDGC)*. IEEE, 2016, pp. 332–336.

[34] S. Dolui, A. Kuurstra, I. C. S. Patarroyo, and O. V. Michailovich, "A new similarity measure for non-local means filtering of mri images," *Journal of visual communication and image representation*, vol. 24, no. 7, pp. 1040–1054, 2013.

[35] J. Donahue, P. Krähenbühl, and T. Darrell, "Adversarial feature learning," *arXiv preprint arXiv:1605.09782*, 2016.

[36] G. Dong, Y. Ma, and A. Basu, "Feature-guided cnn for denoising images from portable ultrasound devices," *IEEE Access*, vol. 9, pp. 28 272–28 281, 2021.

[37] V. Dumoulin, I. Belghazi, B. Poole, O. Mastropietro, A. Lamb, M. Arjovsky, and A. Courville, "Adversarially learned inference," *arXiv preprint arXiv:1606.00704*, 2016.

[38] M. Elhoseny and K. Shankar, "Optimal bilateral filter and convolutional neural network based denoising method of medical image measurements," *Measurement*, vol. 143, pp. 125–135, 2019.

[39] Y. Erez, Y. Y. Schechner, and D. Adam, "Ultrasound image denoising by spatially varying frequency compounding," in *Joint Pattern Recognition Symposium*. Springer, 2006, pp. 1–10.

[40] C.-M. Fan, T.-J. Liu, and K.-H. Liu, "Sunet: swin transformer unet for image denoising," in *2022 IEEE International Symposium on Circuits and Systems (ISCAS)*. IEEE, 2022, pp. 2333–2337.

[41] P. F. Feruglio, C. Vinegoni, J. Gros, A. Sbarbati, and R. Weissleder, "Block matching 3d random noise filtering for absorption optical projection tomography," *Physics in Medicine & Biology*, vol. 55, no. 18, p. 5401, 2010.

[42] M. Geng, X. Meng, J. Yu, L. Zhu, L. Jin, Z. Jiang, B. Qiu, H. Li, H. Kong, J. Yuan *et al.*, "Content-noise complementary learning for medical image denoising," *IEEE transactions on medical imaging*, vol. 41, no. 2, pp. 407–419, 2021.

[43] H. M. Golshan and R. P. Hasanzadeh, "A non-local rician noise reduction approach for 3-d magnitude magnetic resonance images," in *2011 7th Iranian Conference on Machine Vision and Image Processing*. IEEE, 2011, pp. 1–5.

[44] L. Gondara, "Medical image denoising using convolutional denoising autoencoders," in *2016 IEEE 16th international conference on data mining workshops (ICDMW)*. IEEE, 2016, pp. 241–246.

[45] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial networks," *Communications of the ACM*, vol. 63, no. 11, pp. 139–144, 2020.

[46] P. Gravel, G. Beaudoin, and J. A. De Guise, "A method for modeling noise in medical images," *IEEE Transactions on medical imaging*, vol. 23, no. 10, pp. 1221–1232, 2004.

[47] F. Guan, P. Ton, S. Ge, and L. Zhao, "Anisotropic diffusion filtering for ultrasound speckle reduction," *Science China Technological Sciences*, vol. 57, pp. 607–614, 2014.

[48] H. Gudbjartsson and S. Patz, "The rician distribution of noisy mri data," *Magnetic resonance in medicine*, vol. 34, no. 6, pp. 910–914, 1995.

[49] I. Gulrajani, F. Ahmed, M. Arjovsky, V. Dumoulin, and A. C. Courville, "Improved training of wasserstein gans," *Advances in neural information processing systems*, vol. 30, 2017.

[50] P. Guo, P. Wang, R. Yasarla, J. Zhou, V. M. Patel, and S. Jiang, "Anatomic and molecular mr image synthesis using confidence guided cnns," *IEEE transactions on medical imaging*, vol. 40, no. 10, pp. 2832–2844, 2020.

[51] N. Gupta, M. Swamy, and E. Plotkin, "Despeckling of medical ultrasound images using data and rate adaptive lossy compression," *IEEE Transactions on Medical Imaging*, vol. 24, no. 6, pp. 743–754, 2005.

[52] C. Han, H. Hayashi, L. Rundo, R. Araki, W. Shimoda, S. Muramatsu, Y. Furukawa, G. Mauri, and H. Nakayama, "Gan-based synthetic brain mr image generation," in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018, pp. 734–738.

[53] S. Hashemi, N. S. Paul, S. Beheshti, and R. S. Cobbold, "Adaptively tuned iterative low dose ct image denoising," *Computational and mathematical methods in medicine*, vol. 2015, 2015.

[54] D. He, Y. Xia, T. Qin, L. Wang, N. Yu, T.-Y. Liu, and W.-Y. Ma, "Dual learning for machine translation," *Advances in neural information processing systems*, vol. 29, 2016.

[55] R. M. Henkelman, "Measurement of signal intensities in the presence of noise in mr images," *Medical physics*, vol. 12, no. 2, pp. 232–233, 1985.

[56] Y. Hiasa, Y. Otake, M. Takao, T. Matsuoka, K. Takashima, A. Carass, J. L. Prince, N. Sugano, and Y. Sato, "Cross-modality image synthesis from unpaired data using cyclegan: Effects of gradient consistency loss and training data size," in *Simulation and Synthesis in Medical Imaging: Third International Workshop, SASHIMI 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 3*. Springer, 2018, pp. 31–41.

[57] Z. Hu, H. Liu, Z. Li, and Z. Yu, "Data-enabled intelligence in complex industrial systems cross-model transformer method for medical image synthesis," *Complexity*, vol. 2021, pp. 1–7, 2021.

[58] R. Huang, S. Zhang, T. Li, and R. He, "Beyond face rotation: Global and local perception gan for photorealistic and identity preserving frontal view synthesis," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2439–2448.

[59] X. Huang, M.-Y. Liu, S. Belongie, and J. Kautz, "Multimodal unsupervised image-to-image translation," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 172–189.

[60] Y. Huang, L. Beltrachini, L. Shao, and A. F. Frangi, "Geometry regularized joint dictionary learning for cross-modality image synthesis in magnetic resonance imaging," in *Simulation and Synthesis in Medical Imaging: First International Workshop, SASHIMI 2016, Held in Conjunction with MICCAI 2016, Athens, Greece, October 21, 2016, Proceedings 1*. Springer, 2016, pp. 118–126.

[61] Y. Huang, F. Zheng, R. Cong, W. Huang, M. R. Scott, and L. Shao, "Mcmt-gan: Multi-task coherent modality transferable gan for 3d brain image synthesis," *IEEE Transactions on Image Processing*, vol. 29, pp. 8187–8198, 2020.

[62] Y. Huang, F. Zheng, D. Wang, J. Jiang, X. Wang, and L. Shao, "Super-resolution and inpainting with degraded and upgraded generative adversarial networks," in *Proceedings of the Twenty-Ninth International Conference on International Joint Conferences on Artificial Intelligence*, 2021, pp. 645–651.

[63] Z. Huang, J. Zhang, Y. Zhang, and H. Shan, "Du-gan: Generative adversarial networks with dual-domain u-net-based discriminators for low-dose ct denoising," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2021.

[64] Y. Huo, Z. Xu, S. Bao, A. Assad, R. G. Abramson, and B. A. Landman, "Adversarial synthesis learning enables segmentation without target modality ground truth," in *2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018, pp. 1217–1220.

[65] P. Isola, J.-Y. Zhu, T. Zhou, and A. A. Efros, "Image-to-image translation with conditional adversarial networks," in *Proceedings of the*

*IEEE conference on computer vision and pattern recognition*, 2017, pp. 1125–1134.

[66] N. Jetchev, U. Bergmann, and R. Vollgraf, "Texture synthesis with spatial generative adversarial networks (2016)," *arXiv preprint arXiv:1611.08207*.

[67] G. Jiang, Y. Lu, J. Wei, and Y. Xu, "Synthesize mammogram from digital breast tomosynthesis with gradient guided cgans," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22*. Springer, 2019, pp. 801–809.

[68] C.-B. Jin, H. Kim, M. Liu, W. Jung, S. Joo, E. Park, Y. S. Ahn, I. H. Han, J. I. Lee, and X. Cui, "Deep ct to mr synthesis using paired and unpaired data," *Sensors*, vol. 19, no. 10, p. 2361, 2019.

[69] J. Johnson, A. Alahi, and L. Fei-Fei, "Perceptual losses for real-time style transfer and super-resolution," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14*. Springer, 2016, pp. 694–711.

[70] S. A. Kamran, K. F. Hossain, A. Tavakkoli, S. L. Zuckerbrod, and S. A. Baker, "Vtgan: Semi-supervised retinal image synthesis and disease prediction using vision transformers," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 3235–3245.

[71] T. Kim, M. Cha, H. Kim, J. K. Lee, and J. Kim, "Learning to discover cross-domain relations with generative adversarial networks," in *International conference on machine learning*. PMLR, 2017, pp. 1857–1865.

[72] D. Kinga, J. B. Adam *et al.*, "A method for stochastic optimization," in *International conference on learning representations (ICLR)*, vol. 5. San Diego, California;, 2015, p. 6.

[73] D. P. Kingma and M. Welling, "Auto-encoding variational bayes," *arXiv preprint arXiv:1312.6114*, 2013.

[74] L. Kong, C. Lian, D. Huang, Y. Hu, Q. Zhou *et al.*, "Breaking the dilemma of medical image-to-image translation," *Advances in Neural Information Processing Systems*, vol. 34, pp. 1964–1978, 2021.

[75] Y. Korkmaz, S. U. Dar, M. Yurt, M. Özbey, and T. Cukur, "Unsupervised mri reconstruction via zero-shot learned adversarial transformers," *IEEE Transactions on Medical Imaging*, vol. 41, no. 7, pp. 1747–1763, 2022.

[76] Y. Lan and X. Zhang, "Real-time ultrasound image despeckling using mixed-attention mechanism based residual unet," *IEEE Access*, vol. 8, pp. 195 327–195 340, 2020.

[77] Y. LeCun, B. Boser, J. S. Denker, D. Henderson, R. E. Howard, W. Hubbard, and L. D. Jackel, "Backpropagation applied to handwritten zip code recognition," *Neural computation*, vol. 1, no. 4, pp. 541–551, 1989.

[78] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. 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*, 2017, pp. 4681–4690.

[79] J.-S. Lee, "Digital image enhancement and noise filtering by use of local statistics," *IEEE transactions on pattern analysis and machine intelligence*, no. 2, pp. 165–168, 1980.

[80] C. Li and M. Wand, "Precomputed real-time texture synthesis with markovian generative adversarial networks," in *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands,*

*October 11-14, 2016, Proceedings, Part III 14*. Springer, 2016, pp. 702–716.

[81] G.-T. Li, C.-L. Wang, P.-P. Huang, and W.-D. Yu, "Sar image despeckling using a space-domain filter with alterable window," *IEEE Geoscience and Remote Sensing Letters*, vol. 10, no. 2, pp. 263–267, 2012.

[82] J. Li, W. Monroe, T. Shi, S. Jean, A. Ritter, and D. Jurafsky, "Adversarial learning for neural dialogue generation," *arXiv preprint arXiv:1701.06547*, 2017.

[83] M. Li, W. Hsu, X. Xie, J. Cong, and W. Gao, "Sacnn: Self-attention convolutional neural network for low-dose ct denoising with self-supervised perceptual loss network," *IEEE transactions on medical imaging*, vol. 39, no. 7, pp. 2289–2301, 2020.

[84] Y. Li, W. Li, J. Xiong, J. Xia, Y. Xie *et al.*, "Comparison of supervised and unsupervised deep learning methods for medical image synthesis between computed tomography and magnetic resonance images," *BioMed Research International*, vol. 2020, 2020.

[85] K. Lin, D. Li, X. He, Z. Zhang, and M.-T. Sun, "Adversarial ranking for language generation," *Advances in neural information processing systems*, vol. 30, 2017.

[86] M.-Y. Liu, T. Breuel, and J. Kautz, "Unsupervised image-to-image translation networks," *Advances in neural information processing systems*, vol. 30, 2017.

[87] P. Liu, M. D. El Basha, Y. Li, Y. Xiao, P. C. Sanelli, and R. Fang, "Deep evolutionary networks with expedited genetic algorithms for medical image denoising," *Medical image analysis*, vol. 54, pp. 306–315, 2019.

[88] M. Long, J. Wang, Y. Cao, J. Sun, and S. Y. Philip, "Deep learning of transferable representation for scalable domain adaptation," *IEEE*

*Transactions on Knowledge and Data Engineering*, vol. 28, no. 8, pp. 2027–2040, 2016.

[89] T. Loupas, W. McDicken, and P. L. Allan, "An adaptive weighted median filter for speckle suppression in medical ultrasonic images," *IEEE transactions on Circuits and Systems*, vol. 36, no. 1, pp. 129–135, 1989.

[90] Y. Luo, Y. Wang, C. Zu, B. Zhan, X. Wu, J. Zhou, D. Shen, and L. Zhou, "3d transformer-gan for high-quality pet reconstruction," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part VI 24*. Springer, 2021, pp. 276–285.

[91] A. Luthra, H. Sulakhe, T. Mittal, A. Iyer, and S. Yadav, "Eformer: Edge enhancement based transformer for medical image denoising," *arXiv preprint arXiv:2109.08044*, 2021.

[92] J. Ma, J. Huang, Q. Feng, H. Zhang, H. Lu, Z. Liang, and W. Chen, "Low-dose computed tomography image restoration using previous normal-dose scan," *Medical physics*, vol. 38, no. 10, pp. 5713–5731, 2011.

[93] A. Makhzani, J. Shlens, N. Jaitly, I. Goodfellow, and B. Frey, "Adversarial autoencoders," *arXiv preprint arXiv:1511.05644*, 2015.

[94] A. Manduca, L. Yu, J. D. Trzasko, N. Khaylova, J. M. Kofler, C. M. McCollough, and J. G. Fletcher, "Projection space denoising with bilateral filtering and ct noise modeling for dose reduction in ct," *Medical physics*, vol. 36, no. 11, pp. 4911–4919, 2009.

[95] J. V. Manjón, J. Carbonell-Caballero, J. J. Lull, G. García-Martí, L. Martí-Bonmatí, and M. Robles, "Mri denoising using non-local means," *Medical image analysis*, vol. 12, no. 4, pp. 514–523, 2008.

[96] J. V. Manjón, P. Coupé, and A. Buades, "Mri noise estimation and denoising using non-local pca," *Medical image analysis*, vol. 22, no. 1, pp. 35–47, 2015.

[97] J. V. Manjón, P. Coupé, A. Buades, D. L. Collins, and M. Robles, "New methods for mri denoising based on sparseness and self-similarity," *Medical image analysis*, vol. 16, no. 1, pp. 18–27, 2012.

[98] J. V. Manjón, P. Coupé, L. Martí-Bonmatí, D. L. Collins, and M. Robles, "Adaptive non-local means denoising of mr images with spatially varying noise levels," *Journal of Magnetic Resonance Imaging*, vol. 31, no. 1, pp. 192–203, 2010.

[99] X. Mao, Q. Li, H. Xie, R. Y. Lau, Z. Wang, and S. Paul Smolley, "Least squares generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2794–2802.

[100] M. Mardani, E. Gong, J. Y. Cheng, S. Vasanawala, G. Zaharchuk, M. Alley, N. Thakur, S. Han, W. Dally, J. M. Pauly *et al.*, "Deep generative adversarial networks for compressed sensing automates mri," *arXiv preprint arXiv:1706.00051*, 2017.

[101] M. Maspero, M. H. Savenije, A. M. Dinkla, P. R. Seevinck, M. P. Intven, I. M. Jurgenliemk-Schulz, L. G. Kerkmeijer, and C. A. van den Berg, "Dose evaluation of fast synthetic-ct generation using a generative adversarial network for general pelvis mr-only radiotherapy," *Physics in Medicine & Biology*, vol. 63, no. 18, p. 185001, 2018.

[102] C. McCollough, B. Chen, D. Holmes, X. Duan, Z. Yu, L. Xu, S. Leng, and J. Fletcher, "Low dose ct image and projection data [data set]," *The Cancer Imaging Archive*, vol. 10, 2020.

[103] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[104] D. Mishra, S. Chaudhury, M. Sarkar, and A. S. Soin, "Ultrasound image enhancement using structure oriented adversarial network," *IEEE Signal Processing Letters*, vol. 25, no. 9, pp. 1349–1353, 2018.

[105] D. Mittal, V. Kumar, S. C. Saxena, N. Khandelwal, and N. Kalra, "Enhancement of the ultrasound images by modified anisotropic diffusion method," *Medical & biological engineering & computing*, vol. 48, pp. 1281–1291, 2010.

[106] V. Mnih, K. Kavukcuoglu, D. Silver, A. A. Rusu, J. Veness, M. G. Bellemare, A. Graves, M. Riedmiller, A. K. Fidjeland, G. Ostrovski *et al.*, "Human-level control through deep reinforcement learning," *nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[107] T. R. Moen, B. Chen, D. R. Holmes III, X. Duan, Z. Yu, L. Yu, S. Leng, J. G. Fletcher, and C. H. McCollough, "Low-dose ct image and projection dataset," *Medical physics*, vol. 48, no. 2, pp. 902–911, 2021.

[108] J. Mohan, V. Krishnaveni, and Y. Guo, "A survey on the magnetic resonance image denoising methods," *Biomedical signal processing and control*, vol. 9, pp. 56–69, 2014.

[109] T.-T. Nguyen, D.-H. Trinh, and N. Linh-Trung, "An efficient example-based method for ct image denoising based on frequency decomposition and sparse representation," in *2016 International Conference on Advanced Technologies for Communications (ATC)*. IEEE, 2016, pp. 293–296.

[110] D. Nie, R. Trullo, J. Lian, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with context-aware generative adversarial networks," in *Medical Image Computing and Computer Assisted Intervention- MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part III 20*. Springer, 2017, pp. 417–425.

[111] D. Nie, R. Trullo, J. Lian, L. Wang, C. Petitjean, S. Ruan, Q. Wang, and D. Shen, "Medical image synthesis with deep convolutional adversarial networks," *IEEE Transactions on Biomedical Engineering*, vol. 65, no. 12, pp. 2720–2730, 2018.

[112] S. Nowozin, B. Cseke, and R. Tomioka, "f-gan: Training generative neural samplers using variational divergence minimization," *Advances in neural information processing systems*, vol. 29, 2016.

[113] A. Odena, C. Olah, and J. Shlens, "Conditional image synthesis with auxiliary classifier gans," in *International conference on machine learning*. PMLR, 2017, pp. 2642–2651.

[114] H. Petzka, A. Fischer, and D. Lukovnicov, "On the regularization of wasserstein gans," *arXiv preprint arXiv:1709.08894*, 2017.

[115] G.-J. Qi, "Loss-sensitive generative adversarial networks on lipschitz densities," *International Journal of Computer Vision*, vol. 128, no. 5, pp. 1118–1140, 2020.

[116] A. Radford, L. Metz, and S. Chintala, "Unsupervised representation learning with deep convolutional generative adversarial networks," *arXiv preprint arXiv:1511.06434*, 2015.

[117] S. Ramani and J. A. Fessler, "A splitting-based iterative algorithm for accelerated statistical x-ray ct reconstruction," *IEEE transactions on medical imaging*, vol. 31, no. 3, pp. 677–688, 2011.

[118] M. L. P. Rani, G. Sasibhushana Rao, and B. Prabhakara Rao, "Ann application for medical image denoising," in *Soft Computing for Problem Solving: SocProS 2017, Volume 1*. Springer, 2019, pp. 675–684.

[119] L. J. Ratliff, S. A. Burden, and S. S. Sastry, "Characterization and computation of local nash equilibria in continuous games," in *2013 51st*

*Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE, 2013, pp. 917–924.

[120] S. Reed, Z. Akata, X. Yan, L. Logeswaran, B. Schiele, and H. Lee, "Generative adversarial text to image synthesis," in *International conference on machine learning*. PMLR, 2016, pp. 1060–1069.

[121] M. Ren, N. Dey, J. Fishbaugh, and G. Gerig, "Segmentation-renormalized deep feature modulation for unpaired image harmonization," *IEEE transactions on medical imaging*, vol. 40, no. 6, pp. 1519–1530, 2021.

[122] N.-C. Ristea, A.-I. Miron, O. Savencu, M.-I. Georgescu, N. Verga, F. S. Khan, and R. T. Ionescu, "Cytran: Cycle-consistent transformers for non-contrast to contrast ct translation," *arXiv preprint arXiv:2110.06400*, 2021.

[123] A. Rowland, M. Burns, T. Hartkens, J. Hajnal, D. Rueckert, and D. L. Hill, "Information extraction from images (ixi): Image processing workflows using a grid enabled image database," *Proceedings of DiDaMIC*, vol. 4, pp. 55–64, 2004.

[124] S. Roy, A. Carass, and J. L. Prince, "Magnetic resonance image example-based contrast synthesis," *IEEE transactions on medical imaging*, vol. 32, no. 12, pp. 2348–2363, 2013.

[125] H. Shan, A. Padole, F. Homayounieh, U. Kruger, R. D. Khera, C. Nitiwarangkul, M. K. Kalra, and G. Wang, "Competitive performance of a modularized deep neural network compared to commercial algorithms for low-dose ct image reconstruction," *Nature Machine Intelligence*, vol. 1, no. 6, pp. 269–276, 2019.

[126] H. Shan, Y. Zhang, Q. Yang, U. Kruger, M. K. Kalra, L. Sun, W. Cong, and G. Wang, "3-d convolutional encoder-decoder network for low-

dose ct via transfer learning from a 2-d trained network," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1522–1534, 2018.

[127] A. Sharma and G. Hamarneh, "Missing mri pulse sequence synthesis using multi-modal generative adversarial network," *IEEE transactions on medical imaging*, vol. 39, no. 4, pp. 1170–1183, 2019.

[128] L. Shen, W. Zhu, X. Wang, L. Xing, J. M. Pauly, B. Turkbey, S. A. Harmon, T. H. Sanford, S. Mehralivand, P. L. Choyke *et al.*, "Multi-domain image completion for random missing input data," *IEEE transactions on medical imaging*, vol. 40, no. 4, pp. 1113–1122, 2020.

[129] H.-C. Shin, A. Ihsani, S. Mandava, S. T. Sreenivas, C. Forster, J. Cha, and A. D. N. Initiative, "Ganbert: Generative adversarial networks with bidirectional encoder representations from transformers for mri to pet synthesis," *arXiv preprint arXiv:2008.04393*, 2020.

[130] V. Soumya, A. Varghese, T. Manesh, and K. Neetha, "Denoising multi-coil magnetic resonance imaging using nonlocal means on extended lmmse," in *Advances in Signal Processing and Intelligent Recognition Systems: Proceedings of Second International Symposium on Signal Processing and Intelligent Recognition Systems (SIRS-2015) December 16-19, 2015, Trivandrum, India*. Springer, 2015, pp. 187–198.

[131] P. Sudeep, P. Palanisamy, C. Kesavadas, and J. Rajan, "Nonlocal linear minimum mean square error methods for denoising mri," *Biomedical Signal Processing and Control*, vol. 20, pp. 125–134, 2015.

[132] C. Tian, Y. Xu, and W. Zuo, "Image denoising using deep cnn with batch renormalization," *Neural Networks*, vol. 121, pp. 461–473, 2020.

[133] J. Tian and L. Chen, "Image despeckling using a non-parametric statistical model of wavelet coefficients," *Biomedical Signal Processing and Control*, vol. 6, no. 4, pp. 432–437, 2011.

[134] C. Tomasi and R. Manduchi, "Bilateral filtering for gray and color images," in *Sixth international conference on computer vision (IEEE Cat. No. 98CH36271)*. IEEE, 1998, pp. 839–846.

[135] D.-H. Trinh, T.-T. Nguyen, and N. Linh-Trung, "An effective example-based denoising method for ct images using markov random field," in *2014 International Conference on Advanced Technologies for Communications (ATC 2014)*, 2014, pp. 355–359.

[136] S. Tulyakov, M.-Y. Liu, X. Yang, and J. Kautz, "Mocogan: Decomposing motion and content for video generation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1526–1535.

[137] D. Ulyanov, A. Vedaldi, and V. Lempitsky, "It takes (only) two: Adversarial generator-encoder networks," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 32, no. 1, 2018.

[138] D. Wang, Z. Wu, and H. Yu, "Ted-net: Convolution-free t2t vision transformer-based encoder-decoder dilation network for low-dose ct denoising," in *Machine Learning in Medical Imaging: 12th International Workshop, MLMI 2021, Held in Conjunction with MICCAI 2021, Strasbourg, France, September 27, 2021, Proceedings 12*. Springer, 2021, pp. 416–425.

[139] F. Wang, J. Li, Q. Yuan, and L. Zhang, "Local–global feature-aware transformer based residual network for hyperspectral image denoising," *IEEE Transactions on Geoscience and Remote Sensing*, vol. 60, pp. 1–19, 2022.

[140] G. Wang, J. C. Ye, K. Mueller, and J. A. Fessler, "Image reconstruction is a new frontier of machine learning," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1289–1296, 2018.

[141] J. Wang, T. Li, H. Lu, and Z. Liang, "Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose x-ray computed tomography," *IEEE transactions on medical imaging*, vol. 25, no. 10, pp. 1272–1283, 2006.

[142] J. Wang, H. Lu, T. Li, and Z. Liang, "Sinogram noise reduction for low-dose ct by statistics-based nonlinear filters," in *Medical Imaging 2005: Image Processing*, vol. 5747. SPIE, 2005, pp. 2058–2066.

[143] T.-C. Wang, M.-Y. Liu, J.-Y. Zhu, A. Tao, J. Kautz, and B. Catanzaro, "High-resolution image synthesis and semantic manipulation with conditional gans," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 8798–8807.

[144] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, Y. Qiao, and C. Change Loy, "Esrgan: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, pp. 0–0.

[145] Y. Wang, L. Zhou, L. Wang, B. Yu, C. Zu, D. S. Lalush, W. Lin, X. Wu, J. Zhou, and D. Shen, "Locality adaptive multi-modality gans for high-quality pet image synthesis," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2018: 21st International Conference, Granada, Spain, September 16-20, 2018, Proceedings, Part I*. Springer, 2018, pp. 329–337.

[146] Y. Wang, L. Zhou, B. Yu, L. Wang, C. Zu, D. S. Lalush, W. Lin, X. Wu, J. Zhou, and D. Shen, "3d auto-context-based locality adaptive multi-modality gans for pet synthesis," *IEEE transactions on medical imaging*, vol. 38, no. 6, pp. 1328–1339, 2018.

[147] Z. Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: from error visibility to structural similarity," *IEEE transactions on image processing*, vol. 13, no. 4, pp. 600–612, 2004.

[148] S. Watanabe, "Generative image transformer (git): unsupervised continuous image generative and transformable model for [$^{123}$i] fp cit spect images," 2022.

[149] J. M. Wolterink, A. M. Dinkla, M. H. Savenije, P. R. Seevinck, C. A. van den Berg, and I. Išgum, "Deep mr to ct synthesis using unpaired data," in *Simulation and Synthesis in Medical Imaging: Second International Workshop, SASHIMI 2017, Held in Conjunction with MICCAI 2017, Québec City, QC, Canada, September 10, 2017, Proceedings 2*. Springer, 2017, pp. 14–23.

[150] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Generative adversarial networks for noise reduction in low-dose ct," *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2536–2545, 2017.

[151] J. C. Wood and K. M. Johnson, "Wavelet packet denoising of magnetic resonance images: importance of rician noise at low snr," *Magnetic Resonance in Medicine: An Official Journal of the International Society for Magnetic Resonance in Medicine*, vol. 41, no. 3, pp. 631–635, 1999.

[152] D. Wu, K. Kim, G. E. Fakhri, and Q. Li, "A cascaded convolutional neural network for x-ray low-dose ct image denoising," *arXiv preprint arXiv:1705.04267*, 2017.

[153] W. Wu, J. Shi, H. Yu, W. Wu, and V. Vardhanabhuti, "Tensor gradient l-norm minimization-based low-dose ct and its application to covid-19," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–12, 2021.

[154] L. Xu, X. Zeng, H. Zhang, W. Li, J. Lei, and Z. Huang, "Bpgan: Bidirectional ct-to-mri prediction using multi-generative multi-adversarial nets with spectral normalization and localization," *Neural Networks*, vol. 128, pp. 82–96, 2020.

[155] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose x-ray ct reconstruction via dictionary learning," *IEEE transactions on medical imaging*, vol. 31, no. 9, pp. 1682–1697, 2012.

[156] H. Yang, J. Sun, A. Carass, C. Zhao, J. Lee, J. L. Prince, and Z. Xu, "Unsupervised mr-to-ct synthesis using structure-constrained cyclegan," *IEEE transactions on medical imaging*, vol. 39, no. 12, pp. 4249–4261, 2020.

[157] H. Yang, J. Sun, L. Yang, and Z. Xu, "A unified hyper-gan model for unpaired multi-contrast mr image translation," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part III 24*. Springer, 2021, pp. 127–137.

[158] Q. Yang, N. Li, Z. Zhao, X. Fan, E. I. Chang, Y. Xu *et al.*, "Mri cross-modality neuroimage-to-neuroimage translation," *arXiv preprint arXiv:1801.06940*, 2018.

[159] Q. Yang, P. Yan, Y. Zhang, H. Yu, Y. Shi, X. Mou, M. K. Kalra, Y. Zhang, L. Sun, and G. Wang, "Low-dose ct image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1348–1357, 2018.

[160] Z. Yi, H. Zhang, P. Tan, and M. Gong, "Dualgan: Unsupervised dual learning for image-to-image translation," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 2849–2857.

[161] H. Yin and S. Ma, "Csformer: Cross-scale features fusion based transformer for image denoising," *IEEE Signal Processing Letters*, vol. 29, pp. 1809–1813, 2022.

[162] B. Yu, L. Zhou, L. Wang, J. Fripp, and P. Bourgeat, "3d cgan based cross-modality mr image synthesis for brain tumor segmentation," in

*2018 IEEE 15th international symposium on biomedical imaging (ISBI 2018)*. IEEE, 2018, pp. 626–630.

[163] B. Yu, L. Zhou, L. Wang, Y. Shi, J. Fripp, and P. Bourgeat, "Ea-gans: edge-aware generative adversarial networks for cross-modality mr image synthesis," *IEEE transactions on medical imaging*, vol. 38, no. 7, pp. 1750–1762, 2019.

[164] ——, "Sample-adaptive gans: linking global and local mappings for cross-modality mr image synthesis," *IEEE transactions on medical imaging*, vol. 39, no. 7, pp. 2339–2350, 2020.

[165] H. Yu, M. Ding, and X. Zhang, "Laplacian eigenmaps network-based nonlocal means method for mr image denoising," *Sensors*, vol. 19, no. 13, p. 2918, 2019.

[166] L. Yu, W. Zhang, J. Wang, and Y. Yu, "Seqgan: Sequence generative adversarial nets with policy gradient," in *Proceedings of the AAAI conference on artificial intelligence*, vol. 31, no. 1, 2017.

[167] Y. Yu and S. T. Acton, "Speckle reducing anisotropic diffusion," *IEEE Transactions on image processing*, vol. 11, no. 11, pp. 1260–1270, 2002.

[168] D. Zhang and F. Zhou, "Self-supervised image denoising for real-world images with context-aware transformer," *IEEE Access*, vol. 11, pp. 14 340–14 349, 2023.

[169] H. Zhang, I. Goodfellow, D. Metaxas, and A. Odena, "Self-attention generative adversarial networks," in *International conference on machine learning*. PMLR, 2019, pp. 7354–7363.

[170] H. Zhang, T. Xu, H. Li, S. Zhang, X. Wang, X. Huang, and D. N. Metaxas, "Stackgan: Text to photo-realistic image synthesis with stacked generative adversarial networks," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 5907–5915.

[171] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising," *IEEE transactions on image processing*, vol. 26, no. 7, pp. 3142–3155, 2017.

[172] Z. Zhang, L. Yang, and Y. Zheng, "Translating and segmenting multimodal medical volumes with cycle-and shape-consistency generative adversarial network," in *Proceedings of the IEEE conference on computer vision and pattern Recognition*, 2018, pp. 9242–9251.

[173] J. Zhao, M. Mathieu, and Y. LeCun, "Energy-based generative adversarial network," *arXiv preprint arXiv:1609.03126*, 2016.

[174] M. Zhao, G. Cao, X. Huang, and L. Yang, "Hybrid transformer-cnn for real image denoising," *IEEE Signal Processing Letters*, vol. 29, pp. 1252–1256, 2022.

[175] X. Zheng, S. Ravishankar, Y. Long, and J. A. Fessler, "Pwls-ultra: An efficient clustering and learning-based approach for low-dose 3d ct image reconstruction," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1498–1510, 2018.

[176] B. Zhou, C. Liu, and J. S. Duncan, "Anatomy-constrained contrastive learning for synthetic segmentation without ground-truth," in *Medical Image Computing and Computer Assisted Intervention–MICCAI 2021: 24th International Conference, Strasbourg, France, September 27–October 1, 2021, Proceedings, Part I 24*. Springer, 2021, pp. 47–56.

[177] F. Zhu, G. Chen, and P.-A. Heng, "From noise modeling to blind image denoising," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 420–429.

[178] J.-Y. Zhu, T. Park, P. Isola, and A. A. Efros, "Unpaired image-to-image translation using cycle-consistent adversarial networks," in *Proceed-*

*ings of the IEEE international conference on computer vision*, 2017, pp. 2223–2232.

[179] Q. Zuo, J. Zhang, and Y. Yang, "Dmc-fusion: Deep multi-cascade fusion with classifier-based feature synthesis for medical multi-modal images," *IEEE Journal of Biomedical and Health Informatics*, vol. 25, no. 9, pp. 3438–3449, 2021.