Deep Learning Based CT Image Reconstruction

Ruiwen Xing

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2020

Reading Committee:

Dong Si

Thomas Humphries

Clark Olson

Program Authorized to Offer Degree:

Computing and Software Systems

University of Washington

Abstract

Deep Learning Based CT Image Reconstruction

Ruiwen Xing

Chair of the Supervisory Committee:
Dr. Dong Si
Computing and Software Systems

As a common medical imaging method, Computed Tomography (CT) can create tomographic images using X-ray data acquired from around the human body. However, high quality and adequately sampled X-ray measurement data are not always available. In this scenario, the tomographic image created by conventional reconstruction algorithms will be noisy, or contain artifacts. The goal of our study is to reconstruct high-quality tomographic images from noisy or incomplete scan data, including low-dose, sparse-view, and limited-angle scenarios, by utilizing novel deep learning techniques. In this project, we trained a Generative Adversarial Network (GAN) and used it as a signal prior in the Simultaneous Algebraic Reconstruction Technique (SART) iterative reconstruction algorithm. The GAN we trained includes a self-attention block to model long-range dependencies in the scan data. Compared with the state-of-the-art denoising

cycle GAN, CIRCLE GAN, and a conventional mathematical reconstruction algorithm incorporating total variation minimization, our Self-Attention GAN for CT image reconstruction produces competitive results on solving limited-angle data reconstruction problems. On sparse view and low-dose scenarios, our method is not always the best among compared methods, but produces competitive results in some situations.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

First and foremost, I would like to express my deepest gratitude and appreciation to Dr. Si and Dr. Humphries, for the continuous support throughout my study and research. Your kind guidance and feedback are the key for me to complete this thesis. I would also like to thank my committee member, Dr. Olson for his guidance and supervision.

I would like to thank all staff members at University of Washington Bothell Writing and Communication Center, for their consistent guidance and valuable suggestions on my thesis paper.

Last of all, I am always thankful to my family and friends for their encouragement throughout years of study.

Chapter 1. INTRODUCTION

Computed tomography (CT) is a common medical imaging method that generates tomographic images of the human body by using X-ray measurements taken from sensors around a patient. Due to the damage to human body from exposure to ionizing radiation (including X-rays), the US Nuclear Regulatory Commission states that the dose of radiation during exposure should be "As Low As Reasonably Achievable". This is also called the ALARA principle. However, in CT scans, measurements acquired at a lower radiation dose tend to be much noisier than measurements acquired at a higher radiation dose. Also, a lower number of measurements, or a smaller measurement acquisition arc, can cause a distorted image. To reduce the damage of radiation to the human body during scanning while maintaining image quality, CT image noise reduction algorithms are introduced to denoise CT images and remove artifacts. Three typical scenarios will be introduced in section 1.2.

In this project, we introduce a Generative Adversarial Network (GAN) into the image reconstruction process for noise reduction or image correction, providing a feasible method to generate a clear tomographic image using low dose scanning data.

In this section, an introduction of the principles of CT imaging, some current problems in CT imaging, and our proposed solution to those problems are provided. We present the general process of reconstructing images from CT data. Then, three CT imaging scenarios are introduced: low-dose, sparse view, and limited angle imaging. All these scenarios can have a negative effect on the CT image reconstruction process. We next discuss the reason why we want to combine Generative

Adversarial Network (GAN) and Simultaneous Algebraic Reconstruction Technique (SART). And finally, our project goal and benefits for stakeholders is defined in the last part of this section.

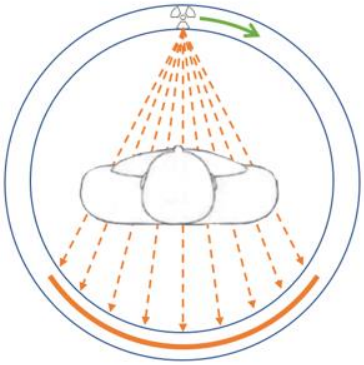
1.1 CT Scan Process

Since the introduction of CT scans in the 1970s, they have become an important medical imaging tool in diagnosis. It allowing us to non-invasively view the interior of an object. This ability results from a corollary of the central slice theorem: The density function of the one-dimensional projection of a two-dimensional image at various angles provides all the information about the image. This means that tomographic images can be reconstructed from a set of one-dimensional projection data that are coplanar with the tomographic image. Secondly, according to Beer's law, the concentration of an analyte is directly proportional to the amount of light absorbed. So, in order to obtain the projection data mentioned previously, X-rays are used in medical CT scans.

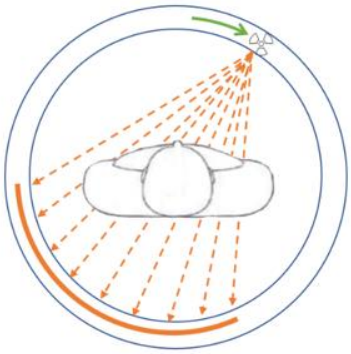
A typical process of a medical CT scan is shown in Figure 1.1. First, the patient will lay down on a hospital bed. Next, the CT machine moves the patient to find the correct position to scan (Figure 1.1(a)). Then, the X-ray source is activated and rotates around the patient. Meanwhile on the opposite side, X-ray detectors are positioned in a sector (Figure 1.1(b)). While the X-ray source rotates, detectors record the one-dimensional projection of the scanned patient. During this process, the projection data are recorded (Figure 1.1(c)). Once the camera system has finished rotating, all the one-dimensional projections are spliced, creating an image that looks like several sinusoidal images superimposed. This image is called a sinogram (Figure 1.1(d)), and is the most common organization form of raw CT scan data. Finally, an image reconstruction algorithm is applied to the sinogram, creating a tomographic image of patient body (Figure 1.1(e)).

(a) *

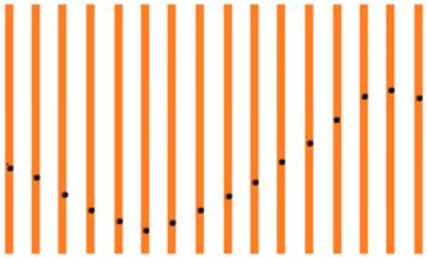
(b)

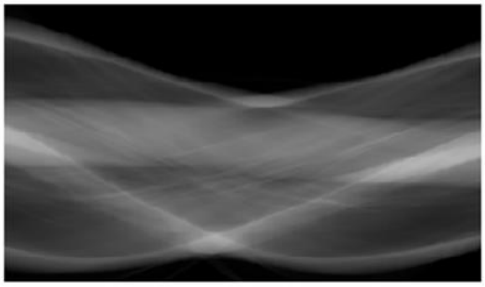


(c)



(d)





(e)

* https://www.mayoclinic.org/tests-procedures/ct-scan/about/pac-20393675

Figure 1.1. Typical process of medical CT scan.

1.2 CHALLENGES IN CT IMAGING

While medical CT scans are used every day in clinical practice with excellent results, several challenging problems remain. These include:

1) Low-dose imaging: In recent years, rising concerns about the effects of radiation on the human body have made low-dose CT scans a more and more popular means of physical examination. Low-dose imaging can be simply achieved by reducing the initial intensity of the X-ray beam. Since the measurements follow a Poisson distribution, lowering the intensity worsens the signal to noise ratio. As a result, the measurement on the sensor side will be noisier, reducing the quality of final the reconstructed image.

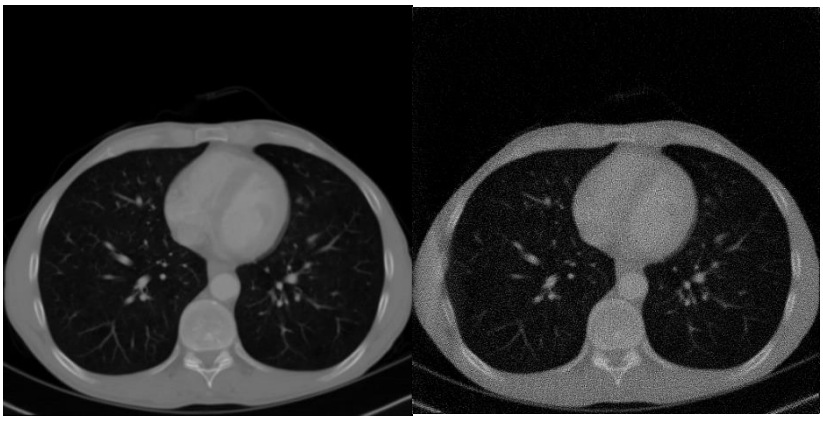


Figure 1.2. Example of normal dose (initial intensity of X-ray beam = 10^6 counts per measurement) and Low-dose (initial intensity of X-ray beam = 10^4 counts per measurement) reconstructed CT image.

2) Sparse-view imaging: Acquiring fewer measurements is another common way of reducing X-ray dose. In this scenario, the intensity of the X-ray beam remains normal to maintain projection quality, but the time of X-ray activations is reduced. If measurements are acquired every 0.2° around the patient in a standard CT imaging

protocol, for example, then in a sparse view protocol, measurements might be acquired in every 2°. In this way, the total dose of X-ray used is reduced, but the object being imaged is not sampled adequately. As a result, some ray-like lines will be seen on the reconstructed CT image.

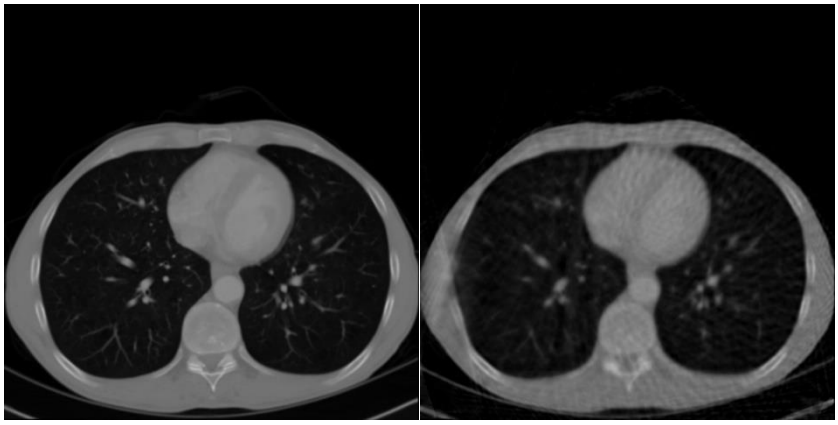


Figure 1.3. Example of standard view (900 view angles) and sparse view (60 view angles) reconstructed CT image.

3) Limited-angle imaging: To properly reconstruct a CT image, at least a 180° scan around the object is required. However, in certain situations, it is possible that the only available data is acquired along a shorter arc. As a result, information about some contents in the image is missing along certain lines, which will produce blurred edges along those directions.

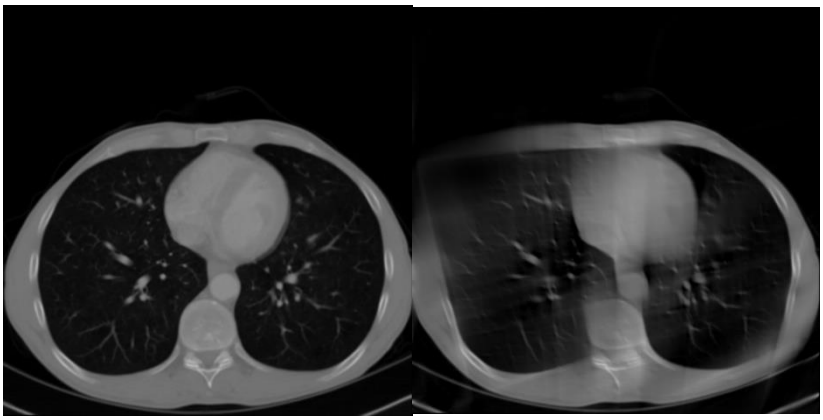


Figure 1.4. Example of full (360°) and limited angle (120°measurement) reconstructed CT image.

1.3 Integration of CT Image Reconstruction and Data Correction

As we have mentioned in the previous section, sinogram data collected from detectors is not always ideal: in low-dose scenarios, the data is inaccurate; in sparse view and limited angle scenarios, the data is insufficient. To create tomographic images from this imperfect data, some denoising or correction algorithms are needed.

Intuitively, we could apply a correction algorithm directly to sinogram data. This practice would act like a pre-processing step for image reconstruction. However, altering sinograms can cause serious problems. In a low-dose scenario, it is acceptable to treat sinogram as two-dimensional image and apply correction algorithm to denoise the sinogram. But in sparse view and limited angle scenarios, the correction algorithm needs to generate some artificial data to fill in the "missing" part of the sinogram. This can be a dangerous practice when the tomographic image is used to diagnose patients – perfect artificial data may cover up the presence of a lesion.

Another way to reconstruct tomographic images from imperfect data is to apply a correction algorithm to those images. This practice will act like a post-processing step for image reconstruction. The advantage is that it keeps the problem in normal two-dimensional image

domain. However, this practice allows noise or blank area in sinograms to affect the image reconstruction process — When sinogram data is very noisy or incomplete, the reconstructed image quality will be poor, making it difficult to recover a high-quality image in post processing. Further, making corrections on final tomographic image, we usually risk losing details in the image. And by losing these details, the mapping relationships between the final tomographic image and the original sinogram data is not guaranteed.

In this project, we chose a middle course. We design and train a Generative Adversarial Network (GAN) to make corrections to tomographic images, and use an iterative algorithm, Simultaneous Algebraic Reconstruction Technique (SART), to reconstruct tomographic images. The GAN is applied to the intermediate images generated during SART, working as a signal prior that guide the optimization process of SART. It provides feedback between tomographic images and original sinogram data. In this way, the problem domain of the noise reduction algorithm still remains in CT image space, while the final tomographic image will agree with sinogram data.

1.4 Project Opportunities

In many existing works applying neural networks to CT imaging, the network is trained only on a low-dose, sparse view or limited angle scenario. Given the computational effort required to train deep neural networks, it is of considerable interest to investigate whether a single network can be trained to improve image quality under a variety of imaging scenarios. Inspired by the architecture of Chang et al. [1][1], we develop a convolutional auto-encoder-based generative adversarial network. The network is trained on a mixed dataset of low-dose, sparse-view, and limited-angle data, and employed as a signal prior between iterations of SART, a well-known iterative reconstruction method for CT. The proposed approach is compared with state-of-the-art denoising

cycle GAN, CIRCLE GAN, and a mathematical approach employing total variation minimization. Our approach has better performance on reconstructing limited angle CT images, and reaches comparable performance on reconstructing low-dose and sparse view CT images.

1.5 PROJECT GOAL

The goal of this project was to create an algorithm that combines SART image reconstruction algorithm and Generative Adversarial Network. This algorithm was designed to reconstruct clean CT image using low-dose, sparse view, and limited angle sinogram data without retraining the neural net when different types of noisy or insufficient data sinograms were applied.

The goal of our project was that the result should be comparable to state-of-the-art CT image reconstruction methods in at least one scenario and demonstrate the ability of the network to solve reconstruction problems in low-dose, sparse view, and limited angle scenario.

Chapter 2. BACKGROUND AND RELATED WORKS

This section provides background knowledge and recent research about CT imaging. Four CT image reconstruction algorithms are discussed. Related previous research directions on CT imaging and CT image noise reduction is provided. Several research articles and their limitations are analyzed. Finally, we discuss the opportunity and advancement of our work.

2.1 CURRENT METHODS OF CT IMAGE RECONSTRUCTION

To reconstruct a tomographic image from a sinogram, a reconstruction algorithm is required. There are three main approaches to doing the reconstruction:

1) Fourier Transform

The first approach is using Fourier Transform. The central slice theorem states that the Fourier Transform of a one-dimensional projection is equal to the two-dimensional Fourier Transform of the whole tomographic image, sliced through its origin parallel to the projection line. Thus, in order to reconstruct a tomographic image, we can first Fourier Transform all one-dimensional projection data, then arrange these transformed data by scanning angle. Finally, we perform a two-dimensional Inverse Fourier Transform on these arranged data. The result will be the desired tomographic image.

The Fourier Transform approach is the most intuitive mathematical application of the central slice theorem, but it is not easy to implement in practice. This is mainly because the transformation between continuous function domain and discrete pixel positions will introduce systematic error.

2) Filtered Back Projection (FBP)

The second approach is Filtered Back Projection, which is based on the back-projection operation. In back projection, each sample point in the sinogram is back projected by setting all the image pixels in the tomographic image along the ray pointing to the sample to the sample value. The final pixel value in tomographic image is the average of the back projected values from all rays passing through the pixel.

Back projection is simple, but has significant side effect: a dense area in tomographic image adds a small value evenly to other area of the image when being back projected from sinogram. However, when we set those values into pixels, the neighboring area of dense area will have a higher possibility being selected in back projection ray. As a result, the neighboring area will look denser than it should be, blurring the border of dense area. As we can see in Figure 2.1, sample points near the object tends to be much lighter, which makes the boundary of object blurry.

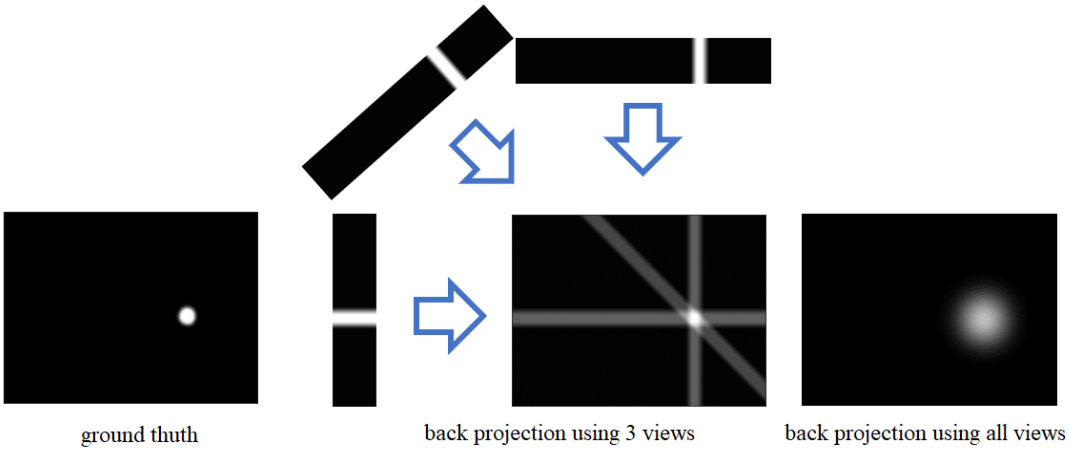


Figure 2.1. Principle of back projection.

To avoid this blurry border effect, it is necessary to reconstruct images using Filtered Back Projection. That is, for each sample point in sinogram, we apply a negative value filter to its

neighboring sample point. As shown in Figure 2.2, the filter offsets the blurry part and keeps the border sharp.

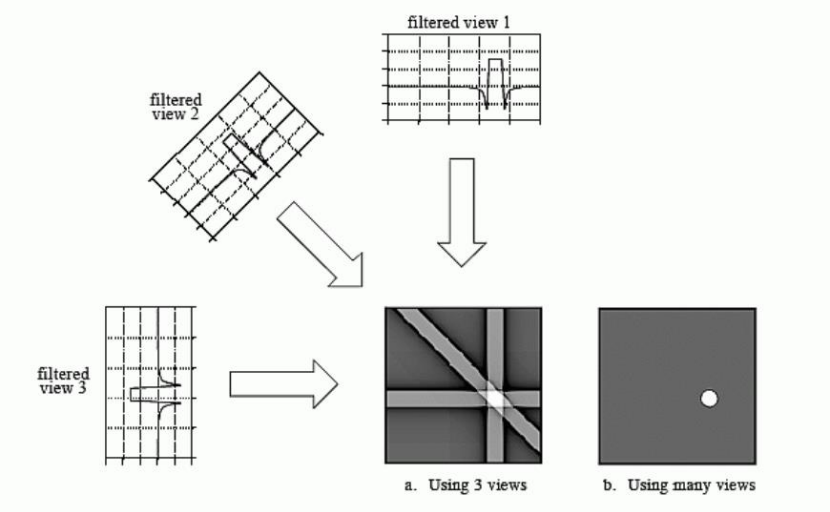


Figure 2.2. Filter back projection.[1]

As well as back projection, filter back projection has the advantage of being very fast. However, it relies on well sampled measurement data. When sparse view, or limited angle data is provided, the filter will set pixels in the area that may not be covered by measurements to negative values, leaving deep black artifacts in tomographic image. When low-dose noisy measurement data is provided, the reconstructed CT image will also be noisy.

3) Iterative Techniques

The last approach is using Iterative Techniques. According to Beer's Law, the relationship between the attenuation of X-ray beam intensity and the spatially varying attenuation function of the object being scanned can be defined as:

$$I = I_0 e^{(-\int_l \ \mu(y)dy)} \tag{1}$$

[1] http://www.dspguide.com/ch25/5.htm

where I_0 is the intensity of X-ray beam, I is the intensity of beam measured at the sensor side, l is the line traveled by the X-ray beam, and $\mu(y)$ represents the attenuation function. After log-transform, the equation can be described as:

$$\int_l \mu(y)dy = ln(\frac{I_0}{I}) \tag{2}$$

Collecting a total of j measurements and discretizing $\mu(y)$ as a two-dimensional image consisting of k pixels yields a linear system of equations:

$$b = Ax, \tag{3}$$

where j is the number of pixels and k is the number of measurements, $b \in R^j$ is the vector of measurements after log-transform (in our experiment, the sinogram), $x \in R^k$ is the reconstructed image, and A is a $j \times k$ system matrix. Matrix A is very large, but sparse, because one beam of X-rays intersects with only a relatively small number of pixels. It represents the so-called Radon Transform applied to a discretized image.

Iterative techniques are more time consuming compared with the previous three methods. However, iterative techniques approach creates an ideal clean image by making back projection and fix incorrect part of image step by step. They are more flexible for modeling physical effects and dealing with non-ideal data, giving us a chance to integrate advanced algorithms to reconstruct image when original sinogram data is insufficient. So, in this project, we choose an iterative technique, the Simultaneous Algebraic Reconstruction Technique (SART), to reconstruct tomographic images. Details of SART algorithm will be provided in section 3.4.

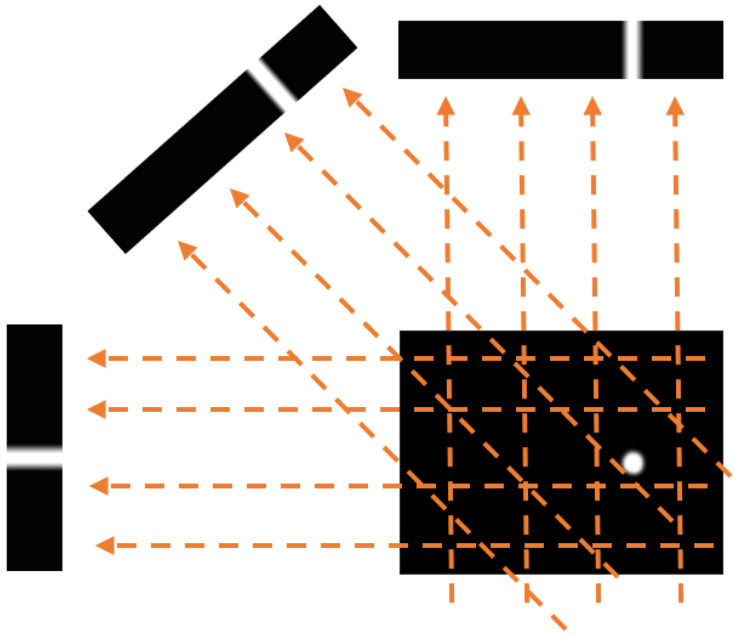


Figure 2.3. Principle of iterative techniques: Current image guess are forward projected to compared with original measurements.

2.2 Current Research on CT Imaging and CT Image Correction

People have realized for several decades that X-rays in CT scans can do harm to human bodies. This effect is particularly significant in patients with weakened immune systems [2], and these people are in need of CT scans the most. To help these people, researchers have been trying to reduce the radiation dose needed in CT scans while obtaining reliable medical images in a variety of ways. Some commercially available hardware solutions include automatic exposure control [3][4][5], automatic tube potential selection [6], beam-shaping filters, and dynamic z-axis collimators [7][8]. These techniques significantly reduce the radiation dose required for imaging, but require an update to the whole CT system, which does not make good use of existing medical equipment resources.

Historically, these types of problems have been addressed by not only hardware-based solutions, but also software-based approaches such as pre-processing or iterative reconstruction.

On the software side, pre-processing approaches to noise reduction typically involve filtering or smoothing the sinogram [9][10][11]. Alternatively, iterative reconstruction of low-dose, sparse-view and limited angle data may include regularization terms to improve image quality, which are typically based on a smoothing or sparsity-enforcing prior [12]-[16].

Over the last few years, there has been significant interest in applying techniques from deep learning to improve image quality in these imaging scenarios, as highlighted recently in a study from Wang et al. [17]. Chen et al. and Kang et al. used Convolutional Neural Networks (CNN) as a post-processing step to remove the noise in CT images generated by low-dose data, introducing this widely used technique for general image information extraction and optimization into CT image processing [18][19]. Cheng et al. also discussed the usage of the encoder-decoder structure in a CNN based low-dose CT denoising network [20]. Yang et al. and Wolterink et al. introduced Generative Adversarial Networks (GAN) into CT denoising and optimized the training process [21][22]. This work made good use of the characteristics of neural network algorithms, and has achieved state-of-art results. Other research has focused on reconstruction from sparse view data [23]. Zhang et al. used a combined structure of DenseNet and deconvolution [24] and Chen et al. worked on expert assessment-based reconstruction network [25]. In recent years, Gu et al. have also tried to solve limited angle reconstruction problem with neural network [26]. However, these works mostly focus on processing the final CT image without fully utilizing the feature that the texture of noise is highly related to the image reconstruction method.

Neural network algorithms have also been integrated into CT image reconstruction. Ma et al. use a simple three-layer neural network to reconstruct CT image with limited amount of

projection data [27], but their work mainly focuses on the reconstruction process itself.

An enormous problem in image correction, denoising, and reconstruction is the utilization of long-range dependency relationships. A single convolution layer can only utilize features within its kernel size; to utilize long range dependency relationships, a deeper convolutional neural net is required. Such networks may cause other issues such as vanishing gradients. A possible solution is developed from the natural language processing field. Vaswani et al. used *attention* to calculate the influence of remote morpheme [28]. Later, this concept was introduced into the field of image processing. Wang et al. developed a non-local neural net, arranging self-attention to each pixel in the image [29]. Zhang and Goodfellow et al. combined the non-local structure with a generative adversarial network and made the generator more sophisticated [30]. Their work motivated us to use the concept of self-attention GAN (SAGAN) in our project to denoise or correct CT image during reconstruction process.

Chapter 3. METHODS

This section discusses the software implementation of our experiment. Our runtime environment is provided first. Then our general structure design and detailed design is provided for each module.

3.1 SOFTWARE REQUIREMENT

This project is implemented in Python 3.6 due to related package dependency.

Some TensorFlow-related code needs to be modified when running on higher Python version environment.

The following Python packages are required:

- Numpy
- Scipy
- TensorFlow
- ASTRA toolkit
- Matplotlib
- Cuda
- Cudnn

Numpy and Scipy are common scientific computing packages in Python. Tensorflow, Cuda and Cudnn provide vector computing service to construct a neural network. ASTRA toolkit simulates CT scan projection and back projection process [31]. Matplotlib provides plotting functionality.

For better performance, we run our code on CentOS 7.5.1804 with NVIDIA GeForce GTX 1080 Ti GPU.

3.2 General structure

In this project, we propose to combine the process of CT image reconstruction with the process of image correction and noise reduction to reconstruct high quality CT images in sparse view, low dose radiation, and limited angle scenarios. Based on Zhang and Goodfellow's research on self-attention GAN [30], a framework with Generative Adversarial Network (GAN) and Simultaneous Algebraic Reconstruction Technique (SART) is introduced. This structure is used to train a single network capable of dealing with problems in all three scenarios.

3.3 Generative adversarial network

Generative Adversarial Network (GAN), introduced by Ian Goodfellow et al. in 2014 [32], is a class of artificial intelligence algorithms widely used in machine learning. A typical GAN structure includes two neural networks that contest with each other – one network generates target data and the other tries to distinguish it from ground truth. In this way the performance of these two networks is continually improved. Compared to classical straightforward neural networks, the discriminator network makes GAN able to handle highly abstract data generation problems. Because of its abilities to generate data, GAN is widely used in image processing problems. It has significant advantages over other network structures in tasks like image synthesis, semantic image editing, and style transfer. [33]

In this project, the GAN we plan to use has a generator network to generate a clean image from a reconstructed image from low-dose/sparse view/limited angle data, and a discriminator network to evaluate the generated image. As a generative model, GAN is designed to create new data based on information provided, rather than remove or extract data [33]. Therefore, it is more

suitable than ordinary neural networks for reconstruction tasks. The detailed structure inside our GAN model will be discussed in Section 3.5.

3.4 SART

Simultaneous Algebraic Reconstruction Technique (SART) [34][35][36] is one of the most well-known and widely used iterative algorithms in the field of CT image reconstruction. It starts from an initial image guess $x^{(0)}$ (we use a blank image in our experiments) and iteratively solves the linear system defined by CT measurements (the sinogram). In each SART iteration, we apply forward projection to image $x^{(i)}$ in a group of selected angles. Then, these measurements are compared with related measurements in sinogram. We calculate the difference and back project these difference to $x^{(i)}$ and create $x^{(i+1)}$. Then we take the projection image $x^{(i+1)}$ as current image in the next iteration. Because the calculation of vertical projection to certain measurement is linear, the whole algorithm is also linear.

The SART algorithm is defined as follows [37]:

First, two diagonal matrices D and M are defined:

$$D_{k,k} = \frac{1}{\sum_{i=1}^{J} |a_{i,k}|} ; M_{j,j} = \frac{1}{\sum_{i=1}^{K} |a_{j,i}|} \quad (4)$$

where D is a $k \times k$ matrix whose diagonal entries are the reciprocals of the column sums of A, and M is a $j \times j$ matrix whose entries are the reciprocals of the row sums.

Second, N_w equally spaced subsets are created from the sinogram (measurement data). These subsets are indexed by w. For example, subset w=1 contains the first, eleventh, twenty-first, etc. columns of the sinogram when N_w=10; subset w=2 contains the second, twelfth, twenty-second, etc. This approach is called block iteration. It accelerates the convergence of the algorithm

by updating $x^{(i)}$, using the average of difference. A full iteration of this block-iterative approach can then be described as

$$x^{(i+1)} = P_{\Omega}(B_{N_w} \dots B_2 B_1(x^{(i)})) \quad (5)$$

where,

$$B_w(x) = x - D_w(A_w)^T M_w(A_w x - b_w) \quad (6)$$

For each iteration, the *w* indicates that only the rows corresponding to subset w are used. P_{Ω} represents projection onto the feasible set of solutions (Ω). Provided that such a solution lies within Ω, this iteration can be shown to converge to a weighted least squares solution of the problem [35]. However, this solution is not considered desirable since it neglects noise in the sinogram and assumes the sinogram is well sampled. In cases of noisy or under sampled data, the original SART algorithm will create noisy or incomplete CT images. Thus, using prior information in the CT image reconstruction process is necessary.

3.5 CT IMAGE RECONSTRUCTION WITH GAN

In this thesis, we propose to combine the process of CT image reconstruction with CNN-based prior information to reconstruct high quality CT images in sparse view, low dose, and limited angle scenarios. Based on Chang et al.'s research on linear inverse problems [1], a GAN optimized SART framework is introduced.

The conceptual framework is shown in Figure 3.1. Our system contains two main modules – a GAN module and a SART module. The GAN module follows the general generative framework to train a neural network: The generator creates the clean image and the discriminator tries to discriminate between the generated clean CT image and the real high-dose, standard-view, full-angle clean CT image. After training, the generator network is connected to the SART module.

In the SART module, the generator network is applied to image x^i (in the i^{th} iteration) between each SART iteration and this guides the reconstruction direction. The pseudocode of GAN guided SART algorithm is shown in Figure 3.2

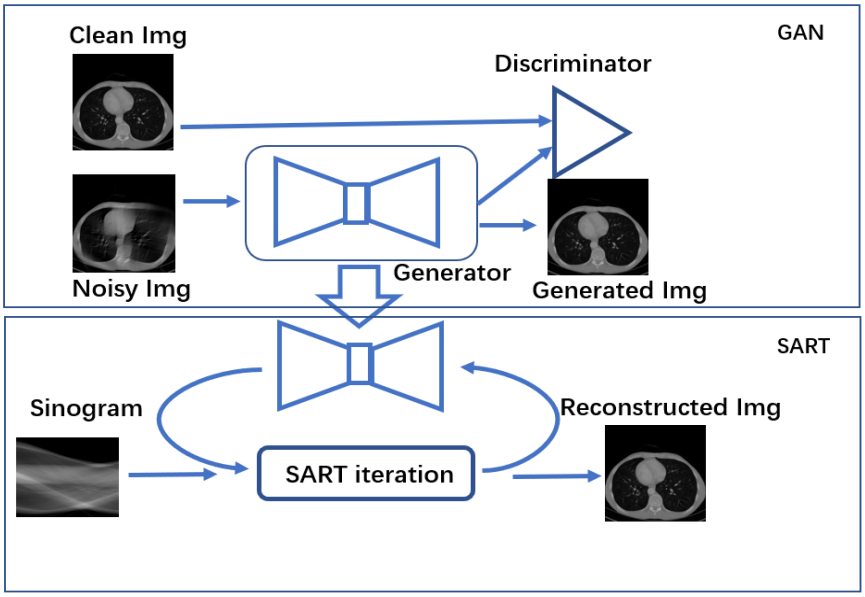


Figure 3.1. General structure of our proposed method.

GAN-SART Start
x = image
sino = sinogram
k = 0
β = 1.0

while k < maxIter:
-- // SART
-- for w = 0→subsetNum:
-- -- fp = Ax_k *// Forward projection, A is system matrix*
-- -- Δ = (sino[w] – fp)/sinoScale
-- -- bp = ΔA^T *// Back projection*
-- -- $x_{k+1} = x_k + \beta \times bp/imageScale$
-- k = k+1
-- // GAN
-- x_k = Generator(x_k)
return $x_{maxIter}$
End

Figure 3.2. Pseudocode of GAN guided SART.

Inside the generator network, we use a convolutional autoencoder structure. Ten convolution layers are used to encode and ten deconvolution layers are used to decode. During the encoding process, the size of the image is decreased and the color dimension of image is increased.

Inspired by recent work of Wang et al. [29] and Zhang and Goodfellow [30], we also apply the self-attention model in SAGAN [30] into our model. Unlike classical convolutional models that requires multiple layers to utilize long range dependencies, the self-attention model arrange attentions through the whole image, providing information of long-range dependency directly to recovered pixels. As shown in Figure 3.3, after encoding, a self-attention block is added to increase the ability of generator to recognize long range dependencies in image. To control the amount of data in the attention map, we assign one attention to each nine adjacent pixels. We apply an average pooling technique to these nine-pixel-blocks, and then calculate the self-attention value for these blocks.

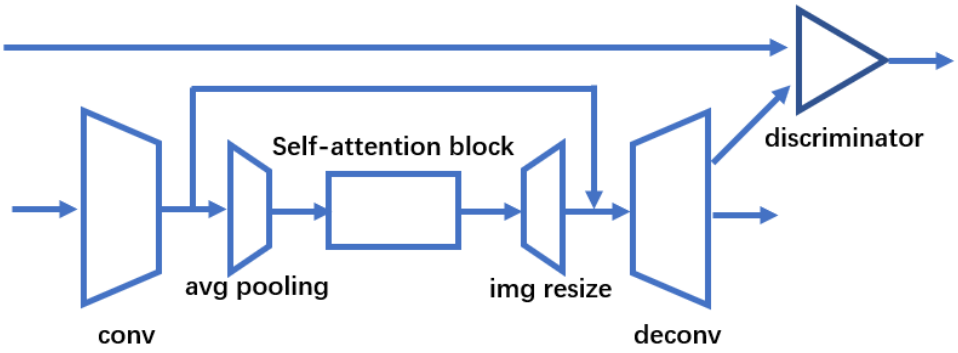


Figure 3.3. Structure of our GAN model.

The inner structure of encoding part of proposed generator is shown in Figure 3.4. One convolution layer, one batch normalization layer, and one ReLU layer form a block. The

convolution layer has 5*5 kernel, scanned with (1, 1) stride, and valid padding strategy. The first five blocks double the number of color channels.

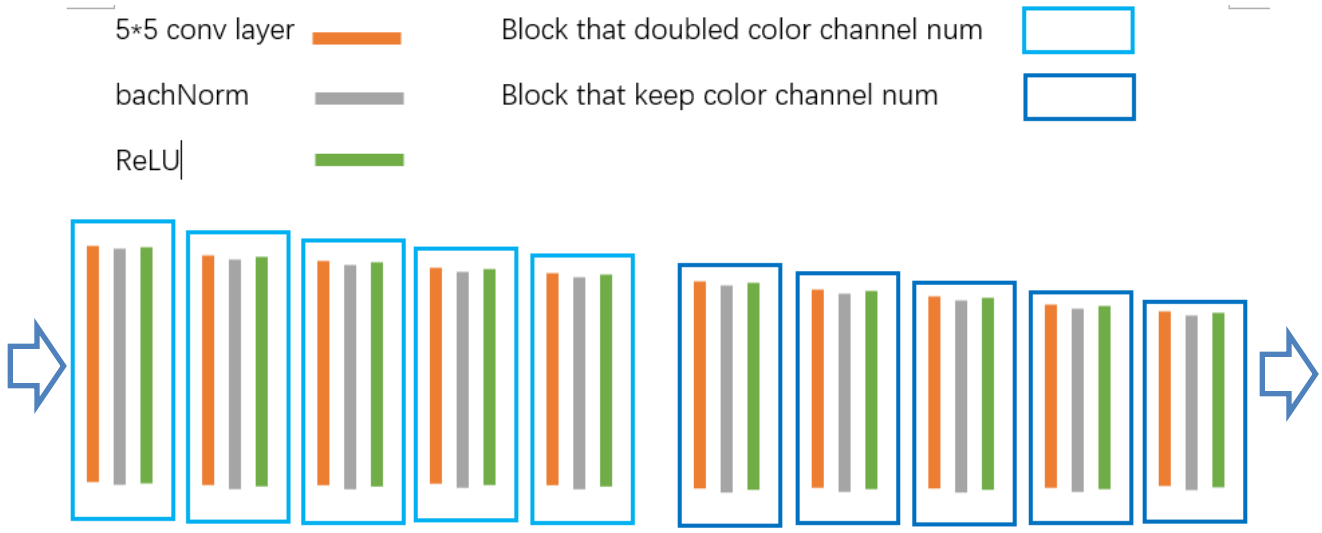


Figure 3.4. Encoding part of proposed generator.

The inner structure of decoding part of proposed generator is shown in Figure 3.5. One deconvolution layer, one batch normalization layer, and one ReLU layer form a block. The settings of deconvolution layers are same as settings of convolution layers in encoding part. The last five blocks reduce the color channel number to one at the end.

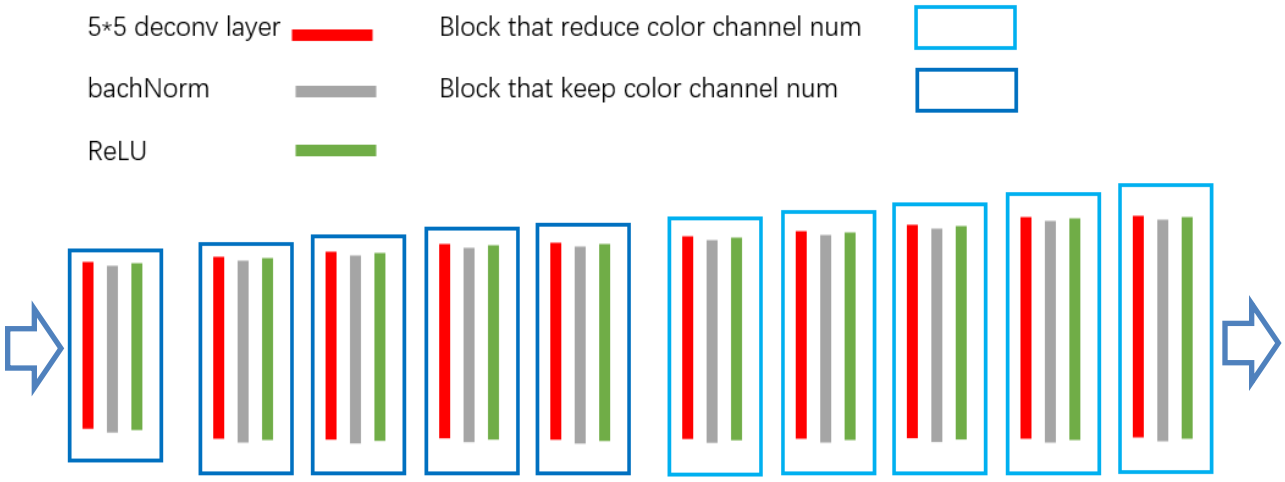


Figure 3.5. Decoding part of proposed generator.

The structure of the self-attention block is shown in Figure 3.6. In this block, the image feature tensor $x^{(m\times n\times c)}$ from the previous coded layer is transformed into two feature spaces *f* and *g*. Here, *m* and *n* are the number of lines and rows in image feature tensor *x*, and *c* is the dimension number of image feature. The feature spaces are generated as

$$f(x) = W_f x, \tag{7}$$

$$g(x) = W_g x, \tag{8}$$

where W_f and W_g are two trainable weight that applied to all features in *x*.

Then, *f(x)* and *g(x)* are flattened to two $(m*n)\times c$ matrices.

The attention map is calculated by

$$a = softmax(flatten(f(x))^T flatten(g(x))) \tag{9}$$

Here, the attention map *a* is a $(m*n)\times(m*n)$ matrix, where $a_{i,j}$ $(i\in Z^+ | 1\le i\le m,\ j\in Z^+ | 1\le j\le n)$ is the attention on the j^{th} image feature when we handle the i^{th} image feature. A total of $(m*n)^2$ attentions are calculated.

Afterwards, the attention map *a* is applied back to a transformed previous layer image feature *h(x)* by

$$o = \gamma \times reshape(flatten\big(h(x)\big)a) + x \tag{10}$$

where γ is a learnable scalar. Similar to *f* and *g*,

$$h(x) = W_h x \tag{11}$$

where W_h is a trainable weight, *x* is the input image feature tensor with shape $m\times n\times c$. The *flatten* function also reshapes *h(x)* to shape $(m*n)\times c$, and the reshape function recovers it to shape $m\times n\times c$ after attention map *a* is applied.

In this way, we arrange a set of attention weights for each pixel block. These attention weights provide information from pixels far away from the current pixel block, and thus give the model

the ability to handle long range dependencies in the image. A shortcut across the self-attention block is also created to preserve information from the previous layer.

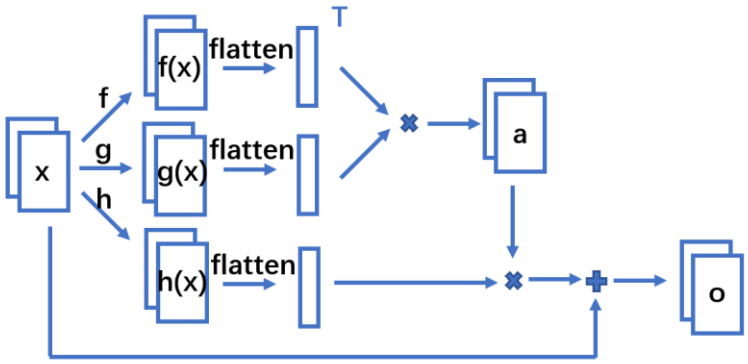


Figure 3.6. Structure of self-attention block. [29][30]

3.6 DISCUSSION ON STRUCTURE

Inspired by work of Rick Chang et al. [1], we designed a two-stage module that contains one neural network and one iterative algorithm. At the start of this project, we used the original iterative algorithm in Chang's work, the Alternating Direction Method of Multipliers (ADMM). However, this algorithm is designed to solve general linear problems. It does not utilize the principle of CT scan and requires a more sophisticated design compares to algorithms designed specifically for tomographic images reconstruction. So, we switch to SART algorithm in later experiments.

For the neural network part, we have tried simple straight forward CNN, straight forward CNN with shortcuts, and convolutional autoencoder structure for the generator of GAN. Our experiments showed that the autoencoder structure performs better than other two structures. After we read the work of Zheng et al. [30], we find that using self-attention in GAN could be a good solution to utilize long range dependencies in CT images. Thus, we add a self-attention block into our structure. The ablation study discussed in Section 4.3.4 showed that this design has also enhanced the performance of our method.

3.7 EXPERIMENT

3.7.1 *Data Preparation*

To train our neural net model, paired clean and noisy or incomplete CT images (including reconstructed image from low-dose, sparse view, and limited angle data) are needed. However, it is impractical to ensure that two real-life CT scans with different settings are in exactly the same position. Therefore, we simply use a set of normal-dose clean CT image as the ground truth, and use the ASTRA toolbox [38]-[39] to simulate the CT scan process with different settings to generate sinogram data.

The training data set consisted of 3900 CT image slices downloaded from the Cancer Imaging Archive[2]. 74.93% are images of lung, 18.34% are images of abdomen, and 6.73% are images of shoulder. Rotation was used to augment the dataset, resulting in 15600 different images. Then we used the ASTRA toolbox to generate low-dose, sparse view, and limited angle sinograms, and reconstruct them to noisy or incomplete CT images by applying the classical SART algorithm. For low-dose training set, we generate sinograms with intensity of X-ray beam $I_0 = 10^5$ and 10^4. For sparse view trainset, we generate sinograms with 100 and 50 measurements for each image. For limited angle trainset, we generate sinograms with 140 degrees measurement data. We apply 20 SART iterations to reconstruct incomplete or noisy images, with 50 subsets of measurement used in each iteration for all three cases mentioned above. Unless described otherwise, default setting for the sinograms are 180 degrees of data collected, 900 measurements for each sinogram

[2]https://www.cancerimagingarchive.net/

and 729 pixels for each measurement. Finally, the generated CT images were paired with original clean CT images and fed into our neural net for training.

3.7.2 *GAN Training*

The Adaptive Moment Estimation (Adam) [40] optimizer was used to optimize the model. During training, the learning rate is reduced. From epoch 1 to 7, learning rate is 0.0001. From epoch 8 to 11, learning rate is reduced to 0.00002. From epoch 12 to 15, learning rate is 0.00001. From epoch 16 to 20, learning rate is 0.000002. The initial learning rate is an experience-based choice on classical convolutional neural network and Adam optimizer. We then reduced it to one fifth of the original value, then half of the second value, then one fifth of the third value. L2 losses are used for both generator and discriminator. Smooth loss is utilized to reduce high frequency noise when we train the generator. It compares every pixel with its left and above pixel in the generated image. The generator loss can be defined as:

$$loss_G = \| Y - G(X) \|_2^2 + \left(\sum_{i=0}^{h}\sum_{j=0}^{w}(G(X_{i,j}) - G(X_{i-1,j}))^2 + \sum_{i=0}^{h}\sum_{j=0}^{w}(G(X_{i,j}) - G(X_{i,j-1}))^2\right) \quad (12)$$

and discriminator loss can be defined as:

$$loss_D = \| D(Y) - 1 \|_2^2 + \| D(G(X)) - 0 \|_2^2 \quad (13)$$

Where *h* and *w* are the height and width of image, *X* is input images from SART iterations, *Y* is ground truth images, *G* is the function of generator.

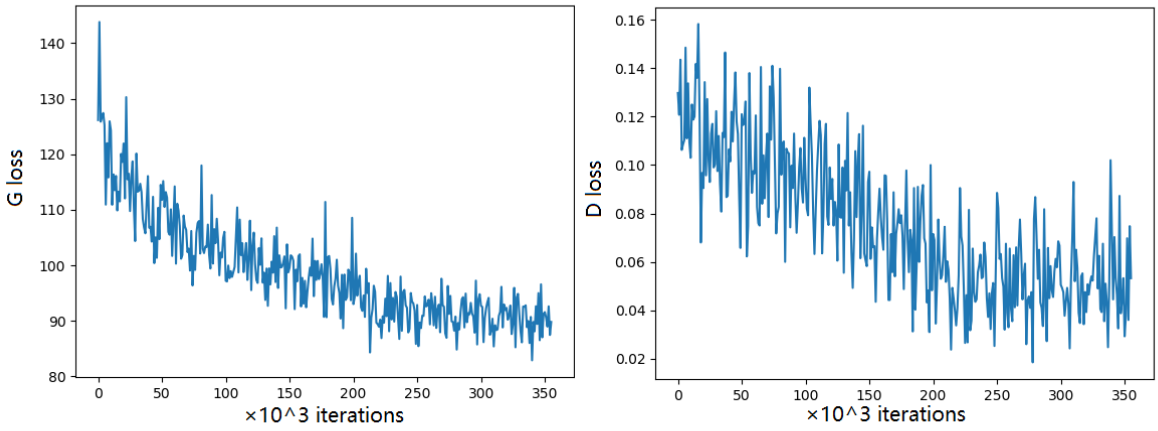


Figure 3.7. Generator loss and discriminator loss during training.

Figure 3.7 shows how generator loss and discriminator loss were reduced during training. The training is stopped after both generator loss and discriminator loss reached a stable stage. We can see from the figure, both generator loss and discriminator loss don't go down further after 250 $\times 10^3$ iterations.

Figure 3.8 shows the increase of the average peak signal-to-noise ratio (PSNR) of the validation set during training. The PSNR value (which will be explained in Section 4.1) measures the quality of our reconstructed CT image. The growth-stable curve of the validation result indicates that our model does not overfit and has reached a steady state. The significant growth in the middle part of the curve implies the reduction of learning rate has optimized the model.

To avoid overfitting, an early stopping strategy is used in our experiment. The training of neural network will stop when the generator loss of validation set bounce between a limited interval and does not goes down further for 10,000 iterations. The total training time is more than 48 hours, and testing time is about 5 minutes per image.

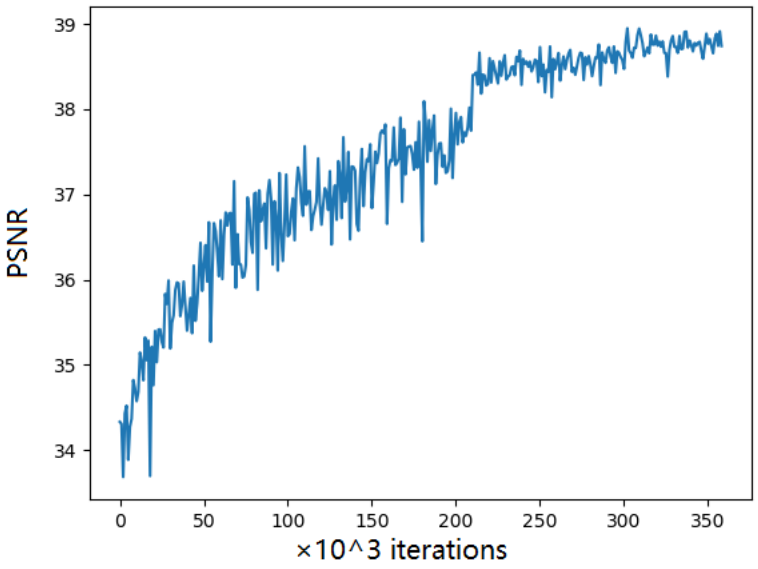


Figure 3.8. PSNR of validation set during training.

Chapter 4. RESULTS

This section describes the results of our experiments. First, in section 4.1, we discuss the method of evaluating the results. Then, in section 4.2, we give an introduction on three compared state-of-the-art algorithms: denoising cycle GAN, CIRCLE GAN, and SART-TV. In section 4.3, the test result of our algorithm on normal dose, low dose, sparse view, and limited angle CT data is compared with original SART reconstruction algorithm and three methods mentioned previously.

4.1 Evaluation Methods

The goal of this project is to use low-dose, sparse view, and limited angle CT data to generate a clean CT image, which can usually only be produced by high-dose, fully sampled data. In our project, we use a high-dose clean CT image as the ground truth. The better the performance of the algorithm, the higher should be the similarity between the reconstructed image and high dose image.

A few methods are used in Image Quality Assessment field to quantitatively describe the quality of a reconstructed image when ground truth is given. We choose two of the most commonly used algorithms, Peak Signal-to-Noise Ratio (PSNR) and Structural SIMilarity (SSIM) [41], to evaluate our reconstructed images.

4.1.1 *Peak Signal-to-Noise Ratio (PSNR)*

Peak Signal-to-Noise Ratio (PSNR) is the ratio between the maximum possible power of a signal and the power of corrupting noise that affects the fidelity of its representation. It is most commonly used to measure the quality of a reconstructed image. PSNR can be defined as:

$$PSNR = 10 \cdot log_{10}(\frac{MAX_I^2}{MSE}), \quad (14)$$

where MSE is the mean squared error, and MAX_I is the maximum possible pixel value in the evaluated image. Given a noise-free $m \times n$ monochrome image I and its noisy approximation K, the mean squared error can be defined as:

$$MSE = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}[I(i,j) - K(i,j)]^2 \quad (15)$$

4.1.2 *SSIM*

The structural similarity (SSIM) index is a method for predicting the perceived quality of digital television and cinematic pictures, as well as other kinds of digital images and videos [42].

The structural similarity can be defined as:

$$SSIM(x,y) = \frac{(2\mu_x\mu_y+c_1)(2\sigma_{xy}+c_2)}{(\mu_x^2+\mu_y^2+c_1)(\sigma_x^2+\sigma_y^2+c_2)} \quad (16)$$

Where μ_x is the average pixel value of image x, μ_y is the average pixel value of image y, σ_x is the variance of image x, σ_y is the variance of image y, σ_{xy} is the covariance of image x and image y, c_1 and c_2 are variables to stabilize the division with weak denominator.

4.2 COMPARED METHODS

To measure the performance of our method, we compare our results with a state-of-the-art denoising cycle GAN [43], a CIRCLE GAN [44], the original SART algorithm, as well as a "superiorized" approach [13], denoted SART-TV. A brief introduction of these methods is provided in this section.

4.2.1 *Denoising cycle GAN and CIRCLE GAN*

Denoising cycle GAN [43] and CIRCLE GAN [44] are two similar state-of-the-art neural networks based on cycle GAN. They treat clean images as one type of data and noisy images as another. Two generators and two discriminators are trained simultaneously in cycle GAN structure to convert images between clean and noisy. Differences between denoising cycle GAN and CIRCLE GAN are present within the structure of generator.

The general structure of denoising cycle GAN is shown in Figure 4.1.

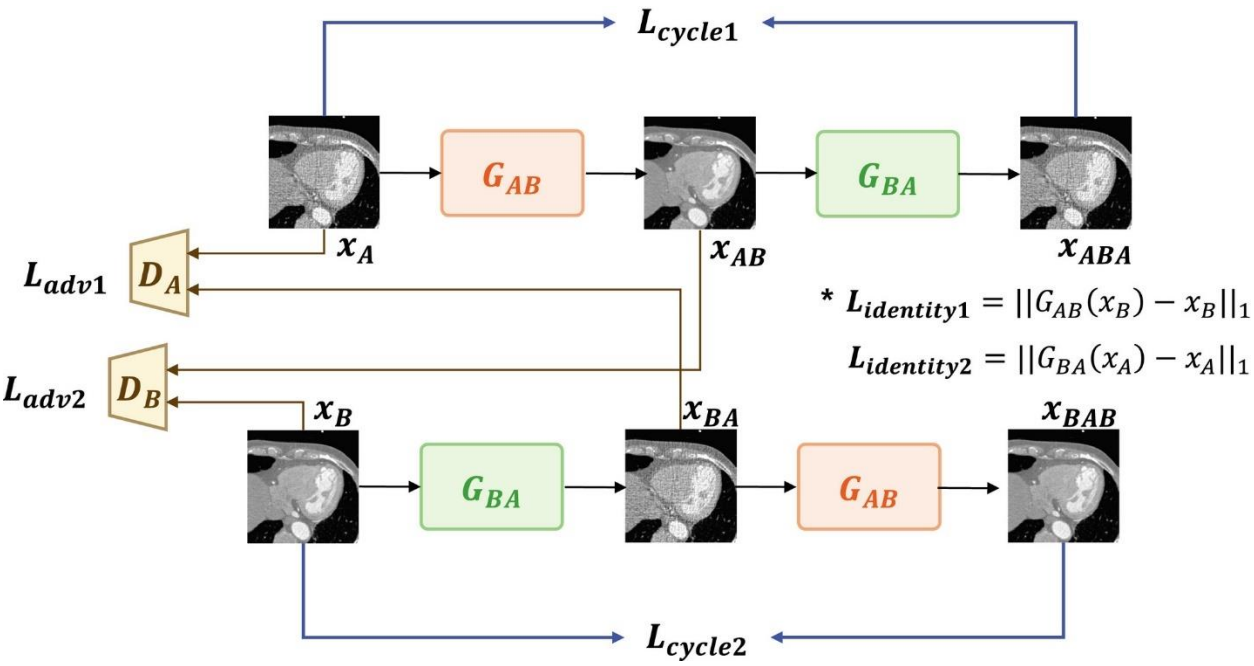


Figure 4.1. General structure of denoising cycle GAN. [43]

G_{AB} and G_{BA} are two generator networks with the same inner structure; D_A and D_B are two discriminator networks with the same inner structure. x_A is a low-dose noisy CT image; x_B is a normal dose clean CT image. In this structure, the author treats noisy images as status *A*, and clean images as status *B*. Generator G_{AB} transfer image from status *A* to status *B*, creating generated clean image x_{AB}. Generator G_{BA} does the reverse work and creates generated noisy image x_{BA}. Then, G_{BA}

and G_{AB} are applied to each other's generated result: x_{AB} is fed into G_{BA}, creating generated noisy image x_{ABA}. x_{BA} is fed into G_{AB}, creating generated clean image x_{BAB}.

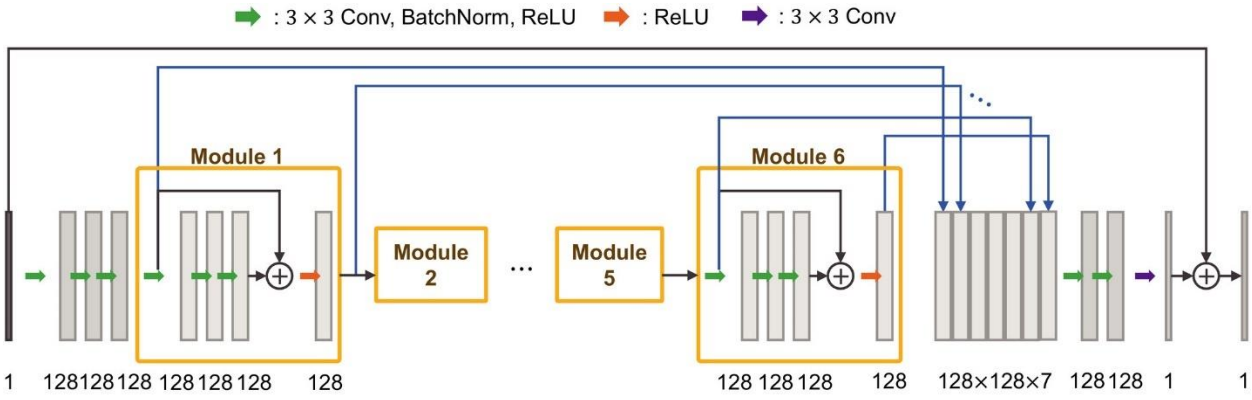


Figure 4.2. Inner structure of generator GAB and GBA in denoising cycle GAN. [43]

The inner structure of networks G_{AB} and G_{BA} is shown in Figure 4.2. The author used 3×3 convolution layers, a batch normalization layer, and a ReLU layer to build a module. Six modules were connected end to end, and the outputs of those modules were combined by convolution layer and generated final output.

Discriminator networks D_A and D_B are created to distinguish clean images from generated clean images, or distinguish noisy images from generated noisy images. It uses 4×4 convolution layers and Leaky ReLU layers to shrink data size. A fully connected layer at the end gives the output. The inner structure of discriminators D_A and D_B are shown in Figure 4.3.

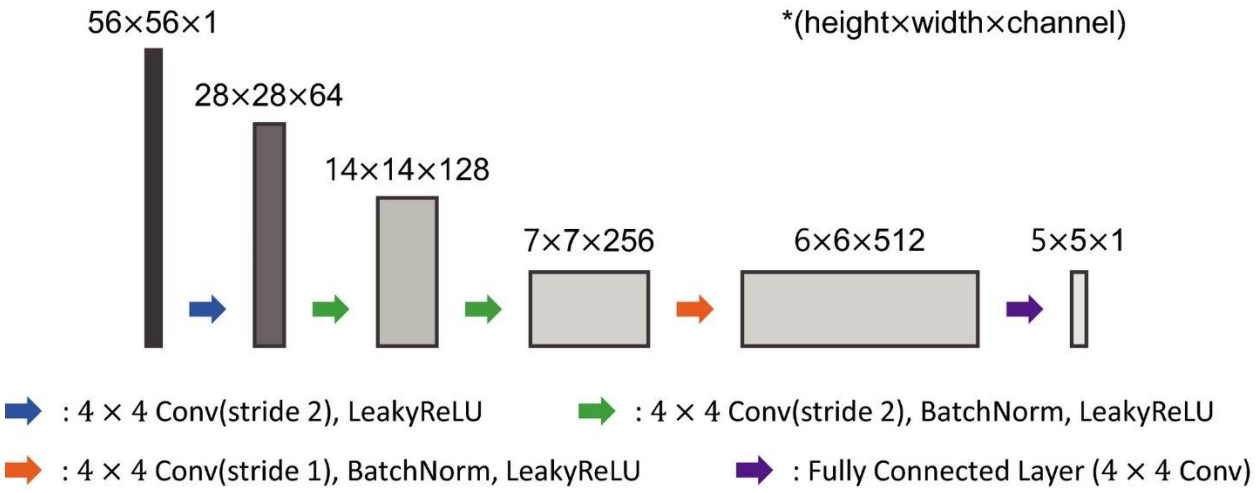


Figure 4.3. Inner structure of discriminator DA and DB in denoising cycle GAN. [43]

The advantage of using denoising cycle GAN is that noisy and clean images do not need to be paired during the training process. The cycle loss is calculated by comparing x_A and x_{ABA}, or x_B and x_{BAB}. The identity loss is calculated by comparing x_A and $G_{BA}(x_A)$, or x_B and $G_{AB}(x_B)$. Only one kind of image is needed in one training step. This characteristic gives denoising cycle GAN its ability to utilize large, unpaired datasets. Detailed design of denoising cycle GAN can be seen in reference [43].

The CIRCLE GAN [44] is a similar approach to denoising cycle GAN, but can be used on super resolution tasks. It also follows the basic design of cycle GAN: It regards two kinds of images as two statuses, X and Y. These two statuses can switch through generators G and F. Discriminators D_X and D_Y are designed to distinguish data from the dataset and data generated from the generators.

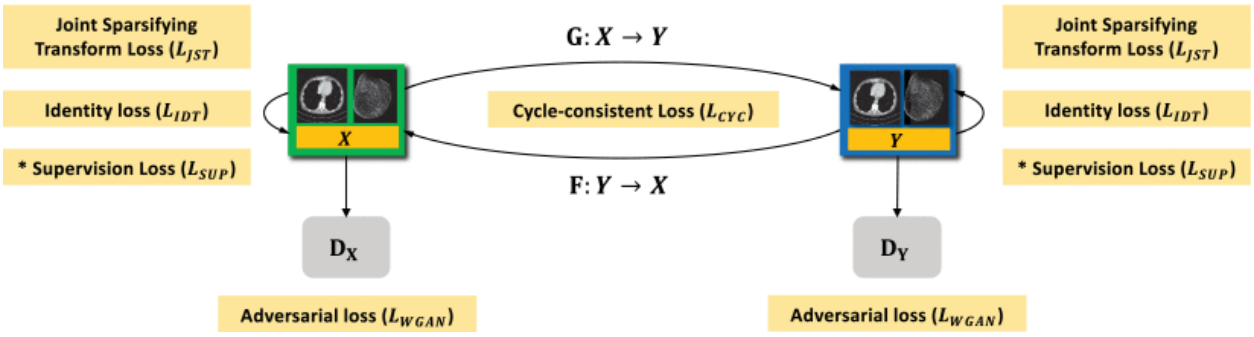


Figure 4.4. General structure of CIRCLE GAN. [44]

As shown in Figure 4.4, the author of CIRCLE GAN project combined 4 kinds of loss functions to train the CIRCLE GAN: adversarial loss (adv), cycle-consistency loss (cyc), identity loss (idt), and joint sparsifying transform loss (jst). These 4 losses regularize the training procedure. Definitions of these losses can be found in reference [44].

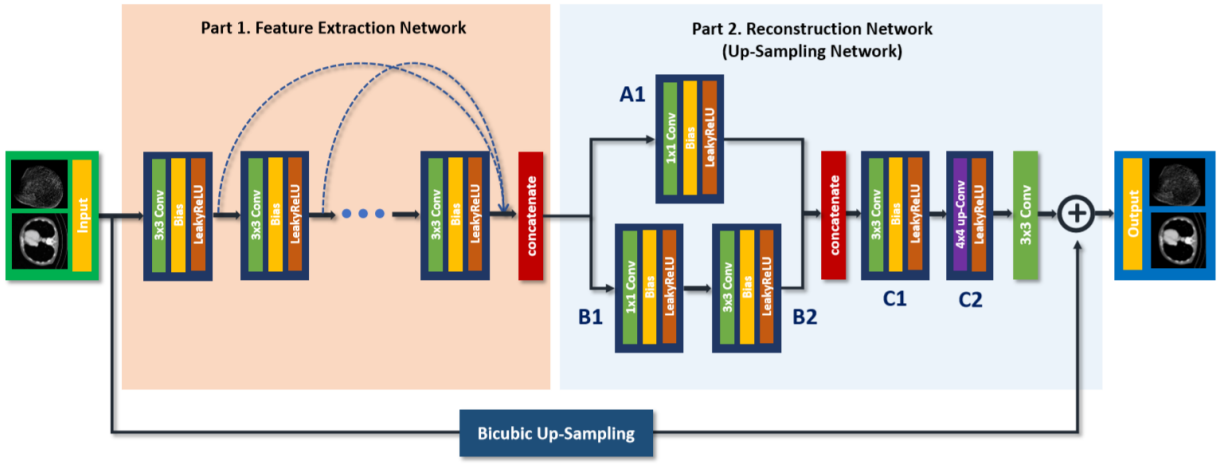


Figure 4.5. Inner structure of generator F and G in CIRCLE GAN. [44]

Figure 4.5 and Figure 4.6 show the inner structure of generators and discriminators in CIRCLE GAN. The generator network has two parts: Part 1 is a feature extraction network. It contains several blocks with 3×3 convolutional and leaky ReLU layers. The shape of the network is straightforward. All outputs of those blocks are combined at the end. Part 2 is a reconstruction network. It uses up sampling method to do the super resolution task. In our comparison, we adjust

all images to 512*512 pixels and calculate PSNR and SSIM.

As shown in Figure 4.6, the discriminator of CIRCLE GAN is also a straight forward network. Here, n stands for the number of convolutional kernels, and s stands for stride.

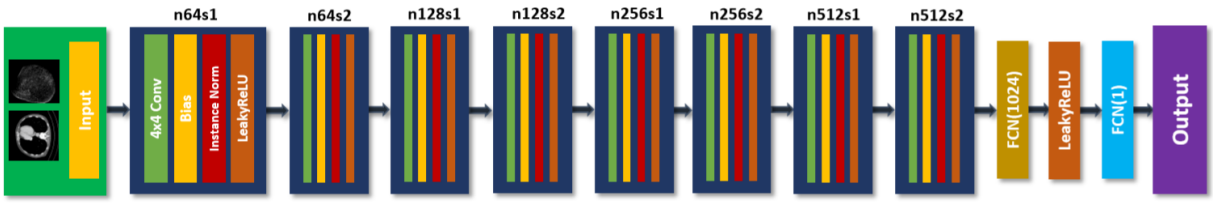


Figure 4.6. Inner structure of discriminator DX and DY in CIRCLE GAN. [44]

As described in the original paper of denoising cycle GAN and CIRCLE GAN, both neural networks work as postprocessing steps on reconstructed CT images, which means they do not participate in iterative reconstruction algorithms. Considering the side effect that applying a denoising neural network multiple times during SART iteration may over sharpen the output image, we followed the original design and applied these two networks only at the end of SART.

4.2.2 *SART-TV*

SART-TV is a mathematical iterative method for CT image reconstruction. It incorporates total variation (TV) minimization between each iteration of SART. The TV minimization mentioned here is a standard mathematical approach to handle low-dose and sparse-view CT imaging superiorization problems. It is designed to reduce noise and preserve edges in image at the same time.

Basically, superiorization is a kind of optimization heuristic applied to iterative algorithms like SART. It optimizes the result between iterations to produce solutions superior than original one based on secondary objective $\Phi(x)$. In an iterative method, we can define the iteration as:

$$x_{k+1} = R(x_k), \tag{17}$$

where x is the data and R is the iterative function defined in section 3.4 equation (5). Then, a superiorized iteration can be defined as:

$$x_{k+1} = R(x_k + \beta_k v_k), \tag{18}$$

where $\{\beta_k\}$ is a summable sequence of real numbers. It ensures the convergence of SART. [45] $\{v_k\}$ is a sequence of bounded perturbation vector. In this project, we choose $\beta_k = \gamma^k$ $(\gamma \in R \mid 0<\gamma<1)$, $v_k = -\frac{\nabla\Phi(x_k)}{\|\nabla\Phi(x_k)\|_2}$ Here, Φ is the penalty function that used to reduce total variation. By choosing v_k to be negative gradient of Φ, we are making the value of Φ decrease between iteration.

For a 2-dimensional image, the total variation function Φ can be defined as:

$$\phi_{TV}(x) = \sum_{(m,n)} \sqrt{(x_{(m+1,n)} - x_{(m,n)})^2 + (x_{(m,n+1)} - x_{(m,n)})^2 + \delta^2}, \tag{19}$$

where x is the image in form of 2-D array indexed by m and n, δ is a small parameter to ensure the differentiability of function at point $x_{(m,n)}$ when the value of its neighboring pixel is same to the value of itself. [45] Figure 4.7 shows the pseudocode of SART-TV algorithm.

SART-TV Start

x = image
sino = sinogram
α = 0.5
γ = 0.9995
N = 20
k = 0
β = 1.0

while k < maxIter:
-- // TV minimization
-- $v_k = -\frac{\nabla\Phi(x_k)}{\|\nabla\Phi(x_k)\|_2}$
-- $x_{k,0} = x_k$
-- for n = 0 → N:
-- -- while True:
-- -- -- $v_{k,n} = -\frac{\nabla\Phi(x_{k,n})}{\|\nabla\Phi(x_{k,n})\|_2}$
-- -- -- $z = x_{k,n} + \alpha \times v_{k,n}$
-- -- -- $\alpha = \alpha \times \gamma$
-- -- -- if $\Phi(z) < \Phi(x_k)$:
-- -- -- -- $x_{k,n+1} = z$
-- -- -- -- break
--
-- // SART
-- for w = 0→subsetNum:
-- -- fp = Ax_k // Forward projection, A is system matrix
-- -- Δ = (sino[w] – fp)/sinoScale
-- -- bp = ΔA^T // Back projection
-- -- $x_{k+1} = x_k + \beta \times bp/imageScale$
-- k = k+1

return $x_{maxIter}$

End

Figure 4.7. Pseudocode of SART-TV algorithm.

4.3 EXPERIMENTAL RESULTS

Our method aims to train a single neural net model to handle CT image artifacts caused by low-dose, sparse view, or limited angle sinogram data. While we train the network using a mixture of these three data types, 33% each, as described in Section 3.6.1. We assess the performance of our method on the three types of problem separately. We check the reconstructed image quality visually and use both PSNR and SSIM value to evaluate it quantitatively.

We test our data on low dose, sparse view and limited angle data generated from 160 CT images. As with the training data, these were generated by rotating 40 images from the Cancer Imaging Archive, and sinogram data were generated using the ASTRA Toolbox.

4.3.1 *Experiment Result on Limited Angle Data*

Table 4.1. PSNR evaluation for limited angle reconstruction.

PSNR	160 degrees	140 degrees	120 degrees
Proposed method	**38.55**	**36.72**	**31.81**
Denoising cycle GAN	37.96	34.69	29.07
CIRCLE GAN	38.07	35.09	29.53
SART-TV	36.03	30.95	27.66
Pure SART	34.72	29.93	27.40

Table 4.2. SSIM evaluation for limited angle reconstruction.

SSIM	160 degrees	140 degrees	120 degrees
Proposed method	**0.9863**	**0.9799**	**0.9680**
Denoising cycle GAN	0.9854	0.9775	0.9523
CIRCLE GAN	0.9801	0.9795	0.9533
SART-TV	0.9779	0.9555	0.9415
Pure SART	0.9637	0.9302	0.9007

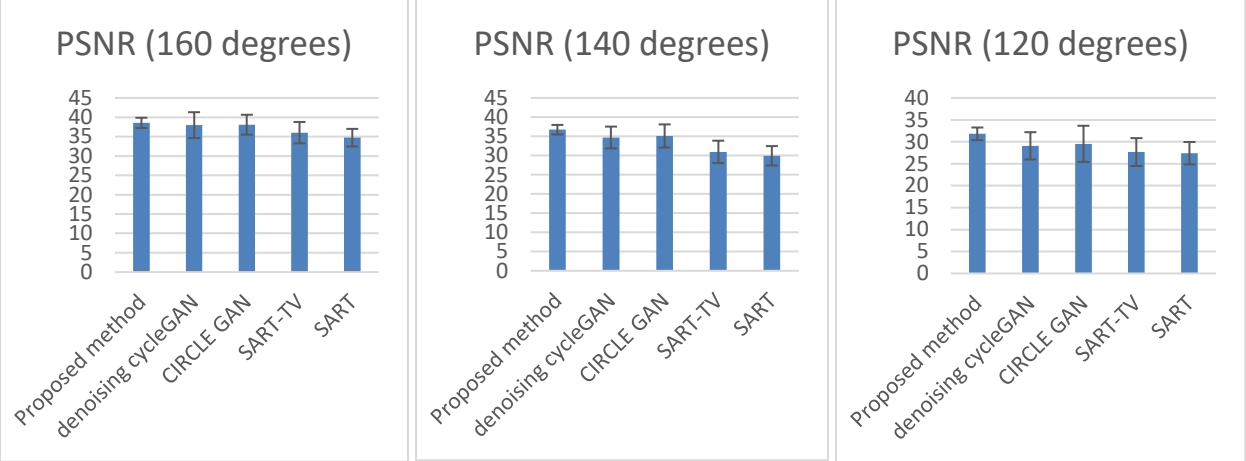


Figure 4.8. PSNR evaluation for limited angle reconstruction.

Table 4.1, Table 4.2, and Figure 4.8 shows the results of each method on 160, 140, and 120 degrees limited angle data. The proposed method gives better results than all other methods in each of the three cases. Generally, the improvement in performance of our method versus the others is greater when fewer angles are available in data. When evaluated by PSNR and compared with static parameter SART-TV algorithm, our approach gives 6.99% enhancement on 160 degrees group, 18.64% enhancement on 140 degrees group, and 15.00% enhancement on 120 degrees group. When compared with CIRCLE GAN, the second well-performed method, our approach gives 1% enhancement on 160 degrees group, 4% on 140 degrees group, and 7% on 120 degrees group. As shown in Figure 4.8, our proposed method is also have the lowest standard deviation,

Both the PSNR and SSIM results show that the reconstructions based on neural network have significantly better performance on limited angle imaging than SART-TV. The mathematical algorithm lacks the ability to recognize artifacts with long-range dependencies, and is only able to smoothen the long smear caused by insufficient data in the local area. Figure 4.9 shows a representative image reconstructed with each approach. As we can see, the proposed method performs better than other methods in this image.

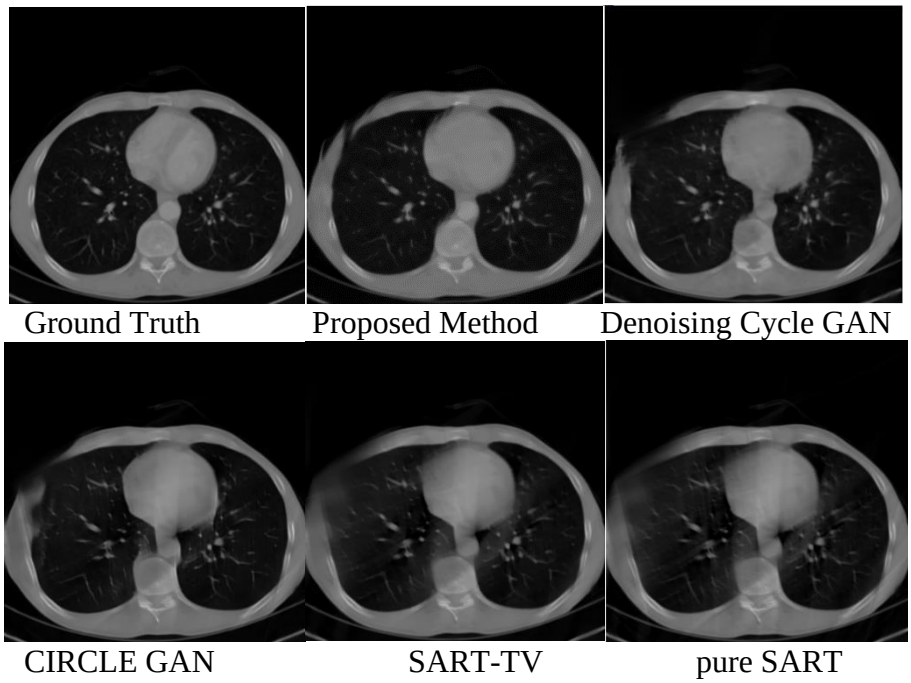


Figure 4.9. Reconstructed image in limited angle 120 degrees.

4.3.2 *Experiment Result on Low-Dose Data*

Table 4.3. PSNR evaluation for low-dose reconstruction.

PSNR	Noise level $I_0 = 10^6$	Noise level $I_0 = 10^5$	Noise level $I_0 = 10^4$
Proposed method	39.79	39.57	**35.26**
Denoising cycle GAN	**41.24**	40.66	26.62
CIRCLE GAN	40.80	**41.81**	34.73
SART-TV	40.53	38.59	33.42
Pure SART	37.07	29.64	20.70

Table 4.4. SSIM evaluation for low-dose reconstruction.

SSIM	Noise level $I_0 = 10^6$	Noise level $I_0 = 10^5$	Noise level $I_0 = 10^4$
Proposed method	0.9886	0.9874	**0.9571**
Denoising cycle GAN	**0.9907**	0.9879	0.7414
CIRCLE GAN	0.9905	**0.9914**	0.9479
SART-TV	0.9913	0.9861	0.9451
Pure SART	0.9783	0.9676	0.8781

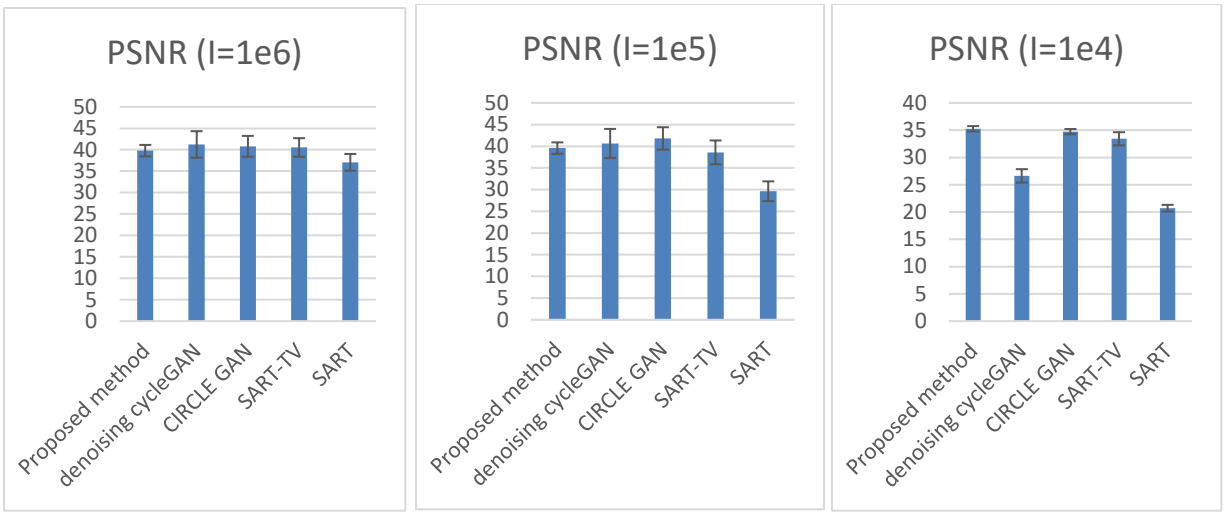


Figure 4.10. PSNR evaluation for low-dose reconstruction.

As shown in Table 4.3, Table 4.4, and Figure 4.10, the denoising cycle GAN and CIRCLE GAN methods work slightly better than the proposed method on lower noise level images (X-ray beam intensity $I_0 = 10^6$). The group with best performance is highlighted. At the highest noise level ($I_0 = 10^4$), however, their performance is diminished, while our approach has the best performance in those 4 compared methods. When evaluated by PSNR, our proposed method is 1% better than the second well-performed method, the CIRCLE GAN, in $I_0=10^4$ group. However, in $I_0=10^5$ and $I_0=10^6$ groups, CIRCLE GAN is 5% and 2% better than our method. But in these two groups, the consistency of our proposed method is better. The standard deviations of our method in these groups are 1.32 and 1.33, significantly lower than denoising cycle GAN (3.34 and 3.08), CIRCLE GAN (2.57 and 2.43), and SART-TV (2.75 and 2.17). Figure 4.11 shows reconstructed images, zoomed in to show fine detail (veins in the lungs).

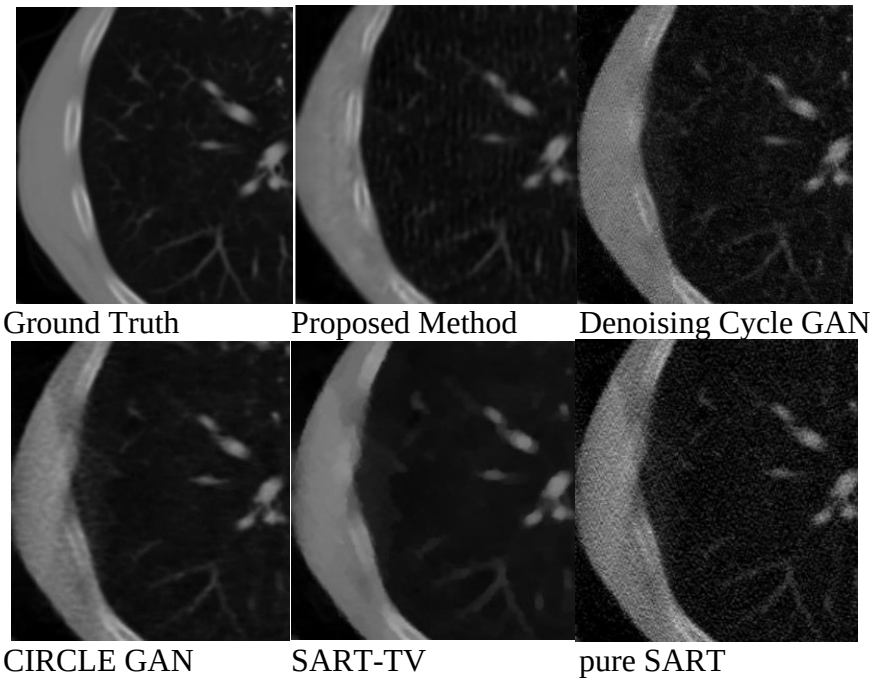


Figure 4.11. Reconstructed image in noise level $I_0 = 10^4$.

4.3.3 *Experiment Result on Sparse View Data*

Table 4.5. PSNR evaluation for sparse view reconstruction.

PSNR	100 views	60 views	50 views
Proposed method	39.31	**38.11**	36.55
Denoising cycle GAN	40.11	36.96	37.07
CIRCLE GAN	**40.33**	37.80	37.30
SART-TV	39.73	37.85	**37.80**
Pure SART	38.20	35.41	34.81

Table 4.6. SSIM evaluation for sparse view reconstruction.

SSIM	100 views	60 views	50 views
Proposed method	0.9868	**0.9833**	0.9744
Denoising cycle GAN	0.9877	0.9805	0.9752
CIRCLE GAN	**0.9886**	0.9809	0.9764
SART-TV	0.9883	0.9829	**0.9793**
Pure SART	0.9715	0.9550	0.9463

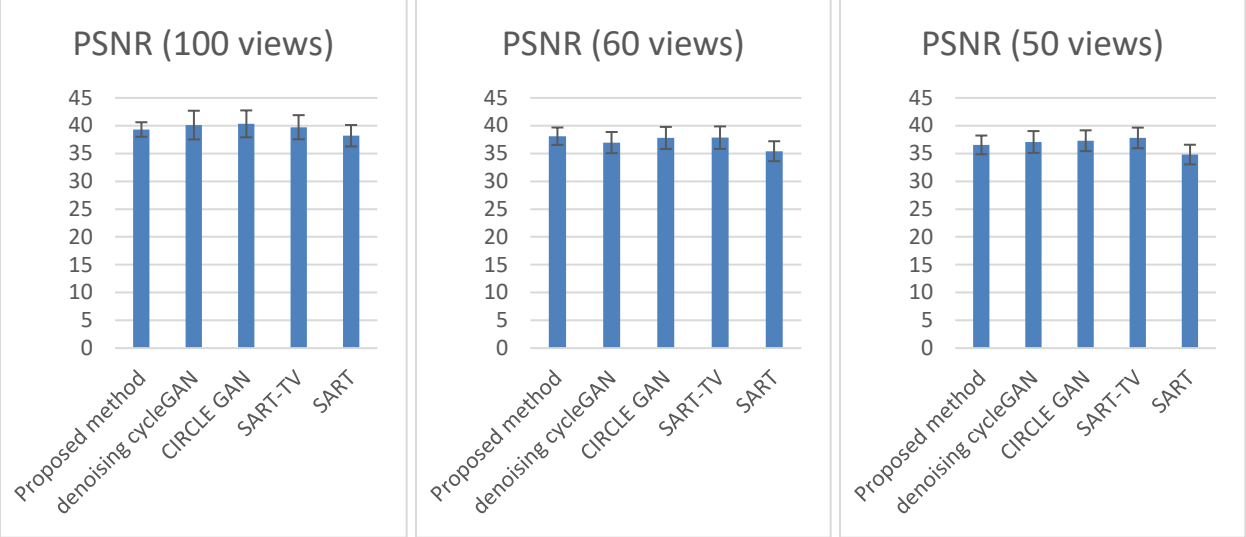
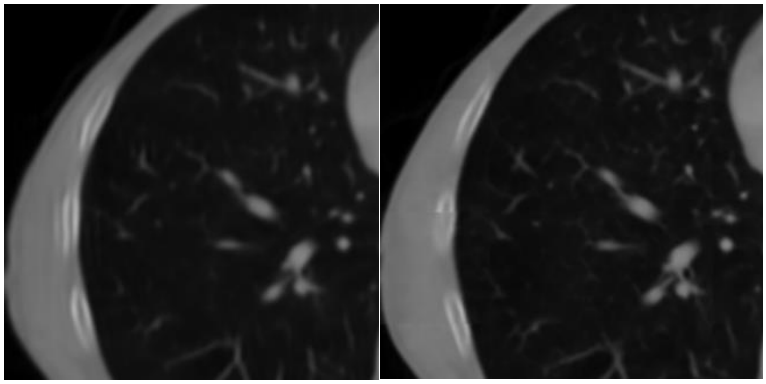


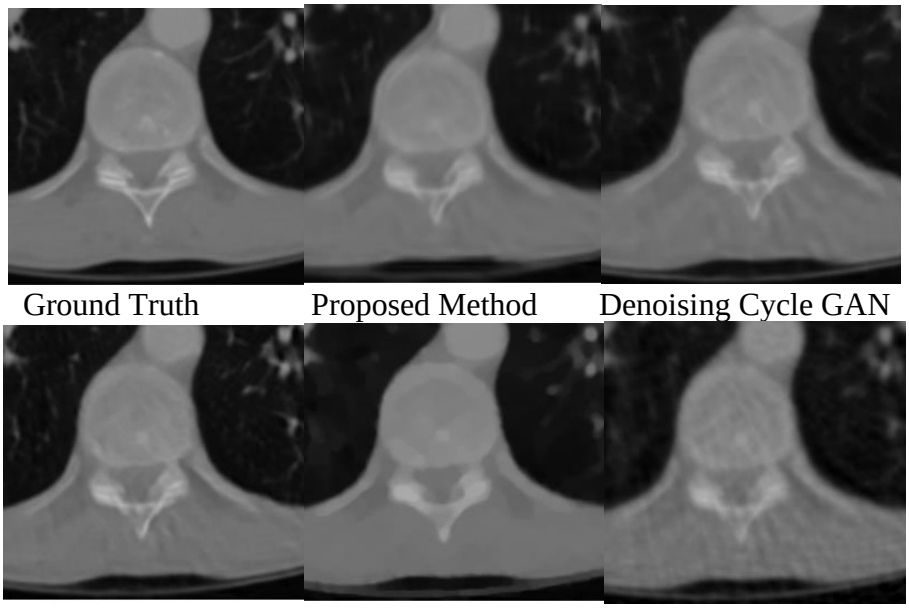
Figure 4.12. PSNR evaluation for sparse view reconstruction.

We tested the performance of each approach on sparse-view data using 100, 60, and 50 views. As shown in Table 4.5, Table 4.6, and Figure 4.12, the group with best performance is highlighted. In the 100-view group, the CIRCLE GAN gave 2.5% better performance on PSNR. And in the 50-view group, the SART-TV algorithm gave us a 3% better result on PSNR. As shown in Figure 4.13, in this case, our method generated some artifacts beside the white tissue. As the number of views was reduced, however, the performance of the proposed method was more consistent. The standard deviation of proposed method in 50 views group is 1.68, smaller than denoising cycle GAN (1.96), CIRCLE GAN (1.86), and SART-TV (1.85). The SART-TV approach was also competitive for sparse-view imaging, yielding (by a small margin) the best results for the 50-view case. Some reconstructed images are shown in Figure 4.14, zoomed in to show more details. The characteristic streaking artifacts caused by sparse-view data are still visible in the images reconstructed using cycle GAN and CIRCLE GAN, but have been mostly eliminated by the proposed approach.



Proposed Method CIRCLE GAN

Figure 4.13. Sample reconstruction image for sparse view 100 view dataset.



Ground Truth Proposed Method Denoising Cycle GAN

CIRCLE GAN SART-TV pure SART

Figure 4.14. Reconstructed image for sparse view 60 view dataset.

4.3.4 *Experiment Result on Self-Attention Block*

To test the effectiveness of using self-attention and utilizing long-range dependencies, we made an ablation study on the self-attention block; that is, we remove the self-attention block from our generator network. We train the generator network only with encoding and decoding layers and do the same test. Not surprisingly, the self-attention block does have a positive influence on denoising and recovering CT images. The result of ablation study on self-attention block is shown in Table 4.7 – Table 4.12.

Table 4.7. PSNR evaluation on proposed method for limited angle reconstruction.

PSNR	160 degrees	140 degrees	120 degrees
With self-attention	**38.55**	**36.72**	31.81
Without self-attention	37.65	36.36	**32.06**

Table 4.8. SSIM evaluation on proposed method for limited angle reconstruction.

SSIM	160 degrees	140 degrees	120 degrees
With self-attention	**0.9863**	0.9799	**0.9680**
Without self-attention	0.9843	**0.9809**	0.9679

Table 4.9. PSNR evaluation on proposed method for low-dose reconstruction.

PSNR	Noise level $I_0 = 10^6$	Noise level $I_0 = 10^5$	Noise level $I_0 = 10^4$
With self-attention	**39.79**	**39.57**	**35.26**
Without self-attention	39.01	38.90	33.77

Table 4.10. SSIM evaluation on proposed method for low-dose reconstruction.

SSIM	Noise level $I_0 = 10^6$	Noise level $I_0 = 10^5$	Noise level $I_0 = 10^4$
With self-attention	**0.9886**	**0.9874**	**0.9571**
Without self-attention	0.9870	0.9863	0.9119

Table 4.11. PSNR evaluation on proposed method for sparse view reconstruction.

PSNR	100 views	60 views	50 views
With self-attention	**39.31**	**38.11**	**36.55**
Without self-attention	38.26	36.72	35.81

Table 4.12. SSIM evaluation on proposed method for sparse view reconstruction.

SSIM	100 views	60 views	50 views
With self-attention	**0.9868**	**0.9833**	**0.9744**
Without self-attention	0.9848	0.9774	0.9722

As we can see in Table 4.7 – Table 4.12, the model with self-attention block performs better in nearly all scenarios, except limited angle 140 and limited angle 120. On these two scenarios, the evaluation results of PSNR and SSIM are inconsistent. Considering the different focus points of PSNR and SSIM, and those small differences on evaluation value, we believe that, on these two scenarios, model with self-attention and model without self-attention block give us very similar results. In general, using self-attention enhanced the performance of our neural net.

Chapter 5. CONCLUSION

This section discusses the conclusion of this project. An overview of our project is given, followed by our contributions to the CT image reconstruction problem. After that, we analyze the limitations of our algorithm and give some possible future directions.

5.1 OVERVIEW

In this paper, we combined the SART reconstruction algorithm and Generative Adversarial Network, creating a module to reconstruct and denoise or recover CT image from low-dose, sparse view, and limited angle sinogram data. In SART iterations, the generator of the GAN is used as a signal prior, helping us continually optimize the CT image during reconstruction process. In the generator network, we make use of the self-attention block in SAGAN, helping recognizing long range dependencies in noisy image. Pixel blocks are formed by each nine neighbor pixels. Attentions are calculated for each pixel block, providing remote position information to better reconstruct pixels in pixel blocks.

In order to test the effectiveness of our model, we compared it with a state-of-the-art CT image denoising models -- denoise cycle GAN, CIRCLE GAN, and a traditional mathematical optimization algorithm (SART-TV). The CT images reconstructed in three different scenarios are used in our experiment to ensure that we covered common situations in CT scans. The result showed that our method and other state-of-the-art methods have advantages on different scenarios and different influence level: For limited angle data, our model performs better on all influence levels. For low-dose data, the performance of our method is better than denoise cycle GAN when the noise level is very high, but not as good as cycle GAN when noise level is lower. Visually, the denoise cycle GAN performs better on image detail recovering. For sparse view data, the pure

mathematical optimization algorithm performs quite well. The score of our model is not as high as those compared algorithms, but still acceptable. This result means, our proposed method can provide competitive result to state-of-the-art methods in some of the common scenarios, like CT image reconstruction from limited angle data.

5.2 CONTRIBUTIONS

In this project, we made the following contributions.

- A generative adversarial neural network that learned the decision boundary to identify a clean and complete CT image is trained. For any CT image data feed into the generative network, it improves the quality of noisy or incomplete images during reconstruction.
- The concept of applying self-attention into generative adversarial network is utilized in our model. We use self-attention to identify and handle artifacts with long range dependence. Experiment results shows that this design is significantly helpful in reconstruction from limited angle data.
- We integrated the generative network into a Simultaneous Algebraic Reconstruction Technique (SART) algorithm. The generative network worked as a projection operator which can correct error during the SART process.

Our proposed method can handle CT image reconstruction problem when low-dose, limited angle, and sparse view sinogram is given. This feature reduces software complexity when the system is designed to reconstruct CT images from different data sources.

5.3 LIMITATIONS

Although our proposed method has showed significant improvement in some scenarios, like reconstruction from limited angle data, some limitations could still be seen.

First, our proposed projection network still has some difficulties in solving challenging problems, such as CT image reconstruction from less than 120 degrees limited angle data, which requires the algorithm make a random guess on tissue in unknown spaces. Although our method already has a better performance compared to other state-of-the-art methods in this scenario, the actual reconstructed CT image still lacks some information in the blurry part.

Second, our proposed reconstruction model still shows some bias on different scenarios. Experimental results show that it performs better than cycle GAN when limited angle data is given. However, the result of our model recovers less detail than cycle GAN when other types of data (such as low-dose data) is given. The reason is our model focuses more on the long-range dependencies while cycle GAN focus more on local detail. It is possible that a combined structure could take advantages of both and solve this problem.

Third, our proposed structure relies on an iterative SART method, which is often computationally expensive and time consuming compared with other neural-network-based noise reduction algorithms. This feature influenced us a lot during model training. We believed that specifically designed hardware could be a possible solution to handle this iteration problem.

5.4 FUTURE WORK

In general, future work may include training on larger dataset, adding different type and different level of noise or artifacts to sinogram, and try to use image patches on training. On top

of that, based on inspections on our current experiment, some potential directions have been found for us to work on.

Currently, only one neural network, the generator of GAN, is used during all SART reconstruction iterations. However, the effect of applying neural network in each iteration are different: In the first few iterations, the neural network mainly accelerates the SART process. While in the last few iterations, the neural network is trying to make the boundary in image clear and sharp to recover more details. So, it is possible that, we could apply different neural network on different SART reconstruction iterations. Refining the responsibility of the neural network may enhance the final performance.

Also, currently our work mainly focuses on optimizing the generator neural network part of the module. However, the SART iteration part also plays an important role in our structure. The basic setting of SART algorithm, like subset number, influences the speed of reconstruction. It could enhance or weaken the "strength" of SART.

According to our inspection, in the final few iterations, SART and GAN "fight against" each other, trying to control the optimization direction of image. The action of SART is based on some incomplete or noisy data, while the action of GAN is based on some general rules. We need to find a balance point between the strength of SART iteration and the strength of GAN to achieve its best performance. In our experiment, we discovered a group of intuitively feasible parameters. But a quantitative theoretical basis for these parameters is still needed in future research.

BIBLIOGRAPHY

[1] Rick Chang, J. H., Li, C. L., Poczos, B., Vijaya Kumar, B. V. K., & Sankaranarayanan, A. C. (2017). One Network to Solve Them All--Solving Linear Inverse Problems Using Deep Projection Models. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 5888-5897).

[2] Mole, R. H. (1990). Childhood cancer after prenatal exposure to diagnostic X-ray examinations in Britain. *British journal of cancer*, *62*(1), 152.

[3] Kalender, W. A., Wolf, H., Suess, C., Gies, M., Greess, H., & Bautz, W. A. (1998). Dose reduction in CT by anatomically adapted tube current modulation: principles and first results. In *Advances in CT IV* (pp. 27-34). Springer, Berlin, Heidelberg.

[4] Kalra, M. K., Maher, M. M., Toth, T. L., Schmidt, B., Westerman, B. L., Morgan, H. T., & Saini, S. (2004). Techniques and applications of automatic tube current modulation for CT. *Radiology*, 233(3), 649-657.

[5] McCollough, C. H., Bruesewitz, M. R., & Kofler Jr, J. M. (2006). CT dose reduction and dose management tools: overview of available options. *Radiographics*, 26(2), 503-512.

[6] Yu, L., Bruesewitz, M. R., Thomas, K. B., Fletcher, J. G., Kofler, J. M., & McCollough, C. H. (2011). Optimal tube potential for radiation dose reduction in pediatric CT: principles, clinical implementations, and pitfalls. *Radiographics*, 31(3), 835-848.

[7] Deak, P. D., Langner, O., Lell, M., & Kalender, W. A. (2009). Effects of adaptive section collimation on patient radiation dose in multisection spiral CT. *Radiology*, 252(1), 140-147.

[8] Christner, J. A., Zavaletta, V. A., Eusemann, C. D., Walz-Flannigan, A. I., & McCollough, C. H. (2010). Dose reduction in helical CT: dynamically adjustable z-axis X-ray beam collimation. *American journal of Roentgenology*, 194(1), W49-W55.

[9] Kalra, M. K., Maher, M. M., Sahani, D. V., Blake, M. A., Hahn, P. F., Avinash, G. B., ... & Saini, S. (2003). Low-dose CT of the abdomen: evaluation of image improvement with use of noise reduction filters—pilot study. *Radiology*, 228(1), 251-256.

[10] La Riviere, P. J. (2005). Penalized‐likelihood sinogram smoothing for low‐dose CT. *Medical physics*, 32(6Part1), 1676-1683.

[11] Wang, J., Li, T., Lu, H., & Liang, Z. (2006). Penalized weighted least-squares approach to sinogram noise reduction and image reconstruction for low-dose X-ray computed tomography. *IEEE transactions on medical imaging*, 25(10), 1272-1283.

[12] Sidky, E. Y., Kao, C. M., & Pan, X. (2006). Accurate image reconstruction from few-views and limited-angle data in divergent-beam CT. *Journal of X-ray Science and Technology*, 14(2), 119-139.

[13] Ritschl, L., Bergner, F., Fleischmann, C., & Kachelrieß, M. (2011). Improved total variation-based CT image reconstruction applied to clinical data. *Physics in Medicine & Biology*, 56(6), 1545.

[14] Xu, Q., Yu, H., Mou, X., Zhang, L., Hsieh, J., & Wang, G. (2012). Low-dose X-ray CT reconstruction via dictionary learning. *IEEE transactions on medical imaging*, 31(9), 1682-1697.
[15] Herman, G. T., Garduño, E., Davidi, R., & Censor, Y. (2012). Superiorization: An optimization heuristic for medical physics. *Medical physics*, 39(9), 5532-5546.
[16] Chen, Z., Jin, X., Li, L., & Wang, G. (2013). A limited-angle CT reconstruction method based on anisotropic TV minimization. *Physics in Medicine & Biology*, 58(7), 2119.
[17] Wang, G., Ye, J. C., Mueller, K., & Fessler, J. A. (2018). Image reconstruction is a new frontier of machine learning. *IEEE transactions on medical imaging*, 37(6), 1289-1296.
[18] Chen, H., Zhang, Y., Zhang, W., Liao, P., Li, K., Zhou, J., & Wang, G. (2017). Low-dose CT via convolutional neural network. *Biomedical optics express*, 8(2), 679-694.
[19] Kang, E., Min, J., & Ye, J. C. (2017). A deep convolutional neural network using directional wavelets for low‐dose X‐ray CT reconstruction. *Medical physics*, 44(10), e360-e375.
[20] Chen, H., Zhang, Y., Kalra, M. K., Lin, F., Chen, Y., Liao, P., ... & Wang, G. (2017). Low-dose CT with a residual encoder-decoder convolutional neural network. *IEEE transactions on medical imaging*, 36(12), 2524-2535.
[21] Yang, Q., Yan, P., Zhang, Y., Yu, H., Shi, Y., Mou, X., ... & Wang, G. (2018). Low-dose CT image denoising using a generative adversarial network with Wasserstein distance and perceptual loss. *IEEE transactions on medical imaging*, 37(6), 1348-1357.
[22] Wolterink, J. M., Leiner, T., Viergever, M. A., & Išgum, I. (2017). Generative adversarial networks for noise reduction in low-dose CT. *IEEE transactions on medical imaging*, 36(12), 2536-2545.
[23] Han, Y., & Ye, J. C. (2018). Framing U-Net via deep convolutional framelets: Application to sparse-view CT. *IEEE transactions on medical imaging*, 37(6), 1418-1429.
[24] Zhang, Z., Liang, X., Dong, X., Xie, Y., & Cao, G. (2018). A sparse-view CT reconstruction method based on combination of DenseNet and deconvolution. *IEEE transactions on medical imaging*, 37(6), 1407-1417.
[25] Chen, H., Zhang, Y., Chen, Y., Zhang, J., Zhang, W., Sun, H., ... & Wang, G. (2018). LEARN: Learned experts' assessment-based reconstruction network for sparse-data CT. *IEEE transactions on medical imaging*, 37(6), 1333-1347.
[26] Gu, J., & Ye, J. C. (2017). Multi-scale wavelet domain residual learning for limited-angle CT reconstruction. *arXiv preprint arXiv:1703.01382*.
[27] Ma, X. F., Fukuhara, M., & Takeda, T. (2000). Neural network CT image reconstruction method for small amount of projection data. Nuclear Instruments and Methods in Physics Research Section A: Accelerators, *Spectrometers, Detectors and Associated Equipment*, 449(1-2), 366-377.
[28] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998-6008).
[29] Wang, X., Girshick, R., Gupta, A., & He, K. (2018). Non-local neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 7794-7803).
[30] Zhang, H., Goodfellow, I., Metaxas, D., & Odena, A. (2018). Self-attention generative adversarial networks. *arXiv preprint arXiv:1805.08318*.

[31] van Aarle, W., Palenstijn, W. J., De Beenhouwer, J., Altantzis, T., Bals, S., Batenburg, K. J., & Sijbers, J. (2015). The ASTRA Toolbox: A platform for advanced algorithm development in electron tomography. *Ultramicroscopy*, 157, 35-47.

[32] Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Generative adversarial nets. In *Advances in neural information processing systems* (pp. 2672-2680).

[33] Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1), 53-65.

[34] Andersen, A. H., & Kak, A. C. (1984). Simultaneous algebraic reconstruction technique (SART): a superior implementation of the ART algorithm. *Ultrasonic imaging*, 6(1), 81-94.

[35] Censor, Y., & Elfving, T. (2002). Block-iterative algorithms with diagonally scaled oblique projections for the linear feasibility problem. *SIAM Journal on Matrix Analysis and Applications*, 24(1), 40-58.

[36] M. Jiang and G. Wang. Convergence of the simultaneous algebraic reconstruction technique (SART). *IEEE Transactions on Image Processing*, 12(8):957–961, 2003.

[37] T. Humphries, J. Winn and A. Faridani. Superiorized algorithm for reconstruction of CT images from sparse-view and limited-angle polyenergetic data. Phys. Med. Biol. 62(16), p. 6762. 2017.

[38] van Aarle, W., Palenstijn, W. J., Cant, J., Janssens, E., Bleichrodt, F., Dabravolski, A., ... & Sijbers, J. (2016). Fast and flexible X-ray tomography using the ASTRA toolbox. *Optics express*, 24(22), 25129-25147.

[39] Palenstijn, W. J., Batenburg, K. J., & Sijbers, J. (2011). Performance improvements for iterative electron tomography reconstruction using graphics processing units (GPUs). *Journal of structural biology*, 176(2), 250-253.

[40] Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

[41] Wang, Z., & Bovik, A. C. (2002). A universal image quality index. *IEEE signal processing letters*, 9(3), 81-84.

[42] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, 13(4), 600-612.

[43] Kang, E., Koo, H. J., Yang, D. H., Seo, J. B., & Ye, J. C. (2019). Cycle‐consistent adversarial denoising network for multiphase coronary CT angiography. *Medical physics*, 46(2), 550-562.

[44] You, C., Li, G., Zhang, Y., Zhang, X., Shan, H., Li, M., ... & Vannier, M. W. (2019). CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE). *IEEE Transactions on Medical Imaging*.

[45] Humphries, T., Loreto, M., Halter, B., O'Keeffe, W., & Ramirez, L. Comparison of Regularized and Superiorized Methods for Tomographic Image Reconstruction.

APPENDIX A

Readme file

Before run our code, the following python packages need to be installed:

Astra-toolbox	v 1.8.3
Cudatoolkit	v 8.0
CUDNN	v 7.1.3
Keras	v 2.2.4
Matplotlib	v 3.1.1
Numpy	v 1.14.2
Pillow	v 6.0.0
Tensorflow	v 1.10.0

To start a training process, type:

$ python main.py --function trainNN \

--cleanTrainset [trainset clean img path] \

--cleanTrainsetDataType [png or flt] \

--noisyTrainset [trainset noisy img path] \

--noisyTrainsetDataType [png or flt] \

--cleanTestset [validation set clean img path] \

--cleanTestsetDataType [png or flt] \

--noisyTestset [validation set noisy img path] \

--noisyTestsetDataType [png or flt] \

--checkpointFolder [path to save checkpoint] \

--batchSize [default is 1] \

--NNtype [neural net structure in NNstructure, such as NN721]

To test the model, type:

$ python main.py –function autoRecon \

```
--inputFolder [path of ground truth CT images] \
--dataType [png or flt] \
--sinoFolder [path to save sinogram] \
--noiseOption [scenarios, e.g. limited_angle_120] \
--mnnOrder [order to reconstruct image, e.g. ‘sart5|(gan)./checkpoint|sart20|return’] \
--outputFolder [path to save reconstructed image] \
--ns [subset number of SART]
```