2023

# PURSUIT OF ULTIMATE TOMOGRAPHIC IMAGE QUALITY: FROM CLASSICAL METHODS TO DEEP LEARNING

Shiwei Zhou

# PURSUIT OF ULTIMATE TOMOGRAPHIC IMAGE QUALITY:

# FROM CLASSICAL METHODS TO DEEP LEARNING

by

SHIWEI ZHOU

DISSERTATION

Submitted in partial fulfillment of the requirements
for the degree of Doctor of Philosophy at
The University of Texas at Arlington
August, 2023

Arlington, Texas

Supervising Committee:
      Dr. Mingwu Jin, Supervising Professor
      Dr. Wei Chen
      Dr. Yujie Chi
      Dr. Amir Shahmoradi
      Dr. Qiming Zhang

# ABSTRACT

PURSUIT OF ULTIMATE TOMOGRAPHIC IMAGE QUALITY: FROM

CLASSICAL METHODS TO DEEP LEARNING

Shiwei Zhou, PhD

The University of Texas at Arlington, 2023

Supervising Professor: Mingwu Jin

Medical imaging plays a crucial role in modern healthcare, serving as a vital component in the realms of diagnosis and treatment. It encompasses a broad spectrum of techniques and technologies aimed at visualizing the internal structure, physiology and bio-chemical processes inside the human body. Medical imaging has revolutionized medical practice by enabling doctors to diagnose diseases and monitor treatments without resorting to invasive procedures. Computed Tomography (CT) is an important tool of medical imaging. A CT scan employs computer-processed combinations of multiple X-ray images taken from different angles to generate cross-sectional images, providing significantly more detailed structural information compared to 2D X-rays. However, CT relies on ionizing radiation and its image quality can deteriorate due to patient motion and reduced imaging dose. In this work, we aim to improve CT image quality (at lower radiation dose) using advanced methods ranging from traditional modeling to deep learning.

In this work, we first developed a general simultaneous motion estimation and image reconstruction (G-SMEIR) method for 4D cone-beam CT (CBCT) to capture and model lung motion for radiation therapy (Chapter 2). It can overcome the local trapping problem of motion estimation and achieve better 4D CBCT image quality and motion tracking for lung tumors. Secondly, we developed several deep learning methods for CT denoising: cycle generative and adversarial network (CycleGAN) and RecycleGAN for unpaired single low-dose CT image denoising and unpaired low-dose CT image sequences denoising, respectively (Chapter 3), and texture transformer for super-resolution (TTSR) for low-dose CT (Chapter 4). These methods yield unprecedented denoising performance compared to other state-of-the-art denoising methods. This dissertation work not only provides multiple tools to address important issues in CT, but also demonstrates that advanced modeling and deep learning methods are effective in solving challenging problems in medical imaging.

# ACKNOWLEDGEMENTS

The journey that began when I first set foot on the UTA campus in 2018 has spanned almost five transformative years. The landscape of those years was dotted with a unique challenge, as the world grappled with the extraordinary circumstance of a global pandemic. Throughout this time, I learned to navigate the solitude that came with it, but I could not have weathered the storm without the aid of several individuals who were instrumental in shaping my journey.

My first and foremost gratitude goes to my advisor, Dr. Mingwu Jin. His unwavering support in both my research and personal life has been invaluable. His willingness to extend assistance in every possible way, without hesitation, provided a steady foundation for my work. There were moments of self-doubt, times when I questioned the meaningfulness of my research, but Dr. Jin was always there to guide and affirm me. He meticulously arranged each detail from foundational principles to the ultimate application in the project. Although I cannot say he is the world's best advisor, I can confidently say he is the best mentor I could have hoped to encounter.

My dissertation committee - Dr. Mingwu Jin, Dr. Wei Chen, Dr. Yujie Chi, Dr. Amir Shahmoradi, and Dr. Qiming Zhang - also warrant a heartfelt thank you. Their insightful advice during my comprehensive exam served as a compass, guiding the direction of my subsequent steps.

# TABLE OF CONTENTS

# LIST OF TABLES

# LIST OF FIGURES

# LIST OF SYMBOLS AND ABBREVIATIONS

| | |
|---|---|
| AI | Artificial Intelligence |
| A-P | Anterior-Posterior |
| ART | Algebraic Reconstruction Technique |
| CBCT | Cone-Beam Computed Tomography |
| CGLS | Conjugate Gradient Least Squares |
| CPU | Central Processing Unit |
| CT | Computed Tomography |
| CUFED | Curation Of Flickr Events Dataset |
| DNN | Deep Neural Network |
| DVF | Deformation Vector Field |
| FBP | Filtered Backprojection |
| FSIM | Feature Similarity Index |
| GAN | Generative Adversarial Network |
| GAN-CIRCLE | Generative Adversarial Network Constrained By The Identical, Residual, And Cycle Learning Ensemble |
| GM | Image Gradient Magnitude |
| GPU | Graphics Processing Unit |
| G-SMEIR | General Simultaneous Motion Estimation And Image Reconstruction |
| HRCT | High-Resolution Computed Tomography |
| HU | Hounsfield Unit |
| IM4D | Image Domain Motion Estimation |
| L-R | Left-Right |

| | |
|---|---|
| LRCT | Low-Resolution Computed Tomography |
| MAP | Maximum A Posteriori |
| MDCT | Multi-Detector Computed Tomography |
| MF-MTR | Multi-Frame Reconstruction With Motion-Compensated Temporal Regularization |
| MF-PMM | Multi-Frame Reconstruction With The Parametric Motion Model |
| MRD | Multi-Resolution Demons |
| MSD | Multi-Step Demons |
| OD | Original Demons |
| PC | Phase Congruency |
| PET | Positron Emission Tomography |
| PSNR | Peak Signal-To-Noise Ratio |
| PWLS | Penalized Weighted Least-Square |
| RMSE | Root Mean Squared Error |
| SF-PMC | Single Frame Reconstruction With Post-Reconstruction Motion Correction |
| S-I | Superior-Inferior |
| SMEIR | Simultaneous Motion Estimation And Image Reconstruction |
| SPECT | Single-Photon Emission Computed Tomography |
| SR | Super Resolution |
| SRCT | Super Resolution Computed Tomography |
| SRGAN | Super Resolution Generative Adversarial Network |
| SSIM | Structural Similarity Index |
| TACC | Texas Advanced Computing Center |
| TTSR | Texture Transformer For Super Resolution |
| TV | Total Variation |

XCAT   4D Extended Cardiac-Torso

# CHAPTER 1.    INTRODUCTION

The evolution of X-ray Computed Tomography (CT) technology represents a major breakthrough in the history of medical imaging. The development of CT from its conception to its current advanced forms is a journey marked by scientific curiosity, technological advancements, and innovative problem-solving. This part delves into the early beginnings and the key milestones of CT technology, exploring its genesis, development phases, and the pivotal moments that led to its wide use in contemporary clinical practice.

## 1.1   Early beginnings and major milestones in CT technology

X-ray Computed Tomography, more commonly known as CT, is an imaging technology that uses computer-processed combinations of many X-ray measurements taken from different angles to produce cross-sectional (tomographic) images of specific areas of a scanned object. This technology allows the visualization of internal structures within the human body in a non-invasive manner, which has greatly improved the diagnostic abilities of physicians worldwide.

The origins of CT scan technology can be traced back to 1972, with the introduction of the first commercial CT scanner, the EMI-Scanner. Sir Godfrey Hounsfield, a British engineer working at EMI Laboratories in England, developed the EMI-Scanner, for which he was awarded the Nobel Prize in Physiology or Medicine in

1979, along with American physicist Allan Cormack[1], who provided a solution for CT image reconstruction.

Hounsfield's initial device was limited to making images of the brain, but it marked a significant milestone in medical imaging. This pioneering model used a single X-ray tube that revolved around the patient's head, producing a single slice of the brain at a time[2]. The development from single-slice to multi-slice CT scanners represented a significant leap in CT technology. The first single-slice CT scanner, introduced in the early 1970s, took several minutes to acquire one slice of data and nearly five hours to reconstruct a single image. The introduction of multi-slice or multi-detector CT (MDCT) scanners in the late 1990s increased the speed and efficiency of CT imaging by acquiring multiple slices of data in a single rotation, resulting in a significant reduction in scanning and reconstruction time[3].

In 1989, spiral (or helical) CT scanning was introduced. This new scanning method, which involved the continuous rotation of the X-ray tube in a helical path around the patient, allowed for faster, more detailed imaging and was particularly useful for capturing structures such as blood vessels and organs that move with the body's motion, such as the lungs and the heart[4].

The late 1990s and early 2000s saw further improvements with the introduction of cone-beam CT, which offered volumetric acquisition of data and improved 3D imaging capabilities. This method involved the X-ray beam shaping into a cone instead of a fan, which facilitated the acquisition of large volumes of data in a single rotation[5].

Today, CT technology is continually evolving, with improvements in scanner technology, image reconstruction techniques, and the application of artificial intelligence for image enhancement and interpretation. Today's CT scanners are faster, more accurate, and safer, leading to lower radiation dose for patients while maintaining high image quality.

## 1.2 Fundamentals of CT

The mathematical foundation of CT imaging is to reconstruct a 2D or 3D image of an object by analyzing projections of the object in many different directions. The fundamental is an application of the Radon Transform and its inverse. The projection is the process of creating a 2D image (or a 1D line) of a 3D object (or a 2D object). As shown in Figure 1.1 left, an X-ray tube irradiates X-ray beams through the body, which are collected by the detector create a projection [2].
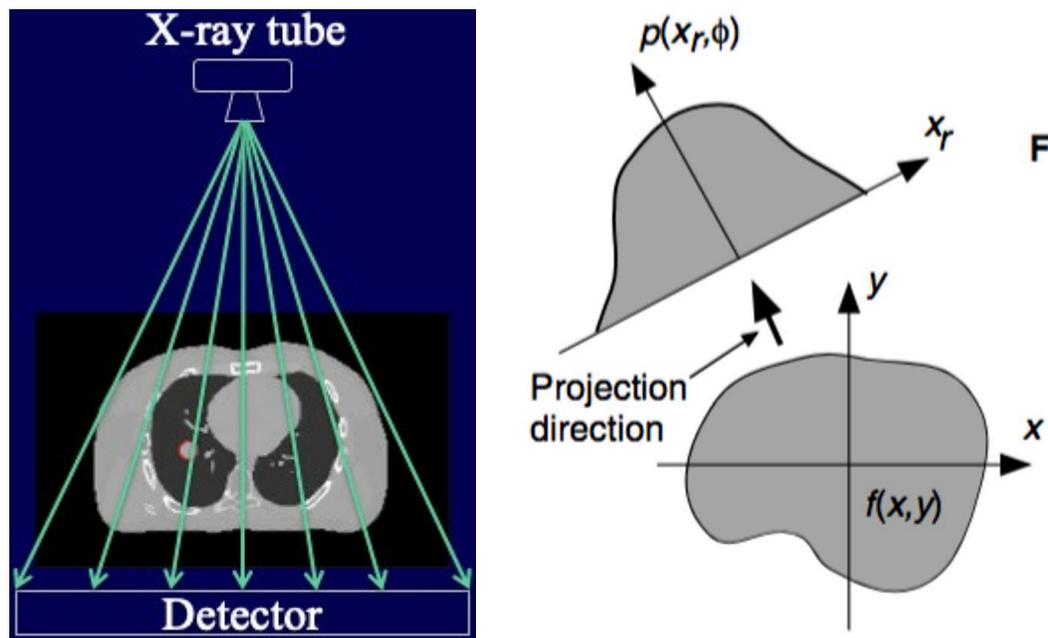


Figure 1.1 The form of CT projections (left) and Radon transform (right) [6]

Given the incident X-ray intensity $I_0$, the detected intensity along a line at angle $\phi$ along $x_r$ (Figure 1.1 middle) can be expressed as:

$$I_{(\phi,x_r)line} = I_0 e^{-\int_{(\phi,x_r)line} f(x,y)ds} \qquad (1.1)$$

where f(x,y) are linear attenuation coefficients at location x and y. Taking a normalized logarithm, the projection can be defined as:

$$p(x_r, \phi) = \ln\frac{I_0}{I_{(\phi,x_r)line}} = \int_{(\phi,x_r)line} f(x,y)ds. \qquad (1.2)$$

Then, the Radon transform [7] (Figure 1.1 right) can be written as:

$$p(x_r, \phi) = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x,y)\delta(x\cos\phi + y\sin\phi - x_r)dx\,dy. \qquad (1.3)$$

The fundamental theorem of CT was introduced in the seminal paper by Radon in 1917[8] and re-invented by Dr. Allan Cormack[1].The theorem can be stated as follows:

If a function $f(x,y)$ is continuous and its Radon Transform $p(x_r, \phi)$ is known for all $x_r$ and $\phi$, then the function $f(x,y)$ can be uniquely determined.

This theorem states that if we know all the projections (i.e., the Radon transform) of an object from every possible direction, we can uniquely reconstruct a complete image of the object (via the inverse Radon transform, which can be used to reconstruct the image $f(x,y)$ from the acquired projections $p(x_r, \phi)$ [7]).

To implement this theorem practically in a CT scanner, projections are acquired from multiple angles around the object, and an algorithm is used to perform the inverse Radon transform (typically in a discrete form of inverse Radon transform with high-pass filtering, i.e. so called filtered backprojection (FBP)) to reconstruct the image. The procedures are shown in Figure 1.2.
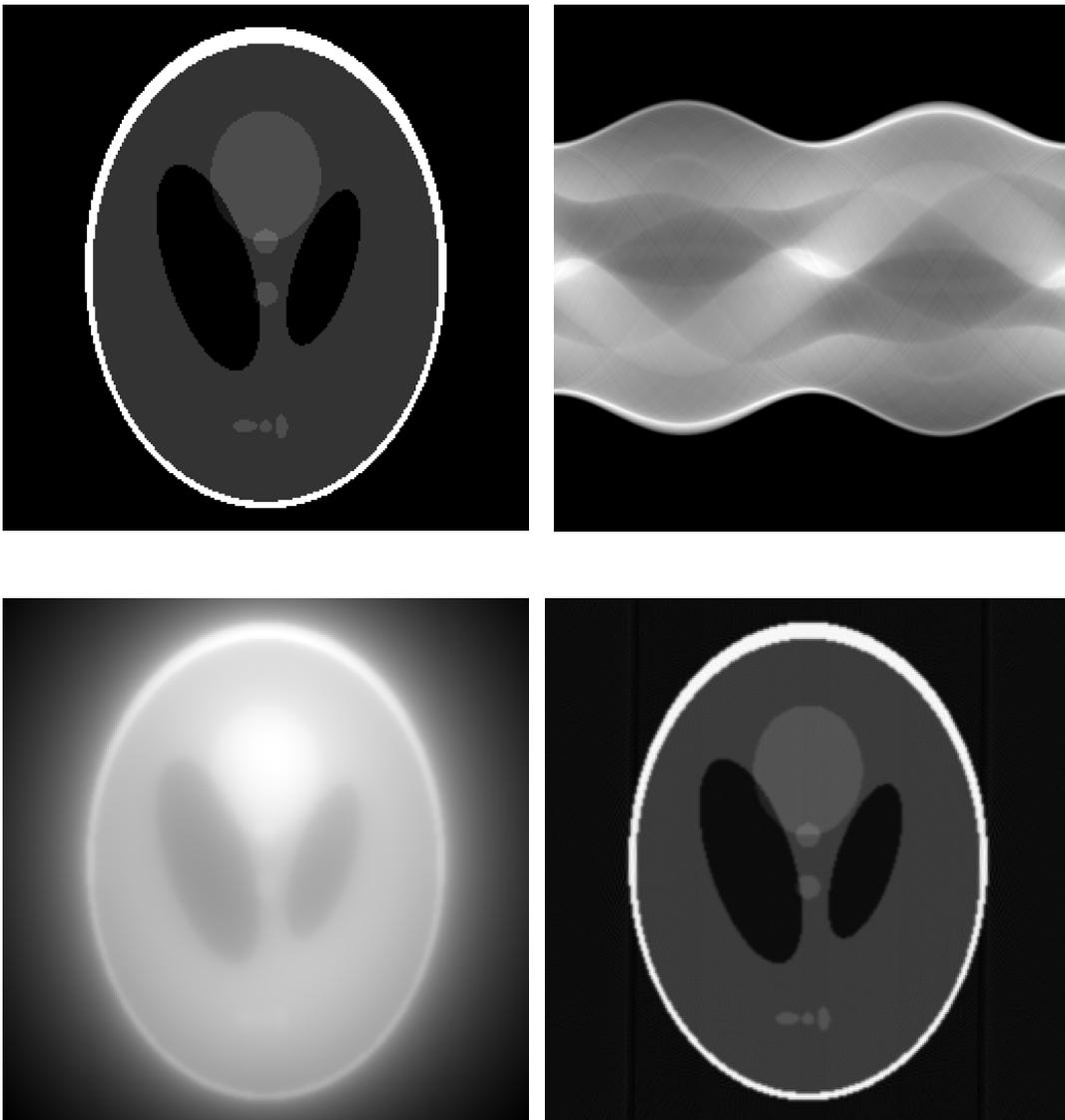


Figure 1.2 Radon transform and CT image reconstruction.(Top left: original image; Top right: Projection ("Sinogram") from Radon transform; Bottom left:

direct inverse Radon transform; Bottom right: inverse Radon transform with high-pass filtering (FBP))

It is important to note that the theorem assumes the object is perfectly sampled with an infinite number of projections, which is not possible in real-world CT imaging. Nevertheless, the theorem provides the mathematical basis for CT image reconstruction, and the computer algorithms work well with a finite number of projections (e.g. Figure 1.2 for 360 views over 360 degrees).

## 1.3 Motion tracking in CT imaging

One inherent challenge in thoracic CT imaging is patient's respiratory motion, which can potentially cause image artifacts and obscure diagnostic information. Respiratory motion causes blurred images, leading to misrepresentation of organ structures and potential misdiagnosis[9]. Tumors can appear smeared or displaced, impairing the accuracy of diagnosis, radiotherapy planning, and treatment assessment. Consequently, techniques for tracking and compensating for respiratory motion have emerged as crucial aspects in improving the quality and diagnostic value of CT images.[10, 11]

Various motion-tracking techniques have been developed, such as retrospective and prospective gating, 4D-CT, and the use of surrogate signals[12]. In retrospective gating, the entire breathing cycle is scanned, and data are retrospectively sorted into different respiratory phases. In contrast, prospective gating only scans during a specific phase of the breathing cycle. These methods can significantly reduce motion artifacts but may also involve higher radiation doses or miss the respiratory motion as a whole[13].

4D-CT involves acquiring images at different phases of the respiratory cycle, providing a "4th dimension" of information – time. 4D-CT enables visualization of the organ's motion trajectory during breathing, which is valuable for radiation therapy planning[14].

Surrogate signals techniques use surrogate signals like chest surface motion or internal fiducial markers to predict the respiratory phase. The prediction can then be used to synchronize image acquisition with the respiratory cycle[15].

The implications and advantages of respiratory motion tracking in CT imaging are multifaceted, improving not only diagnostic accuracy and imaging dose, but also treatment efficacy and safety. Motion tracking compensates for respiratory motion, enabling a more accurate depiction of internal structures. This facilitates better detection and characterization of pathologies, thus significantly enhancing the diagnostic accuracy. In radiation therapy, the accurate targeting of the tumor while sparing healthy tissues is of utmost importance. By integrating motion tracking, the irradiated volume can be determined with precision, leading to safer and more efficacious treatment strategies. Respiratory motion tracking also allows the correlation of image acquisition with specific respiratory phases, minimizing unnecessary irradiation of CT imaging for patients.

As we transition into the future of CT imaging, respiratory motion tracking technology continues to evolve, promising even greater accuracy and efficiency. Emerging techniques such as artificial intelligence (AI)-based algorithms offer the potential to further refine motion tracking, which can not only improve the

diagnostic value of CT imaging but also make radiation therapy more precise and patient-specific. Additionally, the advent of wearable devices for patient monitoring presents a new avenue for non-invasive and continuous tracking of respiratory motion. These technological advancements underscore the dynamic role and importance of respiratory motion tracking in the future of CT imaging. As these new methods and technologies are developed and refined, they have the potential to revolutionize the field, offering safer, more precise, and more efficient diagnostic and treatment options.

## 1.4   Aims of this work

In the following chapters, we will present several methods, which could improve the image quality for CT. These methods can be separated into two categories: 1) conventional methods based on mathematical and physical modelling; and 2) deep learning methods. Although the deep learning methods emerge as a new trend, conventional methods are still effective to solve some problems in CT.

- Chapter 2: To solve the respiratory motion problem in slow CBCT imaging, we proposed a numerical method named general simultaneous motion estimation and image reconstruction (G-SMEIR), which estimates the respiratory motion more accurately. Dr. Yujie Chi helped review the manuscript. Dr. Jing Wang helped review the manuscript and reconstructed images. Dr. Mingwu Jin helped code, analyse the result, and review this manuscript.

- Chapter 3: To denoise CT images using unpaired training, we proposed two deep learning methods, CycleGAN and ReCycleGAN, where the latter enforces temporal connections between cardiac phases to effectively denoise a low-dose CT image sequence. Dr. Jinyu Yang helped validate the phantom simulation, Krishnateja Konduri helped run cross-validation simulation and patient training, Dr. Junzhou Huang helped review the manuscript. Dr. Lifeng Yu helped collect patient data and review the manuscript. Dr. Mingwu Jin helped analyse the result, and review the manuscript.

- Chapter 3: To achieve super-resolution of the CT images, we proposed a deep learning method named texture transformer super-resolution (TTSR), which converts noisy low-resolution CT images to clean high-resolution ones. Dr. Lifeng Yu helped review the manuscript. Dr. Mingwu Jin helped analyse the result, and review the manuscript.

Overall, we aim to improve CT image quality using either conventional methods or deep learning.

.

# CHAPTER 2.    GENERAL SIMULTANEOUS MOTION ESTIMATION AND IMAGE RECONSTRUCTION

Shiwei Zhou[1] , Yujie Chi[1] , Jing Wang[2] and Mingwu Jin[1]

[1] Department of Physics, University of Texas at Arlington, Arlington, TX 76019, United States of America

[2] Department of Radiation Oncology, University of Texas Southwestern Medical Center, Dallas, TX 75390, United States of America

## 2.1 Introduction

Motion-compensated tomographic image reconstruction for X-ray computed tomography (CT), single-photon emission computed tomography (SPECT), and positron emission tomography (PET) has been an active research topic for many years. There are three popular ways to apply motion compensation for improved image quality by suppressing noise and motion artifacts [16]: 1) single frame reconstruction with post-reconstruction motion correction (SF-PMC) [17-19]; 2) multi-frame reconstruction with motion-compensated temporal regularization (MF-MTR) [20-24]; and 3) multi-frame reconstruction with the parametric motion model (MF-PMM) [25-37]. It has been shown that MF-MTR collapsed to a single frame reconstruction if the weight of the temporal regularization is zero and converged to MF-PMM if the weight of the temporal regularization becomes very large using penalized weighted least-square (PWLS) estimators with known nonrigid motion fields [16]. However, the large weight of the temporal regularizers may lead to slow convergence of MF-MTR. It also shows that MF-PMM can lead to a smaller variance than SF-PMC based on Maximum a Posteriori (MAP) estimators with Poisson likelihoods. Furthermore, the specially designed regularizers may improve the nonuniform and anisotropic spatial resolution of these methods at the expense of noise performance [38]. Even though MF-PMM may be preferable to SF-PMC and MF-MTR, the accurate motion estimation in practice is be a key factor to achieve these theoretical predictions of noise and spatial resolution performance and to improve reconstructed image quality.

Cone-beam CT (CBCT) has been widely used in image-guided radiation therapy. Traditionally, CBCT is performed to get 3D images for patient alignment. However, the acquisition of one 3D image is not sufficient for moving targets, such as tumors in the lungs [39]. 4D CBCT has been proposed to acquire a series of 3D images that can track the motion of tumors, e.g. respiratory motion for lung cancer patients [40, 41]. Several clinical studies have shown that 4D CBCT offers a more accurate target location as compared to 3DCBCT for motion-involved sites [42, 43]. Furthermore, 4D CBCT has been used to reconstruct and monitor the actual dose delivered to patients for adaptive radiation therapy [44, 45]. For 4D CBCT, if single frame 3D reconstruction methods are used, the radiation dose to image the patient would have to increase dramatically to keep the image quality since each of the 3D image series needs a regular scan. Not only could the excessive ionizing radiation lead to harmful secondary diseases, especially for young patients [46], but also the acquisition time will be long, leading to reduced clinical workflow. The quality of 4D CBCT images can be greatly improved by using the aforementioned motion-compensated reconstruction methods. In our previous work, we developed a simultaneous motion estimation and image reconstruction (SMEIR) for 4D CBCT [36], which a type of MF-PMM method. In SMEIR, the motion estimation in the projection domain is alternatively updated with image reconstruction iterations.

Although SMEIR can accurately reconstruct the tumor location and motion during patient breath for precise radiotherapy, it suffers a local optimum trapping problem for motion estimation conducted in the projection domain. We proposed a modified SMEIR method to alleviate this problem by adding one more step of image-domain

12

motion estimation [47]. In this work, we further develop a general SMEIR (G-SMEIR) framework, which can flexibly combine SMEIR iterations with image-domain motion estimation to improve estimated deformation vector fields (DVFs) for better 4D reconstruction. To tackle the problem of computational burden, CPU parallel computing is used for SMEIR with multi-reference frames and GPU computing is used for image-domain motion estimation [48]. The G-SMEIR framework can be readily extended to other tomographic modalities, such as CT, SPECT, and PET [34].

## 2.2   Methods

### 2.2.1   SMEIR

Here we briefly introduce the image model and the simultaneous motion estimation and image reconstruction (SMEIR) method, one of MF-PMM methods, which utilizes all projections (for $P$ phases) to reconstruct phase images by exploring the motion correlation of different phases in the projection domain [36]. The image model of $P$-phase images can be described as,

$$p_k = A_k \mu_k + \epsilon_k, for\ k = 1, 2, \dots P, \qquad (2.1)$$

where $p_k$ is the projection data, $A_k$ is the projection matrix for phase $k$, whose element is the intersection length of a particular voxel with the ray reaching a particular detector bin, $\mu_k$ is the image to be reconstructed for phase $k$, and $\epsilon_k$ is the corresponding noise. Note that $A_k$ is different for different phases due to

different view angles. If the motion transform is available, an image at phase $k$ can be obtained from a reference phase, say phase 1 for example,

$$\mu_k = T_{1,k}\mu_1, \qquad for\ k = 1, 2, \qquad ...\ P, \qquad (2.2)$$

where $T_{1,k}$ is the deformation matrix to transform the image at phase 1 to that at phase $k$ . Given $p = [p_1', p_2', ..., p_P']'$ as the projections for all phases, $\epsilon = [\epsilon_1', \epsilon_2', ..., \epsilon_P']'$ as the noise terms, and $A = ((A_1 T_{1,1})', (A_2 T_{1,2})', ..., (A_P T_{1,P})')'$ as the composite projection matrix, all projection data can be used jointly for phase 1 image reconstruction,

$$p = A\mu_1 + \epsilon . \qquad\qquad (2.3)$$

The constrained total variation (TV) minimization [36] can be used to solve for $\mu_1$, whereas the other phase images can be obtained by using (2.2) given the deformation matrix $T_{1,k}$.

In practice, the deformation matrices in (2.3) need to be estimated. The deformation vector field (DVF) $v^{1 \to k}$ is defined as the displacement vector from phase 1 to phase $k$ so that the following Eq. (2.4) holds,

$$T_{1,k}\mu_1 = \mu_1(x + v^{1 \to k}), for\ k = 1, 2, \qquad ...\ P. \qquad\qquad (2.4)$$

The forward $v^{1 \to k}$ and backward $v^{k \to 1}$ DVFs in SMEIR are estimated as follows,

$$v^{1 \to k*} = \arg\min f_1(v^{1 \to k}) = \arg\min \|p_k - A_1\mu_1(x + v^{1 \to k})\|_2^2 + \beta R(v^{1 \to k})$$

$$v^{k \to 1*} = \arg \min f_2(v^{k \to 1}) = \arg \min \|p_1 - A_k \mu_k (x + v^{k \to 1})\|_2^2 + \beta R(v^{k \to 1})$$

$$s.t. \, v^{1 \to k} \circ v^{k \to 1} = v^{k \to 1} \circ v^{1 \to k} = 0, \qquad (2.5)$$

where $R(\bullet)$ is the regularization on DVFs for this ill-posed problem, $\beta$ is used to balance the motion match and regularization, and $\circ$ is a composition operation in the last constraint to enforce the inverse consistency between the forward and backward DVFs. The initial DVFs can be obtained from motion estimation using images reconstructed by 3D phase-by-phase total variation minimization (3D TV). In SMEIR, the motion estimation step (2.5) is alternately updated with the joint reconstruction step of (2.3) until convergence (see yellow blocks in Figure 2.1). The pseudo-code of SMEIR and TV minimization can be found in the Appendix.

### 2.2.2  G-SMEIR

It is observed that the reference phase image reconstructed by SMEIR usually has better image quality than those of other phases transformed using the estimated DVFs, which change little after several iterations in SMEIR. To provide a general framework for 4D reconstruction and address the local optimum trapping problem, we propose a general SMEIR (G-SMEIR) framework to overcome this problem of SMEIR as shown in Figure 2.1, which is equivalent to solve Eq. (2.3) using each phase as the reference phase,

$$\boldsymbol{p_k} = \boldsymbol{A}\mu_k + \boldsymbol{\epsilon}, for \, k = 1, 2, \dots, P. \qquad (2.6)$$

The rationale behind G-SMEIR is that SMEIR is applied on all phases, thus leading to better individual phase images. Then, these images are used to estimate the DVFs, which may jump out the local optimum trapped in SMEIR. Therefore, in G-SMEIR, the image domain motion estimation is conducted after every N SMEIR iterations. In this work, we choose to use Demons non-rigid registration [49], where the DVF can be described as:

$$v = \frac{(m-s)\boldsymbol{\nabla}s}{(\boldsymbol{\nabla}s)^2 + (m-s)^2} \ or \ \mathbf{0} \ if (\boldsymbol{\nabla}s)^2 + (m-s)^2 < \varepsilon \qquad (2.7)$$

where $s$ is the reference image and $m$ is the target image. A Gaussian filter is applied after each iteration to smooth the DVF. The smoothing parameter is defined as the standard deviation of the Gaussian smoothing kernel. The pseudo-code of the Demons algorithm can be found in the Appendix.

Although Demons non-rigid registration was used in this work, any other image domain motion estimation methods can be equally applied. The inner SMEIR iteration number N and the outer G-SMEIR iteration number M can be flexibly combined to achieve a trade-off between reconstruction quality and speed. The pseudo-code for G-SMEIR is listed in the Appendix along with the parameter selection.

It can be seen from the G-SMEIR structure in Figure 2.1 that the SMEIR part for each phase image reconstruction can be easily computed in parallel using $P$ CPUs. One of the computational bottlenecks of G-SMEIR is 3D image domain motion estimation since DVFs for each pair of phases need to be estimated. For

example, Demons registration for 7 pairs of 3D images (for one reference phase of P=8) takes 2 hours Intel(R) Xeon(R) CPU E5-2620 v4 CPU, which makes M iterations of G-SMEIR computationally impractical. In this regard, we investigated the acceleration of image domain motion estimation through both algorithms and GPU acceleration.
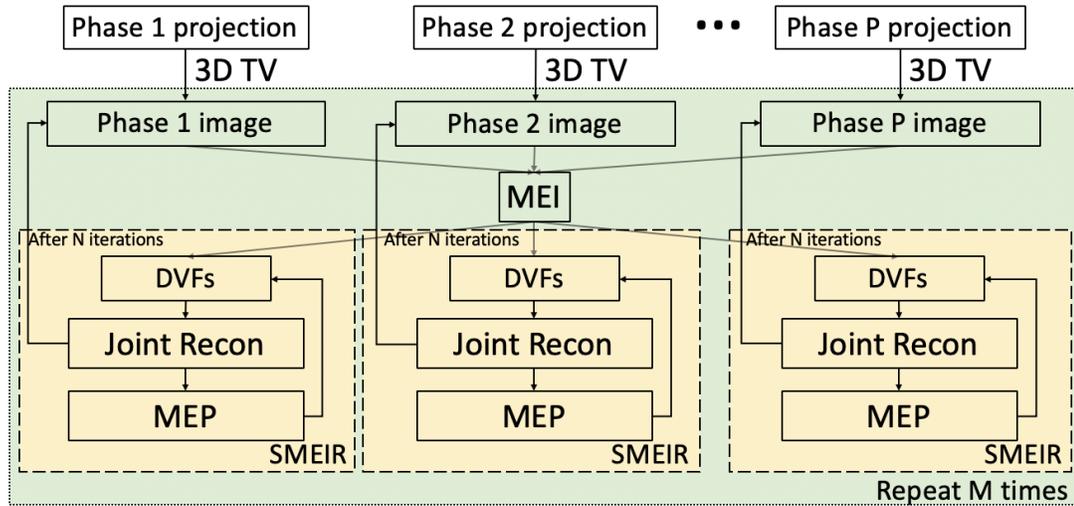


Figure 2.1 Flow chart of G-SMEIR for N inner loops and M outer loops. 3D TV: 3D TV minimization reconstruction; MEI: motion estimation in the image domain; DVFs: deformation vector fields; MEP: motion estimation in the projection domain. The projection data for all P phases are used for joint reconstruction in each SMEIR (connections not shown in the figure for conciseness).Simulation experiments

A 4D extended cardiac-torso (XCAT) phantom [50] with respiratory motion (10 phases) was used to evaluate the performance of G-SMEIR. The XCAT phantom images at two representative phases (Phase 1 and 4) are shown in Figure 2.2. The dimensions of the XCAT phantom were 256×256×100 with a voxel size of 2×2×2 mm$^3$. CBCT acquisition was simulated using 300 projections, i.e. 30 views/phase, distributed evenly over 360° by a fast ray-tracing algorithm [51]. The dimensions of each projection were 384×150 with a detector pixel size of 2×2 mm$^2$. The

Poisson-distributed counting noise ($1 \times 10^5$ photons/incident ray for regular dose, $5 \times 10^4$ photons/incident ray for half dose, $10^4$ photons/incident ray for 10% dose) and Normal-distributed electronic noise (variance of 10) were added to the original noise-free projections. The number of photons per incident ray represents the number of photons that would reach the detector element if there is no any attenuation along the source and the detector. To evaluate the motion tracking performance of each method, a spherical 3D tumor with a diameter of 10 mm was also introduced.

As for the image domain motion estimation methods, we compared both MRD and MSD for convergence and warped image quality to determined MRD as the choice for G-SMEIR in this work.

For reconstruction, we compared the following four methods: 1) phase-by-phase 3D total-variation minimization reconstruction (3D TV); (2) 4D reconstruction with image domain motion estimation (IM4D) (M=24 and N=0); 3) SMEIR (M=0 and N=24); and 4) G-SMEIR with different combinations of N and M (M+MxN = 24). In order to have a fair comparison, we fixed the number of joint reconstructions (Figure 2.1) to be 24 for 4D iterative reconstruction methods. When M=24 and N=0, G-SMEIR becomes IM4D, whereas when M=0 and N=24, G-SMEIR collapses into SMEIR with 24 iterations (10 projection/backprojection pairs for each iteration). For 3D TV, we ran additional 24 iterations (20 projection/backprojection pairs for each iteration) starting with the input images for initial DVF estimation. More projection/backprojection operations help 3D TV obtain improved image quality. For IM4D and SMEIR, Phase 1 was used as the reference phase and the

other phases were obtained by warping the reconstructed image of the reference phase to the target phase using the estimated DVFs, unless otherwise stated. The reconstruction performance is evaluated qualitatively by images and quantitatively by root mean squared error (RMSE) and structural similarity index (SSIM) calculated in the 3D image volume (256x256x100) for each phase, unless otherwise stated. In addition, the tumor motion recovery performance of different methods is measured by the maximum and mean deviations from the phantom tumor motion, which is averaged over the 3x3x3 volume at the tumor center.
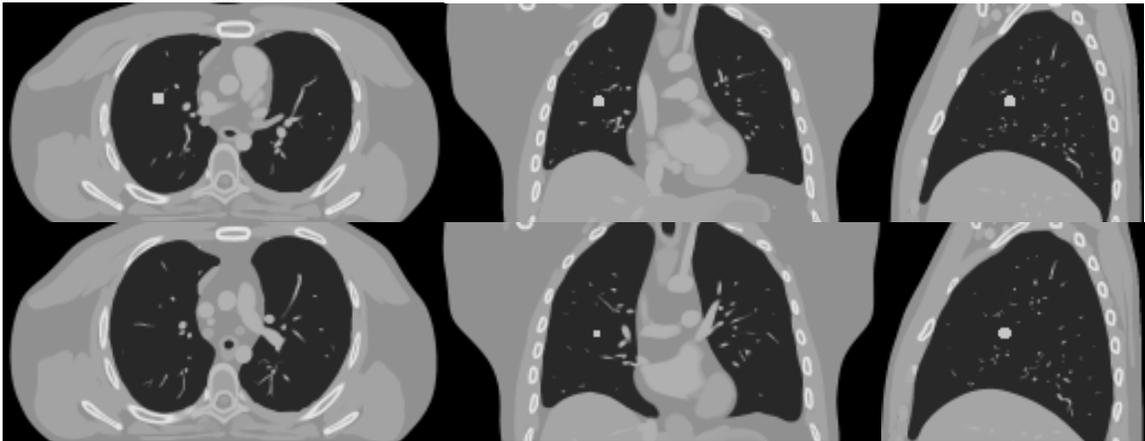


Figure 2.2 The XCAT phantom images for Phase1 (top) and Phase 4 (bottom) in transverse (left), coronal (middle) and sagittal (right) views.

### 2.2.3  Patient experiments

We also test our method on a real patient CBCT data with eight respiratory phases. The use of anonymous projection data from this patient was approved by UTSW IRB (082013-008). The data were acquired using a Varian CBCT system. The acquisition protocol parameters were: 120 kVp and 1.6 mAs per projection, a total of 534 projection views (1024 × 768 pixels with a pixel size of 0.388 × 0.388 mm$^2$ for each view) evenly distributed in 360º, and acquisition time of 1 min. Each

projection was downsampled by a factor of 2 before reconstruction. The source to detector distance was 1500 mm and the source to isocenter distance was 1000 mm. The projection data were sorted into ten phases based on the respiratory signal recorded by Real-time Position Management system (Varian, Inc.). Thirty views per each respiratory phase were selected, leading to a total of 300 projection views for 4D reconstruction. The dimensions of reconstructed images were 150×150×100 with voxel size of 2×2×2 $mm^3$. We compared the patient images reconstructed by 1) 3D TV; 2) IM4D; 3) SMEIR; and 4) G-SMEIR.

## 2.3  Results

### 2.3.1  Comparison of the original Demons, MRD, and MSD

To speed up the image-domain motion estimation, particularly for large motions, we used the XCAT phantom images to compare the original demons (OD) algorithm, multi-resolution demons (MRD), and multi-step demons (MSD). The phase 1 image serves as the reference phase and the phase 3 and 4 images are used as the target phases. After the target images were registered to the reference image using different methods, the RMSE and SSIM values of the registered phase 1 images were calculated. 500 iterations of OD lead to the plateau of RMSE, while 100 iterations for MSD and MRD. In Figure 2.3, we plotted these values along with the changing smoothing parameter of DVF.  The best performance of MRD and MSD is similar for 100 and 500 iterations and better than 500 iterations of OD, indicating their improved convergence over OD. In addition, MSD is robust to a wide range of the smoothing parameter, while MRD works well for a narrower

range. However, MSD uses OD as the first-level estimate, which leads to registration performance inferior to MRD as shown in the registered images in Figure 2.4. In the rest of this work, MRD was used for G-SMEIR.
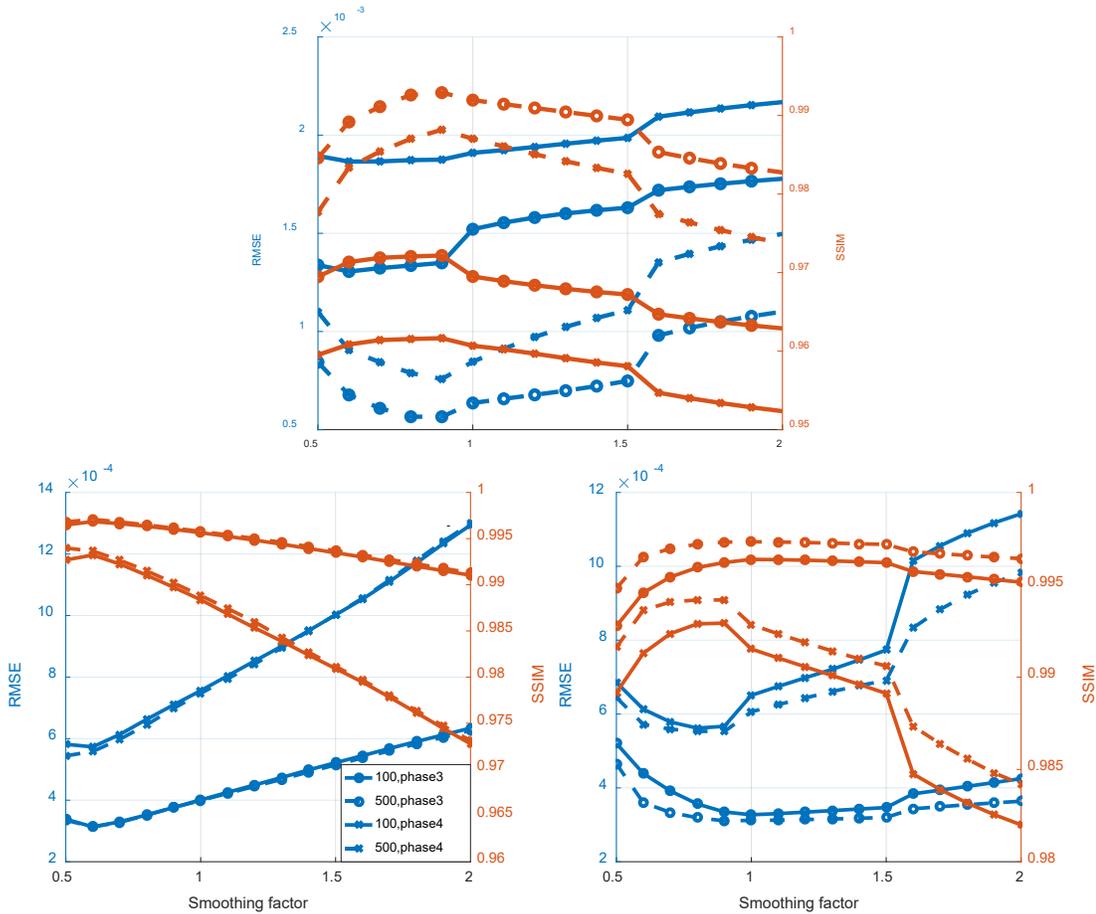


Figure 2.3 The RMSE and SSIM values over different smooth parameters (the horizontal axis) (the legend shows the target phase and the number of iterations in different line styles). (a: OD; b: MRD; c: MSD).

In order to shorten the computation time, the GPU acceleration was used. To achieve a similar performance of RMSE, OD needs 500 iterations at the original resolution (~10.05 s GPU computing averaged over 50 repetitions). MRD needs 100 iterations at each resolution, which is equivalent to around 188 iterations (~3.72 s). MSD only needs 100 iterations (4.08 s = 2.98 s for demons plus 1.10 s

for the sum of DVF). The computation time of image domain motion estimation has been reduced from 17 minutes (CPU) to about 3.72 seconds (GPU) for each pair of 3D images, which makes M times of image-domain motion estimation of G-SMEIR practical.
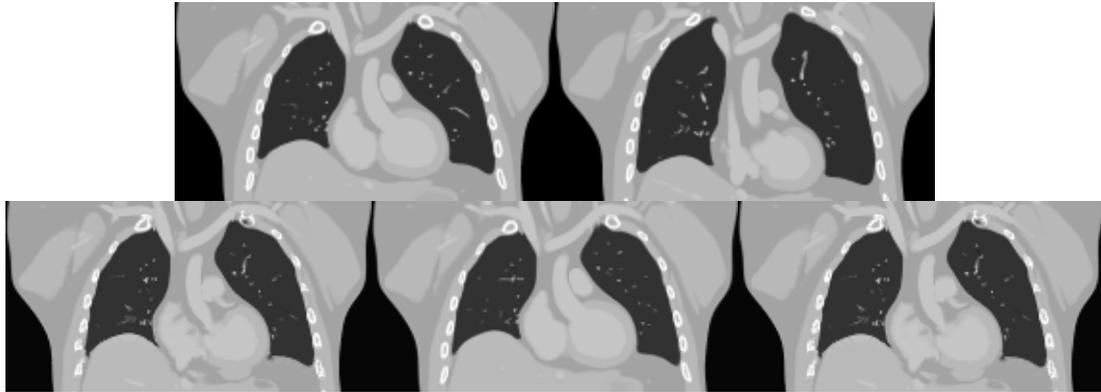


Figure 2.4 Comparison of registered images. (a: Phase 1 image; b: Phase 4 image; c: OD registered Phase 1 image from Phase 4 image; d: MRD registered Phase 1 image from Phase 4 image; e: MSD registered Phase 1 image from Phase 4 image.)

### 2.3.2 Convergence of the motion estimation objective functions

The average values of the forward and backward motion-estimation objective functions in Eq. (2.5) vs. the iteration number are shown in Figure 2.5, where the dashed lines for SMEIR and the solid lines for G-SMEIR (three cases: M=2, N=11; M=3, N=7; and M=4, N=5). Here we only showed the full dose results for conciseness whereas the half dose and 10% dose results followed a similar trend with larger values. As can be seen, the image domain motion estimation in all three G-SMEIR cases breaks the convergence pattern and leads to a smaller objective function value than SMEIR for the same number of iterations. It is worth noting that although there is an initial jump of the curve of G-SMEIR right after the image

domain motion estimation was applied, it drops quickly and becomes lower than SMEIR later on. These results confirm that G-SMEIR can jump out the local motion-estimation optimum trapped in SMEIR.
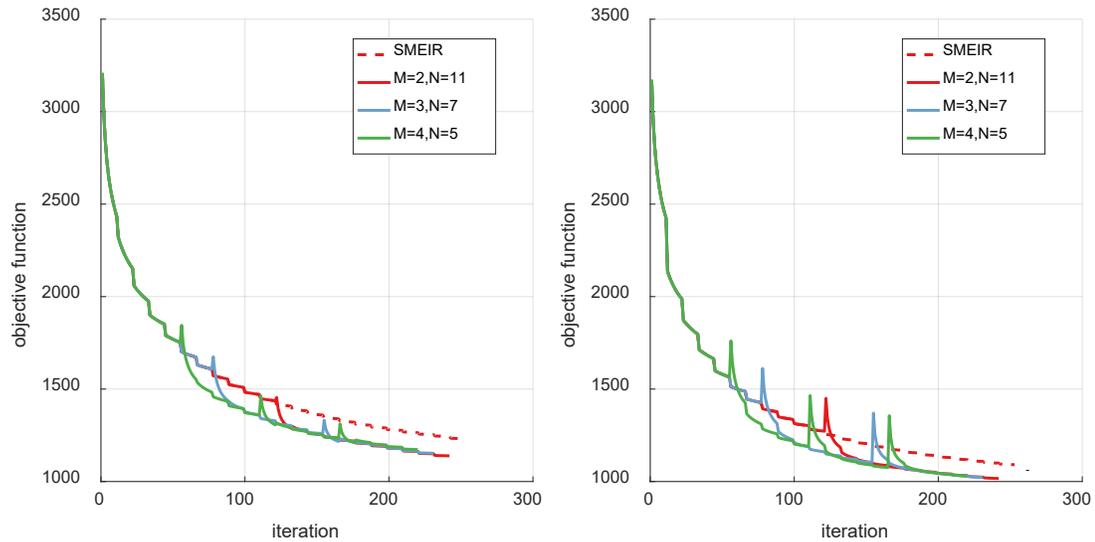


Figure 2.5 Comparison of motion-estimation objective functions between SMEIR and G-SMEIR (M=2, N=11; M=3, N=7; and M=4, N=5). (a: Phase 1 to Phase 2; b: Phase 4 to Phase 5).

### 2.3.3  Reconstruction results for individual phases

The quantitative results of Phase 1 and Phase 4 images at full dose for 3D TV, IM4D, SMEIR, and G-SMEIR with different combinations of M and N are shown in Figure 2.6. For IM4D and SMEIR, Phase 1 was used as the reference phase and Phase 4 was obtained by warping the phase 1 image to Phase 4 using the estimated DVFs. All methods improve RMSE and SSIM along with the iteration. 3D TV is much worse than three motion-compensated reconstruction methods. Among motion-compensated reconstruction methods, SMEIR and G-SMEIR outperform IM4D and seem to have comparable performance in terms of RMSE

23

and SSIM for Phase 1. However, the superior performance of G-SMEIR over SMEIR and IM4D becomes obvious for Phase 4 images.
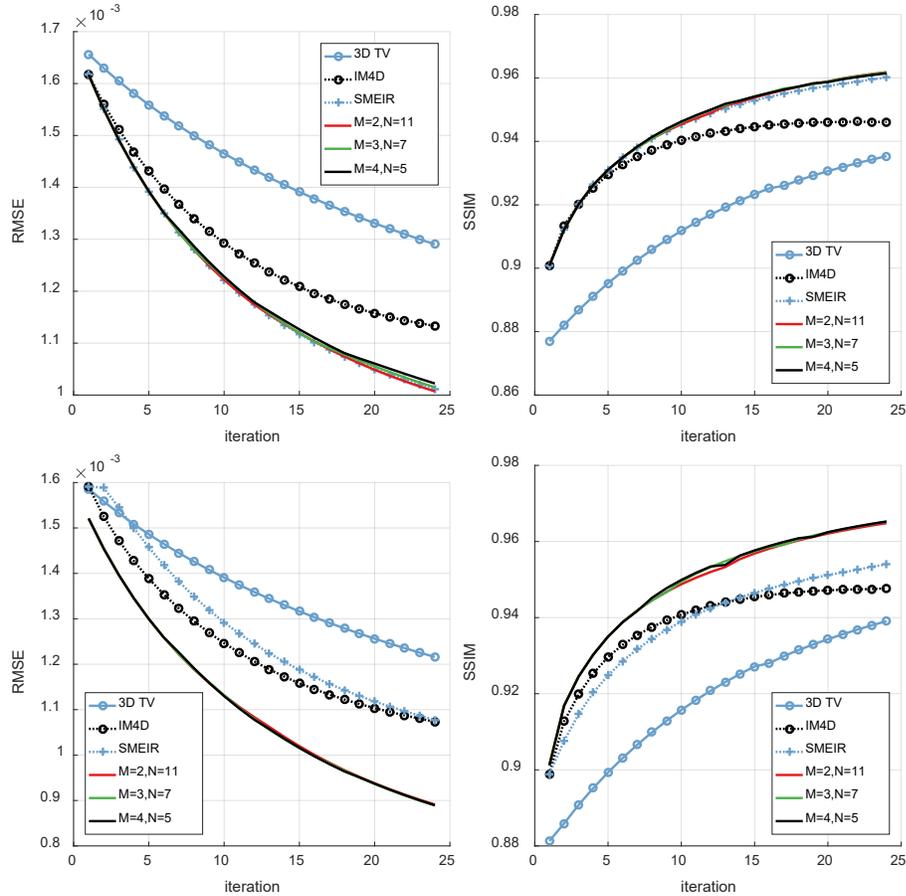


Figure 2.6 Quantitative accuracy for different reconstruction methods at the full dose level for phase 1 (Top) and 4 (Bottom). M and N are combinations of G-SMEIR.

The RMSE and SSIM values for Phase 1 and 4 of the final reconstruction images at full dose are listed in Table 1. Motion-compensated methods (IM4D, SMEIR, and G-SMEIR) outperform 3D-TV by large margins (10~25% reduction in terms of RMSE). G-SMEIR (M=2 and N=11) achieves the best performance on both RMSE and SSIM. For Phase 1 (i.e. the reference phase) SMEIR seems to have better RMSE than the other G-SMEIR combination and worse SSIM than G-SMEIR.

However, the performance of G-SMEIR for Phase 4 becomes notably better than SMEIR for both RMSE ($0.89 \times 10^{-3}$ vs $0.92 \times 10^{-3}$) and SSIM (0.9653 vs. 0.9621). In Table 2.1, the larger SSIM and lower RMSE in Phase 1 than Phase 4 were only observed for 3D TV, IM4D, and G-SMEIR, but not for SMEIR. This was caused by the variation of XCAT phantom at different respiratory phases. For SMEIR, although its RMSE and SIMM were better than 3D TV, their values in the reference phase (Phase 1) were indeed better than Phase 4. This indicates that a large motion error may exist to produce the worse warped image at Phase 4. These results show that G-SMEIR can effectively solve this problem of SMEIR.

Table 2.1 The RMSE and SSIM values of different reconstruction methods at full dose for Phase 1 and Phase 4. (M, N) is for G-SMEIR.

| | Phase 1 | | Phase 4 | |
|---|---|---|---|---|
| | RMSE | SSIM | RMSE | SSIM |
| 3D TV | $1.27 \times 10^{-3}$ | 0.9365 | $1.19 \times 10^{-3}$ | 0.9403 |
| IM4D | $1.13 \times 10^{-3}$ | 0.9461 | $1.07 \times 10^{-3}$ | 0.9477 |
| SMEIR | $1.01 \times 10^{-3}$ | 0.9601 | $1.08 \times 10^{-3}$ | 0.9540 |
| (2,11) | $1.00 \times 10^{-3}$ | 0.9618 | $0.89 \times 10^{-3}$ | 0.9648 |
| (3,7) | $1.01 \times 10^{-3}$ | 0.9617 | $0.89 \times 10^{-3}$ | 0.9652 |
| (4,5) | $1.02 \times 10^{-3}$ | 0.9614 | $0.89 \times 10^{-3}$ | 0.9653 |

For the reconstruction image comparison, we used the combinations for G-SMEIR that achieved the best RMSE (i.e. M=2 and N=11). The reconstructed images for four methods in three orthogonal views for Phase 1 and Phase 4 are shown in Figure 2.7 and Figure 2.8, respectively. For phase 1 images in Figure 2.7, All three motion-compensated reconstruction methods greatly reduce the streak artifacts in 3D TV. IM4D seems to suffer the motion artifacts, e.g. the boundary of the diaphragm. Both SMEIR and G-SMEIR achieve image quality superior to 3D TV

and IM4D. Although the difference between SMEIR and G-SMEIR is small in general, G-SMEIR seems to suffer fewer artifacts than SMEIR, e.g. the liver in the sagittal view. For phase 4 images in Figure 2.8, the overall image quality can be observed similar to phase 1 images. However, both IM4D and SMEIR suffer some motion blur, particularly at the tumor location, due to imperfect DVFs used for image warping. SMEIR seems to perform inferior to IM4D in terms of the tumor recovery (red circles in Figure 2.8). Such deterioration is successfully eliminated by G-SMEIR.
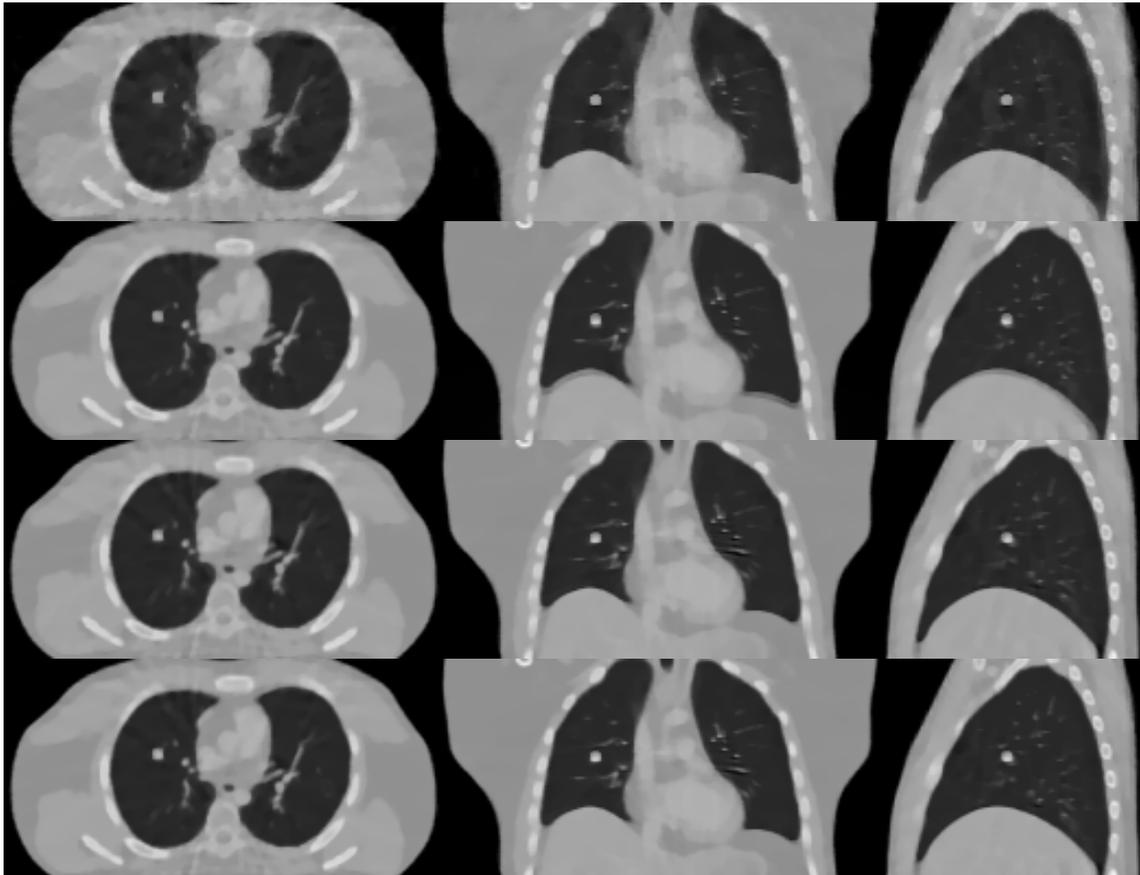


Figure 2.7 Reconstructed XCAT images for different methods for Phase 1 at full dose. From left to right: transverse, coronal, and sagittal; from top to bottom: 3D TV, IM4D, SMEIR, and G-SMEIR. (HU range is [-1000, 1427]).
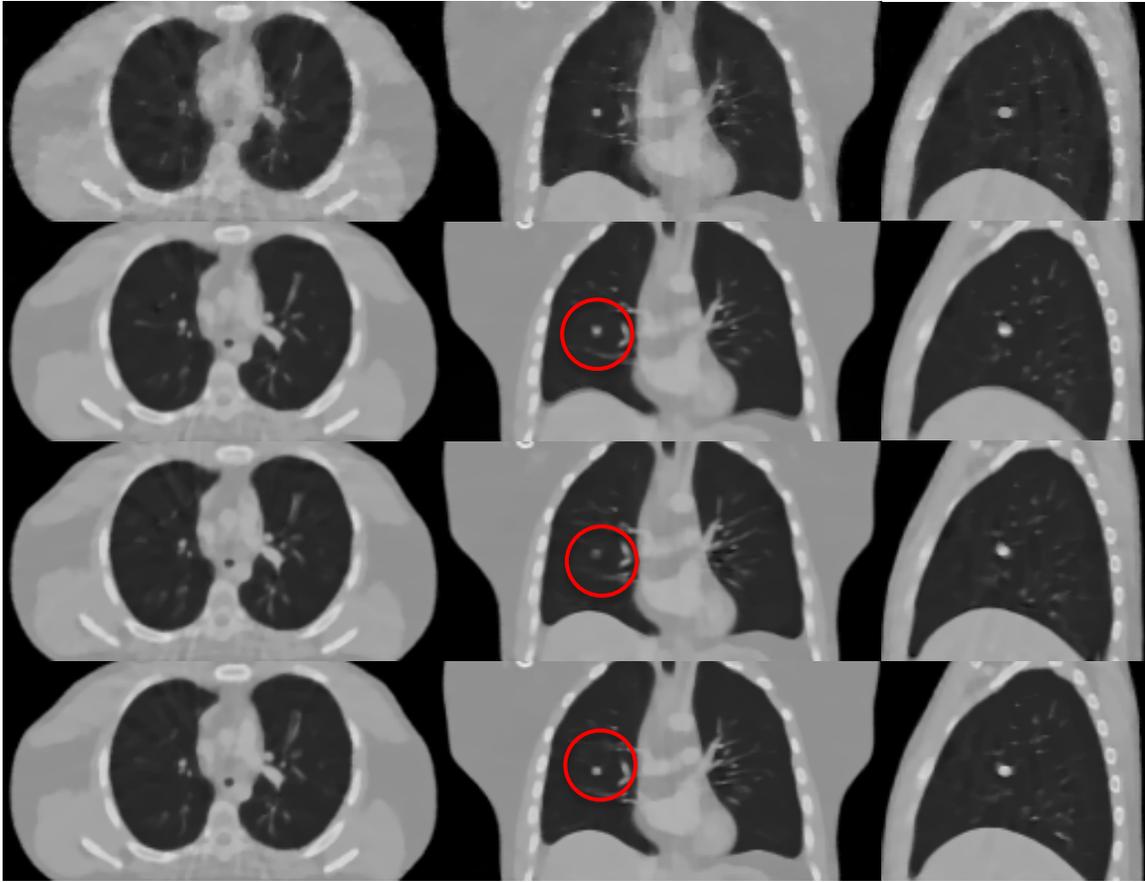
Figure 2.8 Reconstructed XCAT images for different methods for Phase 4 at full dose. From left to right: transverse, coronal, and sagittal; from top to bottom: 3D TV, IM4D, SMEIR, and G-SMEIR. (HU range is [-1000, 1427]).

The results of half dose and 10% dose are similar to that of full dose. For conciseness, here we only show the RMSE and SSIM values of final reconstructed images for Phase 1 and 4 in Table 2.2 and Table 2.3 for half dose and 10% dose, respectively. Again, the performance is in the ascent order for 3D TV, IM4D, SMEIR, and G-SMEIR. The performance is similar between the full dose case and the half dose case, whereas a large degradation is observed from half dose to 10% dose. However, motion-compensated reconstruction methods degrade less than 3D TV (~5% increase in RMSE for the former vs ~10% increase in RMSE for the latter). Although SMEIR works well for the reference phase (Phase 1), its

performance on the non-reference phase (Phase 4) is substantially worse than G-SMEIR.

Table 2.2 The RMSE and SSIM values of different reconstruction methods at half dose for Phase 1 and Phase 4. (M, N) are for G-SMEIR.

|  | Phase 1 | | Phase 4 | |
| --- | --- | --- | --- | --- |
|  | RMSE | SSIM | RMSE | SSIM |
| 3D TV | $1.28 \times 10^{-3}$ | 0.9346 | $1.20 \times 10^{-3}$ | 0.9387 |
| IM4D | $1.13 \times 10^{-3}$ | 0.9451 | $1.08 \times 10^{-3}$ | 0.9468 |
| SMEIR | $1.02 \times 10^{-3}$ | 0.9591 | $1.08 \times 10^{-3}$ | 0.9530 |
| (2,11) | $1.01 \times 10^{-3}$ | 0.9609 | $0.90 \times 10^{-3}$ | 0.9643 |
| (3,7) | $1.02 \times 10^{-3}$ | 0.9608 | $0.89 \times 10^{-3}$ | 0.9647 |
| (4,5) | $1.02 \times 10^{-3}$ | 0.9606 | $0.89 \times 10^{-3}$ | 0.9648 |

Table 2.3 The RMSE and SSIM values of different reconstruction methods at 10% dose for Phase 1 and Phase 4. (M, N) are for G-SMEIR.

|  | Phase 1 | | Phase 4 | |
| --- | --- | --- | --- | --- |
|  | RMSE | SSIM | RMSE | SSIM |
| 3D TV | $1.42 \times 10^{-3}$ | 0.9121 | $1.33 \times 10^{-3}$ | 0.9191 |
| IM4D | $1.18 \times 10^{-3}$ | 0.9385 | $1.11 \times 10^{-3}$ | 0.9408 |
| SMEIR | $1.07 \times 10^{-3}$ | 0.9556 | $1.12 \times 10^{-3}$ | 0.9506 |
| (2,11) | $1.06 \times 10^{-3}$ | 0.9565 | $0.95 \times 10^{-3}$ | 0.9599 |
| (3,7) | $1.07 \times 10^{-3}$ | 0.9565 | $0.94 \times 10^{-3}$ | 0.9603 |
| (4,5) | $1.07 \times 10^{-3}$ | 0.9563 | $0.94 \times 10^{-3}$ | 0.9604 |

### 2.3.4  Reconstruction accuracy across all phases

From the results of individual phases, G-SMEIR seems to gain only a marginal advantage over SMEIR for the reference phase and a greater advantage for the non-reference phase. This confirms our hypothesis that SMEIR's performance may deteriorate for other phases due to the local optimal trap of DVF. In this part,

we evaluated the quantitative measures for all phases. For G-SMEIR (M=1 and N=12), images for all phases were reconstructed simultaneously. For SMEIR, the reference phase was reconstructed directly, whereas the other phases were warped using the estimated DVF. Two phases, Phase 1 and Phase 4, were used as the reference phase as shown in the top row and the bottom row in Figure 2.9, respectively. As can be seen, SMEIR works similar to G-SMEIR in terms of RMSE and a little worse than G-SMEIR in terms of SSIM for the reference phase. However, for the phases other than the reference phase, G-SMEIR performs much better than SMEIR, usually more than 10% on RMSE. It is also interesting to note that the performance of full dose and half dose is comparable, which indicates that SMEIR and G-SMEIR are robust to increased noise and can be used to lower the radiation dose.
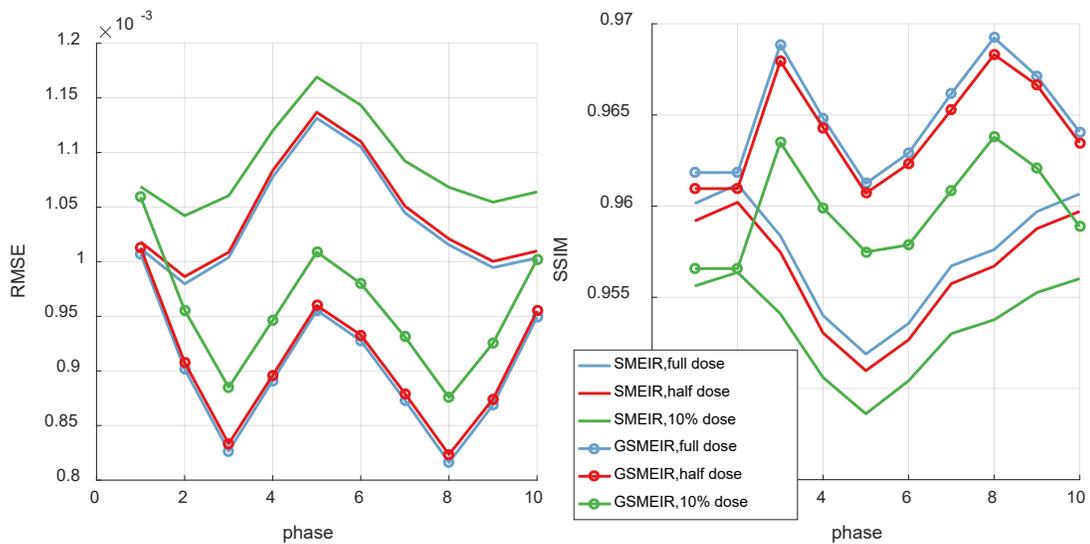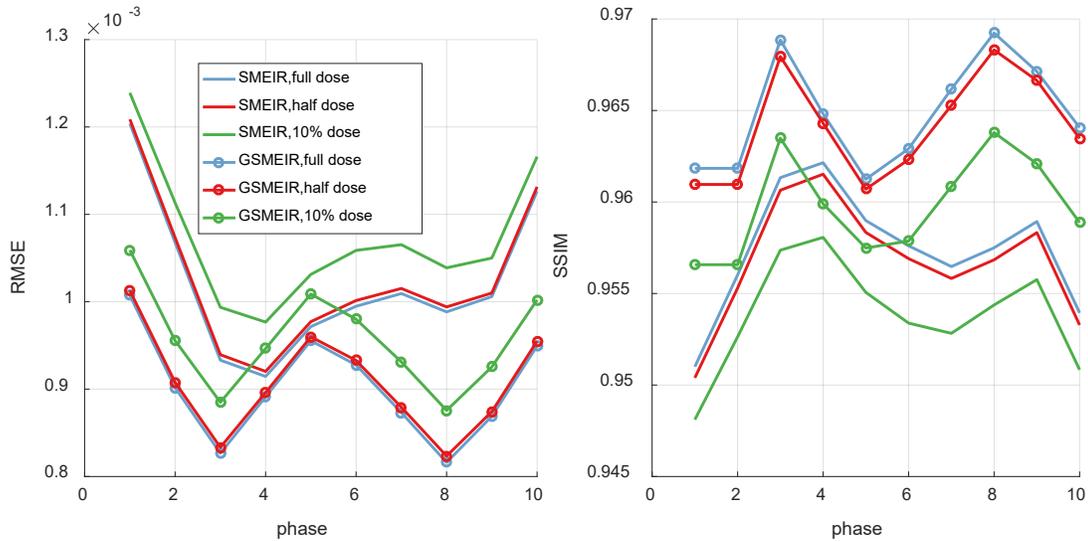
Figure 2.9  Quantitative results for different phases for SMEIR using Phase 1 (Top) and Phase 4 (Bottom) as the reference phase and G-SMEIR.

To further verify that the above behavior is general for using any phase as the reference phase, we used each phase of 10 phases as the reference phase for SMEIR and summarized RMSE and SSIM for each case at different dose levels. Since there are 10 phases, we calculated the mean and standard deviation values and listed them in Figure 2.10. As can be seen, the mean RMSE for SMEIR is always higher than that for G-SMEIR, and the mean SSIM for SMEIR is lower than that for G-SMEIR at all dose levels. The mean RMSE averaged over all ten cases for SMEIR are $1.02 \times 10^{-3}$ at full dose, $1.02 \times 10^{-3}$ at half dose, and $1.07 \times 10^{-3}$ at 10% dose. In contrast, the mean RMSE for G-SMEIR is much lower: $0.9 \times 10^{-3}$ at full dose, $0.91 \times 10^{-3}$ at half dose, and $0.95 \times 10^{-3}$ at 10% dose. In terms of SSIM, the corresponding values are 0.9577 at full dose, 0.9569 at half dose, and 0.9537 at 10% dose for SMEIR, and 0.9648 the full dose, 0.9641 at half dose, and 0.9598 at 10% dose for G-SMEIR. For both SMEIR and G-SMEIR, the performance at full dose and half dose is comparable and notable degradation occurs at 10% dose

This result demonstrates that G-SMEIR can reconstruct better images than SMEIR for all phases. The differences between SMEIR and G-SMEIR on RMSE and SSIM are statistically significant using a two-sample *t*-test.
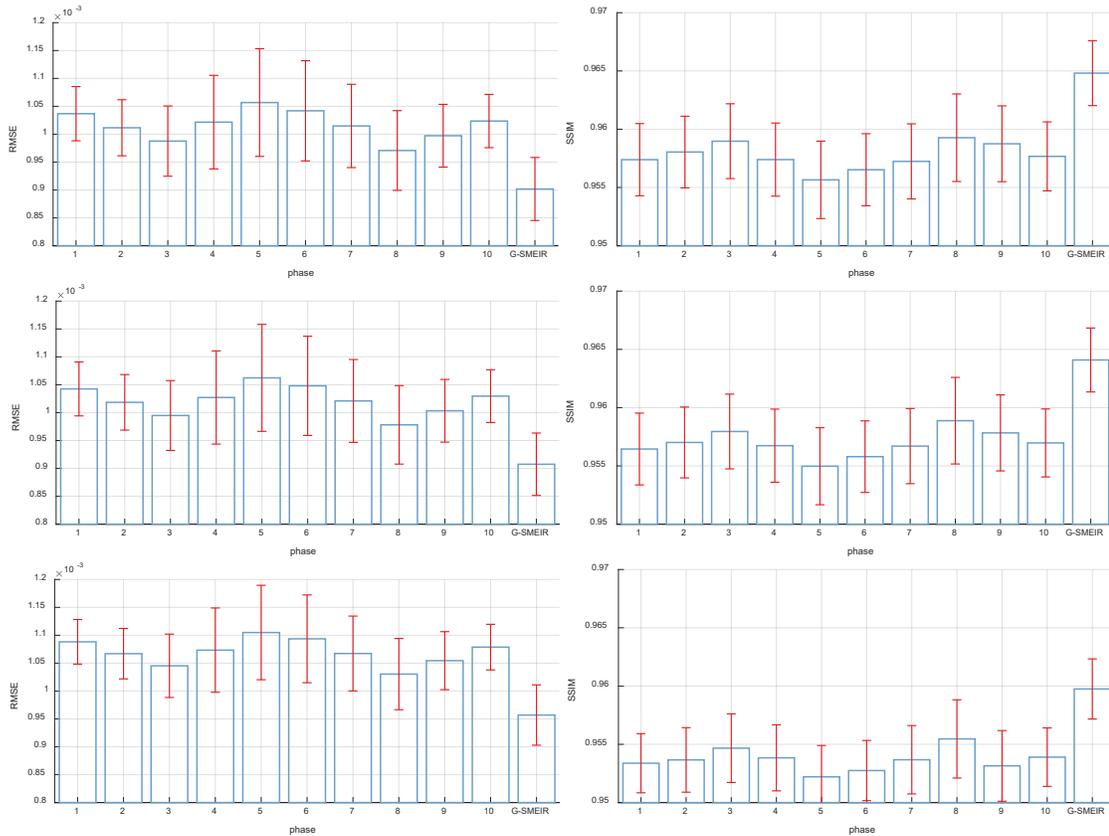


Figure 2.10 RMSE and SSIM for all phases for SMEIR (when varying the reference phase) and G-SMEIR. (Error bar represent standard deviation) (Top: full dose; middle: half dose; bottom: 10% dose)

### 2.3.5 Tumor motion recovery

The maximum ("MAX") and mean deviations (in voxels) from the tumor phantom motion obtained from DVFs are listed in Table 2.4. It can be seen that G-SMEIR usually achieves the smallest maximum and mean deviations compared to SMEIR and IM4D. It is also worth noting that although SMEIR has smaller maximum deviations than IM4D, its mean deviations are worse than IM4D. The overall

improvement of G-SMEIR over SMEIR in terms of mean deviation is 42%~47% at different dose levels. The motion displacements of the tumor in Anterior-Posterior (A-P), Left-Right (L-R), and Superior-Inferior (S-I) at full dose are shown in Figure 2.11, where the ordinate axis is the displacement (in voxels) and the abscissa axis is the number of phases. IM4D seems to recover the motion well in A-P and L-R directions, where the motion is small, but to become incapable of capturing the large motion in the S-I direction. In comparison, G-SMEIR not only recovers best the large motion in the S-I direction, but also improves the motion estimation in the other two directions over SMEIR. Note that the L-R motion was set as zero in the simulation. An abrupt transition around Phase 6 may be caused by residuals from the A-P motion for SMEIR and G-SMEIR. However, the amplitudes of these artificial L-R motions are small (less than ±1 voxel).

Table 2.4 The maximum and mean deviations (voxel) from the phantom tumor motion for different reconstruction methods. (M, N) is for G-SMEIR.

|  |  | MAX | Mean |
| --- | --- | --- | --- |
| Full dose | IM4D | 2.6785 | 0.5915 |
|  | SMEIR | 2.2684 | 0.7220 |
|  | (2,11) | 2.1405 | 0.5071 |
|  | (3,7) | 2.0782 | 0.4354 |
|  | (4,5) | 2.0771 | 0.4211 |
| Half dose | IM4D | 2.6642 | 0.5864 |
|  | SMEIR | 2.3128 | 0.7316 |
|  | (2,11) | 2.1553 | 0.5127 |
|  | (3,7) | 2.0827 | 0.4449 |
|  | (4,5) | 2.0730 | 0.4237 |
| 10% dose | IM4D | 2.6679 | 0.5767 |
|  | SMEIR | 2.3929 | 0.8079 |
|  | (2,11) | 2.1784 | 0.5288 |
|  | (3,7) | 2.0926 | 0.4480 |
|  | (4,5) | 2.0778 | 0.4288 |
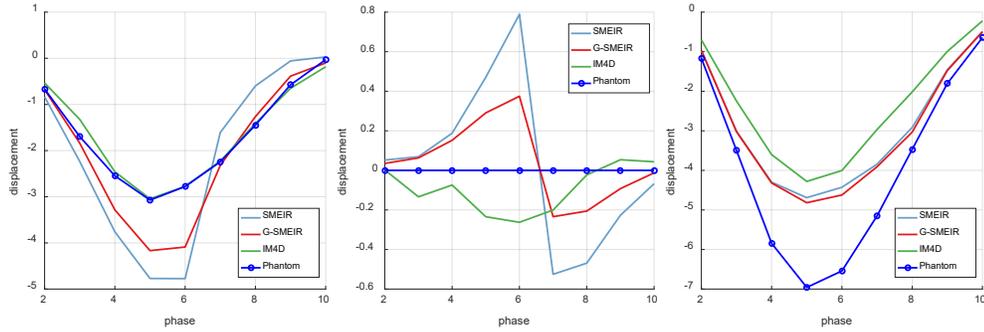
Figure 2.11 Motion trajectories in different directions for different reconstruction methods. (From left to right: A-P, L-R, and S-I).

## 2.3.6 Patient results

The reconstructed patient images for two phases are shown in Figure 2.12 (Phase 1) and Figure 2.13 (Phase 7) (with HU range of [-1000 4434]), respectively. IM4D, SMEIR, and G-SMEIR show much better image quality than 3D TV, which suffers more noise and blocky artifacts due to the limited views for each phase. For the reference phase (Figure 2.12, Phase 1), all three motion-compensated methods achieve a similar image quality, which is consistent with the findings in the phantom study. However, for Phase 7 in Figure 2.13, G-SMEIR not only maintains the contents in the lunges better (yellow arrows), but also suffers less motion artifacts (red arrows) than IM4D and SMEIR. Some ringing artifacts (e.g. blue arrow in Figure 2.12) in the sagittal view of MF-PMM methods can be seen due to non-optimized reconstruction parameters. Since 4D methods are much more time consuming than 3D TV, we were only able to tune parameters (ART step sizes: $\lambda, \lambda_{red}$, and TV minimization step sizes: $\gamma, \gamma_{red}$, see Appendix for definitions) for patient data and used the other parameters from the simulation study for MF-PMM

33

methods. When we lowered the TV regularization by reducing $\gamma$ and $\gamma_{red}$, we observed the dominance of noise and streak artifacts for 3D TV images, while the images from MF-PMM methods (IM4D, SMEIR, and G-SMEIR) are similar. Also, the ringing artifacts in 4D MF-PMM methods are alleviated for smaller TV regularization. This demonstrates that the strong denoising and sparse data recovery of 4D reconstruction than 3D reconstruction. Note that these images were acquired for the patient positioning purpose of radiation therapy and the projection views for each phase was only 30, thus their quality is not as good as diagnostic CT images and serves as a comparative purpose.
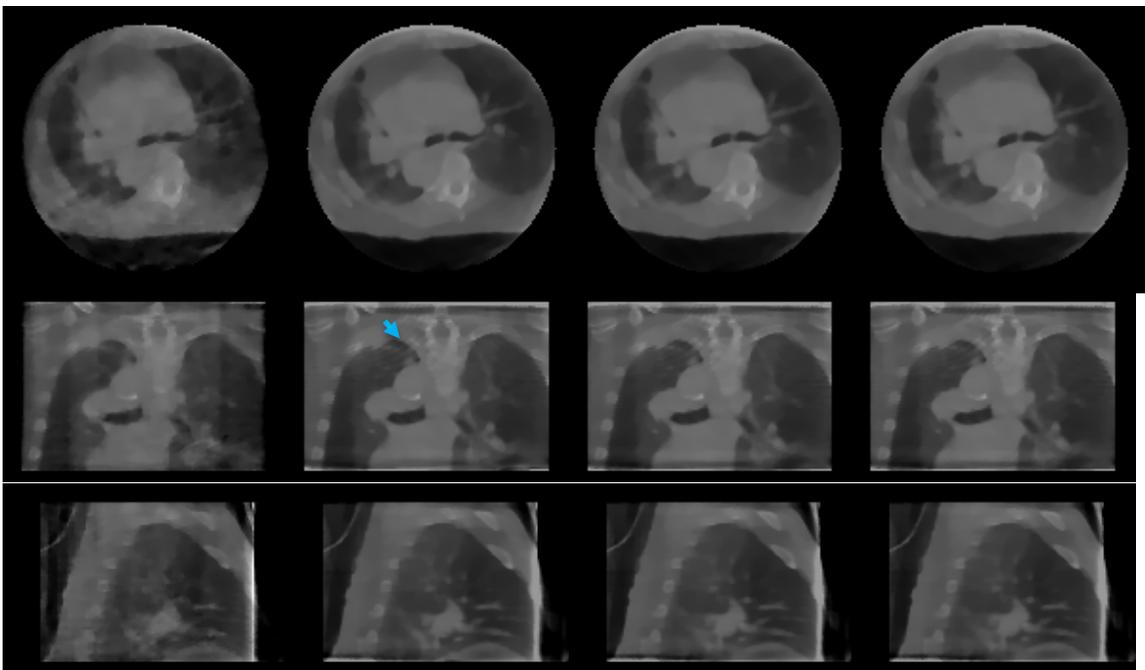


Figure 2.12 Phase 1 images of the patient for different methods. From top to bottom: transverse, coronal, and sagittal; from left to right: 3D TV, IM4D, SMEIR, and G-SMEIR.
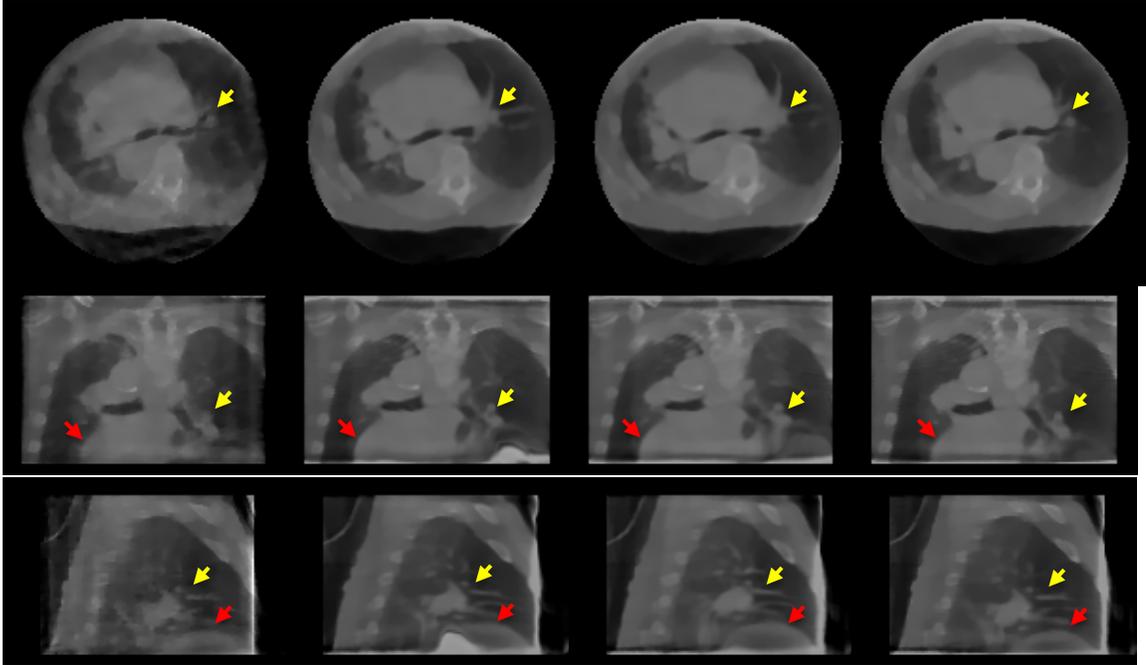
Figure 2.13 Phase 7 images of the patient for different methods. From top to bottom: transverse, coronal, and sagittal; from left to right: 3D TV, IM4D, SMEIR, and G-SMEIR.

## 2.4   Discussion

We also tested other combinations for G-SMEIR, e.g. (M, N) = (6, 3) and (8, 2). The performance is similar to those reported in Section 3. In general, we observed that the smaller M and the larger N leads to better RMSE, and the larger M and the smaller N leads to better SSIM. Although the improvement over RMSE and SSIM seems not to be substantial, G-SMEIR provides much better motion tracking of the tumor as indicated in Table 2.4, where the mean motion tracking error is reduced by more than 40% compared to SMEIR. The flexibility of G-SMEIR may provide an effective tool to boost the 4D reconstruction performance of other imaging modalities, such as CT, PET, and SPECT.

For image domain motion estimation, we mainly focused on improving the speed through the faster convergence and GPU implementation. Although the Demons algorithm was used in this work, more sophisticated motion estimation algorithms can be used to further improve the DVFs, thus the final reconstruction. It seems that the projection domain motion estimation using a symmetric form leads to better reconstruction for SMEIR in terms of RMSE and SSIM, whereas the image domain motion estimation using Demons leads to better tumor motion tracking for IM4D. G-SMEIR takes advantage of both image domain and projection domain motion estimation to achieve the best performance in all quantitative metrics as well as the appearance of reconstruction images.

It is also worth noting that the motion-compensated reconstruction methods belonging to MF-PMM (IM4D, SMEIR, and G-SMEIR) hold great potential for dose reduction. When the imaging dose was reduced from $10^5$ photons/incident ray to $5 \times 10^4$ photons/incident ray, the RMSE and SSIM values changed little. Only when the dose was reduced to $1 \times 10^4$ photons/incident ray, a few percent decrease on RMSE was observed. The strong denoising and data compression capability of these methods are achieved by using both spatial (TV minimization) and temporal (motion-compensated joint reconstruction) correlations in phase images. The superior reconstruction quality can be seen from the patient images, where each phase has only 30 projection views. Among three motion compensated reconstruction methods, G-SMEIR reveals more anatomic details than IM4D and SMEIR, which further demonstrates the power of combining both image domain and projection domain motion estimation for better reconstruction of 4D images.

The flat-panel detector in the XCAT phantom study was larger than the conventional one used in CBCT in order to cover the whole-body projection. By doing this, we can avoid the truncation in the projection domain and focus on studying the behavior of different reconstruction methods under an idealized condition. In real patient data, the half-scan was used due to the size of the flat-panel detector. Nevertheless, the reconstruction performance ranking of different methods is consistent with findings in the simulation study.

In this work, G-SMEIR was run on the Maverick2 GPU server at Texas Advanced Computing Center (TACC). The image domain motion estimation ran on GPUs (NVidia P100 GPU), while the projection domain motion estimation ran on CPUs (Intel® Xeon® Platinum 8160 CPU). The reconstruction of each phase of P phases was implemented in parallel on one of the P CPU cores, providing P times saving on computation time. Both two domain motion estimations run parallel to decrease time consumption. It takes about 1,000 seconds to complete image domain motion estimation, 220 seconds to complete joint ART reconstruction, and 1,600 seconds to complete projection domain motion estimation. Note that since I/O operations of large DVF files are included in the calculation, the time reported for image domain estimation is much longer than the runtime of the MRD algorithm. The biggest computational bottleneck is the projection domain motion estimation, which includes the optimization of motion objective functions and multiple projection and warping operations. It is expected that GPU parallel computing can significantly reduce the runtime for this part similar to image domain motion estimation. The parallelization of projection/backprojections will further reduce the ART operations.

Finally, a clever scheme of using DVFs is in need to avoid excessive I/O operations. In the future, we will investigate these possibilities to further lower the computation cost, which is essential for parameter tuning and selection of deformable registration models of G-SMEIR for better performance.

In summary we develop a G-SMEIR framework for MF-PMM to alleviate the local optimum trapping problem of 4D image reconstruction and accelerated the computational intense image domain motion estimation using GPU. The results using a 4D XCAT phantom and patient CBCT data demonstrate the superior reconstruction performance of G-SMEIR in a manageable time.

# CHAPTER 3.  SPATIOTEMPORAL DENOISING OF LOW-DOSE CARDIAC CT IMAGES USING RECYCLEGAN

Shiwei Zhou[1], Jinyu Yang[2], Krishnateja Konduri[3], Junzhou Huang[2], Lifeng Yu[4], and Mingwu Jin[1]

[1] Department of Physics, University of Texas at Arlington, Arlington, Texas USA

[2] Department of Computer Science and Engineering, University of Texas at Arlington, Arlington, Texas USA

[3] Department of Bioengineering, University of Texas at Arlington, Arlington, Texas USA

[4] Department of Radiology, Mayo Clinic, Rochester, Minnesota USA

## 3.1 Introduction

To avoid risks from cardiac catheterization of invasive coronary angiographies (ICAs) in low- and intermediate-risk coronary artery disease (CAD) patients, multi-detector computed tomography (MDCT) has been used for CT angiography (CTA) to noninvasively assess the presence, location, severity, and characteristics of coronary atherosclerosis [52-54]. In addition, some findings from CTA may not be detectable by ICA [55-57]. The main challenge in CTA is the strong demand on high temporal resolution (to mitigate cardiac motion artifacts) and high spatial resolution (for small coronary structures), which leads to high radiation dose [58]. Electrocardiogram- (ECG-) gated multi-phase CTA (MP-CTA), either in a retrospective helical scan mode or a prospective axial scan mode, can provide much more clinically relevant information than single-phase CTA (SP-CTA). Not only is the important heart function information lost in SP-CTA, but also different parts of the coronary arteries are better seen in different phases [59]_ENREF_49. Thus, MP-CTA may be preferred for much greater diagnostic value than SP-CTA. However, even with ECG tube current modulation (TCM), the average effective dose of a MP-CTA scan could be much higher than 10 mSv [60]_ENREF_50 (6~24 mSv at Mayo Clinic), depending on the width of the pulse window and patient size [61]_ENREF_51. Taking 80% patients with negative findings into account, minimizing the radiation dose becomes a major and urgent need for a broader application of MP-CTA for CAD diagnosis.

Many methods have been developed to reduce radiation dose in CT acquisition, including optimization of tube current, tube potential, and use of dedicated bowtie

filters. However, X-ray dose reduction in general will lead to elevated noise in reconstructed images. The noise in the low-dose CT (LDCT) images can be reduced by either conventional reconstruction methods [62-69], or emerging deep-learning based denoising methods directly on images after regular reconstruction through paired-image training [70-72] or unpaired images training using a cycle-consistent generative adversarial network (CycleGAN) [73-76]. Several CycleGAN variants and a paired deep learning method (RED-CNN)[70] for LDCT denoising were investigated and compared [75]. However, all these deep learning denoising methods treated each CT image independently and failed to count for the temporal correspondence between images, such as that of MP-CTA image sequences.

To our best knowledge, CycleGAN with an identity loss [73] or wavelet-assisted noise disentanglement [76] was the first work to use deep-learning methods to improve low-dose MP-CTA images. Although CycleGAN can achieve the translation between LDCT and full-dose CT (FDCT) without the need of paired training images, the translation is established only in the spatial domain. The temporal connections along the different cardiac phases of a MP-CTA image sequence are not utilized by CycleGAN and may lead to sub-optimal denoising performance. On the other hand, an advanced CycleGAN model with a recurrent loss and a cycle consistency loss over space and time ("recycle loss"), so called RecycleGAN [77], was proposed to achieve video-to-video translation in computer vision, which utilizes both spatial and temporal information to solve the translation problem of temporally related data. Nevertheless, RecycleGAN has never been applied to denoise low-dose CT image sequences including MP-CTA. In this work,

41

we adapt RecycleGAN to take into consideration of the temporal connection of the succeeding cardiac phases of MP-CTA images. This novel deep learning denoising method not only enjoys the advantage of CycleGAN without need of paired training images, but also exploits both spatial and temporal correspondence to boost denoising performance for time series of MP-CTA images. As our aim in this work focuses on comparing the denoising performance of CycleGAN and RecycleGAN for low-dose MP-CTA images, the comparison between CycleGAN and other traditional and deep learning methods for LDCT denoising can be found in the previous works, such as [75].

## 3.2 Methods

### 3.2.1 CycleGAN

To achieve image-to-image translation, CycleGAN [78] is proposed to learn mapping functions between two different domains without the need of paired data. Formally, given a set of images from a source domain $A$ (e.g., low-dose CT images) and a set of images from a target domain $B$ (e.g., full-dose CT images), the goal of CycleGAN is to learn a mapping $G_{AB}: A \rightarrow B$, such that the output $G_{AB}(a)$ is indistinguishable from the images in domain $B$. The architecture of CycleGAN is composed of two generators and two adversarial discriminators (Figure 3.1). Specifically, each generator aims to translate images from one domain to the other domain, while each discriminator is designed to distinguish between real samples in the target domain from the translated images.

The objective of CycleGAN contains two terms: an adversarial loss[79] and a cycle consistency loss[78]. Given data distribution $a \sim p_{data}(a)$ and $b \sim p_{data}(b)$, the adversarial loss is designed to match the distribution of generated images $G_{AB}(a)$ to the distribution of the target domain $B$. The objective of $G_{AB}$ and $D_B$ is defined as:

$$\mathcal{L}_{GAN}(G_{AB}, D_B, A, B)$$
$$= \mathbb{E}_{b \sim P_{data}(b)}[\log D_B(b)] \qquad (3.1)$$
$$+ \mathbb{E}_{a \sim P_{data}(a)}\left[\log\left(1 - D_B\left(G_{AB}(a)\right)\right)\right]$$

where $G_{AB}$ aims to minimize this objective against an adversary $D_B$ that tries to maximize it, i.e., $\min_{G_{AB}} \max_{D_B} \mathcal{L}_{GAN}(G_{AB}, D_B, A, B)$ [78]. Similarly, for the generator $G_{BA}$, the objective is

$$\mathcal{L}_{GAN}(G_{BA}, D_A, A, B)$$
$$= \mathbb{E}_{a \sim P_{data}(a)}[\log D_A(a)] \qquad (3.2)$$
$$+ \mathbb{E}_{b \sim P_{data}(b)}\left[\log\left(1 - D_A\left(G_{BA}(b)\right)\right)\right]$$

To further reduce the space of possible mapping functions [79], cycle-consistency loss is introduced to guarantee the output of each cycle to be close to the input to that cycle, i.e., $G_{BA}\left(G_{AB}(a)\right) \approx a$. The objective is defined as

$$\mathcal{L}_{cycle}(G_{AB}, G_{BA})$$

$$= \mathbb{E}_{a \sim P_{data}(a)} \left[ \left\| G_{BA}(G_{AB}(a)) - a \right\|_1 \right] \qquad (3.3)$$

$$+ \mathbb{E}_{b \sim P_{data}(b)} \left[ \left\| G_{AB}(G_{BA}(b)) - b \right\|_1 \right]$$

This cycle-consistency loss enforces the constraint that $G_{AB}$ and $G_{BA}$ be inverse of each other[73].

Taken together, the overall objective loss for CycleGAN is:

$$\mathcal{L}_{cycleGAN}(G_{AB}, G_{BA}, D_A, D_b)$$

$$= \mathcal{L}_{GAN}(G_{AB}, D_B, A, B)$$

$$+ \mathcal{L}_{GAN}(G_{BA}, D_A, B, A) \qquad (3.4)$$

$$+ \lambda \mathcal{L}_{cycle}(G_{AB}, G_{BA})$$

where $\lambda$ controls the importance of each objective term.

The variants of CycleGAN [78] has been applied to various domains [75] . However, they only use the spatial information in 2D images, and do not consider the temporal information for optimization [77].

The cycle-consistency loss forces the optimization to learn a solution that is closely tied to the input. This is suitable for the situation that only spatial information is available during the translation, while for time-related image sequences, such as CTA images, with only the cycle consistency, the model may be inadequate to generate perceptually unique results. The network structure of CycleGAN used in this work is based on Fig. 4 and Fig. 5 in [75].
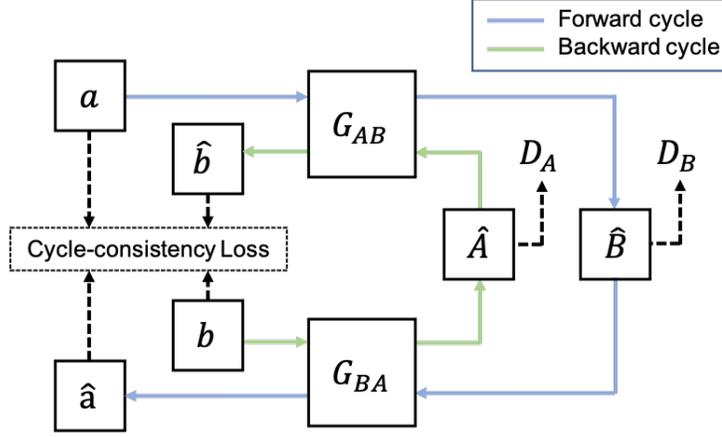
Figure 3.1 The network structure of CycleGAN.

In the forward cycle (blue line), an image $a$ from domain $A$ is translated to domain $B$ by generator $G_{AB}$, expressed as $\hat{B} = G_{AB}(a)$. Then, $\hat{B}$ is translated back to domain $A$, expressed as $\hat{a} = G_{BA}(G_{AB}(a))$. The backward cycle (green line) has similar operations where image $b$ in domain $B$ is mapped to domain $A$ as $\hat{A} = G_{BA}(b)$ and then mapped back to domain $B$ as $\hat{b} = G_{AB}(G_{BA}(b))$.

## 3.2.2 RecycleGAN

RecycleGAN [77] is proposed to learn a mapping between two videos from different domain. It utilizes both spatial and temporal information to solve the reconstruction problem of temporally related data. RecycleGAN shares similar model framework with CycleGAN, while the cycle-consistency loss is replaced by recurrent loss and recycle loss to make use of the temporal ordered images and learn better mapping. The network structure is shown in Figure 3.2.

Given unpaired but ordered streams $(a_1, a_2, \dots, a_t, \dots) \in A$ (e.g., temporally ordered low-dose CT images) and $(b_1, b_2, \dots, b_s, \dots) \in B$ (e.g., temporally ordered full-dose

45

CT images), the recurrent temporal predictor $P_A$ is trained to predict future sample given the past. The recurrent loss is defined as:

$$\mathcal{L}_\tau(P_A) = \sum_t \|a_{t+1} - P_A(a_{1:t})\|^2 \tag{3.5}$$

where $a_{1:t} = (a_1, \ldots, a_t)$. Then the recycle loss can be defined using this temporal prediction model. The objective that across domains and time is expressed as:

$$\mathcal{L}_r(G_{BA}, G_{AB}, P_B) = \sum_t \|a_{t+1} - G_{BA}(P_B(G_{AB}(a_{1:t})))\|^2 \tag{3.6}$$

where $G_{AB}(a_{1:t}) = (G_{AB}(a_1), G_{AB}(a_2), \ldots, G_{AB}(a_t))$. In both forward and backward cycles, the above loss requires a sequence of image frames to map back to the initial domain. The overall loss is defined by:

$$
\begin{aligned}
\mathcal{L}_{recycleGAN}&(G, P, D) \\
&= \mathcal{L}_{GAN}(G_{AB}, D_B, A, B) \\
&+ \mathcal{L}_{GAN}(G_{BA}, D_A, B, A) \\
&+ \lambda_{rx}\mathcal{L}_r(G_{BA}, G_{AB}, P_B) \\
&+ \lambda_{ry}\mathcal{L}_r(G_{AB}, G_{BA}, P_A) + \lambda_{\tau x}\mathcal{L}_\tau(P_A) \\
&+ \lambda_{\tau y}\mathcal{L}_\tau(P_B)
\end{aligned}
\tag{3.7}
$$

where $\lambda$'s control the importance of the losses. We show in the experiments that the proposed method provides an effective translation from low-dose MP-CTA to full-dose MP-CTA images when learning from unpaired CT image series. The detailed network structure of RecycleGAN [77] can be found in Appendix D.
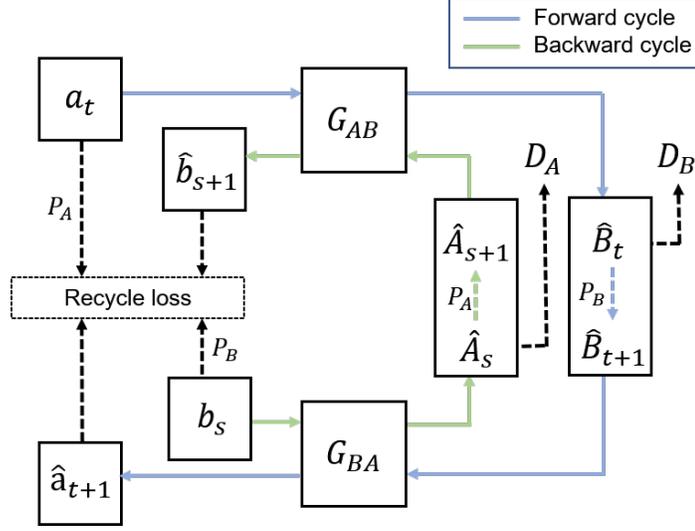
Figure 3.2 The framework of RecycleGAN.

In the forward cycle (blue line), an image $a_t$ at time $t$ from domain $A$ is translated to domain $B$ by generator $G_{AB}$, expressed as $\widehat{B_t} = G_{AB}(a_t)$. Then, a temporal predictor $P_B$ is applied on $\widehat{B_{1:t}}$ to predict a future frame $\widehat{B}_{t+1}$ and then $\widehat{B}_{t+1}$ is translated back to domain $A$, expressed as $\widehat{a_{t+1}} = G_{BA}(P_B(G_{AB}(a_t)))$. The backward cycle (green line) has similar operations where image $b_s$ in domain $B$ is mapped to domain $A$ as $\widehat{A_s} = G_{BA}(b_s)$ and then mapped back to domain $B$ with a temporal predictor $P_A$, expressed as $\widehat{b_{s+1}} = G_{AB}(P_A(G_{BA}(b_s)))$.

## 3.3   Experimental setting

### 3.3.1  Phantom Data

We used the XCAT phantom program [80] based 18 patients' data (nine females and nine males) to generate cardiac CT images (thorax 512x512x128, voxel size of 1 mm³) for two different dose levels: full-dose and low-dose (20% of the full-dose). The number of phases for each cardiac cycle is set to eight. The 18

phantoms were divided into nine pairs of female and male. To generalize the performance of CycleGAN and RecycleGAN, we used the 9-fold cross-validation (CV). For each CV, the training dataset contains seven pairs of female and female phantoms, the validation dataset contains a pair of male and female phantoms, and the testing dataset contains another pair of male and female phantoms. The Table 3.1 shows the patient pairs for nine CV sets. For each CV, the network was trained using the training data and the hyperparameters were tuned using the validation data. Afterward, the optimal hyperparameters were used to train the network using both training and validation datasets. Finally, the denoising performance was evaluated on the test dataset. To account for the temporal relationship among cardiac phases, the images of each slice are viewed as a looped video of eight frames.

Table 3.1 The CV setting

| CV number | Training patients | Validation patients | Testing patients |
|---|---|---|---|
| 1 | #2, #3, #4, #5, #6, #7, #9 | #8 | #1 |
| 2 | #1, #3, #4, #5, #6, #7, #8 | #9 | #2 |
| 3 | #1, #4, #5, #6, #7, #8, #9 | #2 | #3 |
| 4 | #1, #2, #3, #5, #7, #8, #9 | #6 | #4 |
| 5 | #1, #2, #4, #6, #7, #8, #9 | #3 | #5 |
| 6 | #1, #2, #3, #4, #5, #8, #9 | #7 | #6 |
| 7 | #1, #2, #3, #5, #6, #8, #9 | #4 | #7 |
| 8 | #2, #3, #4, #5, #6, #7, #9 | #1 | #8 |
| 9 | #1, #2, #3, #4, #6, #7, #8 | #5 | #9 |

### 3.3.2  Patient Data

We also used the real patient CTA images from Mayo Clinic to evaluate the performance of RecycleGAN. CTA images of 50 patients were retrospectively collected and deidentified (IRB was approved by Mayo Clinic). Intravenous

iodinated contrast (Omnipaque® 350) was injected using a bolus tracking technique, where the volume and injection rate were determined by the patient weight, followed by 10 cc saline chaser. The arterial attenuation enhancement is 200~350 HU. These cases were acquired using a routine retrospectively ECG-gated helical scanning technique on a 3rd generation 192-slice dual-source scanner (Force, Siemens Healthcare): 0.25 sec rotation time, 192x0.6 mm detector configuration, helical pitch automatically selected based on heart rate, tube potential automatically determined (CAREkV), TCM (CAREDose4D, maximum tube current (MTC) 180 mAs in the pulse window and 20% outside), and ECG-pulsing at 40%-70% phases. These parameters may vary for some patients, especially for those with irregular heartbeat. The CTDIvol was varying from patient to patient depending on the patient size, heart rate, and regularity of the heart rate (31~120 mGy, i.e. 6~24 mSv). For irregular heart rate, the pulsing window may be extended automatically, which could dramatically increase radiation dose. 3D volume images (512x512 in plane, 300~375 slices, isotropic 0.4 mm size) at 20 phases (0%-95%) was reconstructed using the Siemens ADMIRE algorithm with a Qr40 kernel (ADMIRE strength setting of 3). Therefore, in 20 phases of CTA images of each patient, roughly 6 phases are of full dose (with MTC) while the remaining 14 phases are of low dose (with 20% MTC). Due to the patient size, heartbeat irregularity, and unbalanced full-dose and low-dose slices (# of full-dose slices << # of low-dose slices), we selected the full-dose slices and the low-dose slices for training (48 patients out of 50) based on the standard deviation (STD) of a square region in the aorta (full dose < 39 HU and low dose > 59 HU) and at least

three consecutive phases falling into either full dose or low dose. To keep the underlying data the same, we selected 16.2 thousand low-dose images and 15.8 thousand full-dose images for CycleGAN training, while we selected 15.8 thousand low-dose frames and 15.2 thousand full-dose frames for RecycleGAN training. The difference was caused by the requirement of three consecutive phases for RecycleGAN training, which was not satisfied by all CycleGAN training images. To tune the model hyperparameters, CTA images of one patient were used for the validation set. The remaining one patient dataset was served as the test set for performance evaluation.

### 3.3.3 Evaluation Metrics

To evaluate the proposed method, peak signal-to-noise ratio (PSNR) [81] and structural similarity index (SSIM) [82, 83] are used as quantitative measurements for the XCAT phantom data. The PSNR is an expression for the ratio between the (denoised) low-dose CT image $x$ and the corresponding full-dose CT image $y$ as follows,

$$PSNR = 10 \log_{10}\left(\frac{MAX_Y^2}{MSE}\right) \qquad (3.8)$$

where $MAX_Y$ is the maximum signal value that is set as 4095 for 12-bit CT images in our experiments. The term "MSE" stands for mean squared error and is defined as,

$$MSE = \frac{1}{mn} \sum_{i=0}^{m-1} \sum_{j=0}^{n-1} [x(i,j) - y(i,j)]^2 \qquad (3.9)$$

Where $i$ and $j$ are the row and column indices of low-dose CT image $x$ and the corresponding full-dose CT image $y$, respectively, and $m$ and $n$ represent the number of rows the number of columns, respectively. The PSNR measures the cumulative difference between two images. The higher the PSNR, the better the performance of the denoising.

In addition to PSNR, the SSIM is designed to compare luminance, contrast, and structure difference between two images and is defined as,

$$SSIM(x,y) = l(x,y)c(x,y)s(x,y) \qquad (3.10)$$

where $l(x,y) = \frac{2\mu_x\mu_y + c_1}{\mu_x^2 + \mu_y^2 + c_1}$, $c(x,y) = \frac{2\sigma_{xy} + c_2}{\sigma_x^2 + \sigma_y^2 + c_2}$, and $s(x,y) = \frac{\sigma_{xy} + c_3}{\sigma_x\sigma_y + c_3}$. The first term $l(x,y)$ measures closeness of mean luminance $\mu_x$ and $\mu_y$. The contrast $c(x,y)$ is measured by standard deviation $\sigma_x$ and $\sigma_y$. The structure similarity $s(x,y)$ is measured by correlation coefficient between images $x$ and $y$. $\sigma_{xy}$ is the covariance between two images. The $c_1$, $c_2$ and $c_3$ are used to stabilize the division operation[82, 83] . The higher SSIM value indicates the closer resemblance of two images.

For patient data, since the ground truth was unknown, the performance was evaluated using STD in a square region of the aorta of CTA images of the test

patient, where the uniform intensity is expected. Therefore, the lower STD, the better denoising performance.

### 3.3.4 Hyperparameters

Hyperparameters of CycleGAN and RecycleGAN were generally kept the same as the previous publications[75, 77]. Specifically, for CycleGAN $\lambda$ was set to 10, while for RecycleGAN, $\lambda_{rx}$ was set to 0.5, and $\lambda_{ry}$ was set to 50, $\lambda_{\tau x}$ was set to 1, and $\lambda_{\tau y}$ was set to 100. The networks were trained with random weights from scratch using the Adam solver. For each model, we searched for the best learning rate in the range of $5.00 \times 10^{-6}$ to $1.26 \times 10^{-3}$ based on the lowest PSNR of the validation set (a pair of female and male patients for the phantom data and one patient for the patient data). For the phantom data, after training each CV data set, the best performing model was applied on the test dataset for performance evaluation. For the patient data, the learning rate was tuned using the validation patient and the best model was applied on the test patient.

## 3.4 Results

### 3.4.1 Phantom results

We compared our proposed spatiotemporal RecycleGAN method with CycleGAN using PSNR and SSIM as quantitative metrics. Figure 3.3 shows PSNR changes of the validation set along with different learning rates for nine CV sets. We separated the female and male validation PSNR as some large differences were found between the genders (see Table 3.2 and Table 3.3). For CycleGAN, the

learning rates $2\times10^{-5}$ to $3\times10^{-4}$ seem to have a PSNR plateau for the validation set. For RecycleGAN, this range narrows to $3\times10^{-5}$ to $3\times10^{-4}$. The best validation PSNR for each CV set was listed in Table 3.2 along with SSIM. First, the different PSNR and SSIM performance can be clearly seen between female and male validation patients. In most cases for CycleGAN, the PSNR differences are 2-6 dB except for CV7 (less than 1 dB), while SSIM difference is ranged from more than 0.01 to about 0.07. This difference is mainly caused by the learning rate was tuned based on the overall PSNR using both female and male validation patients. Although the differences are also observed for RecycleGAN metrics, they are notably smaller. RecycleGAN outperformances CycleGAN in almost all cases, except for CV7 male SSIM (marked as bold blue in Table 3.2). After taking the average values ($\pm$ Standard Deviation) of nine CV sets, the PSNR and SSIM for CycleGAN are $41.23\pm2.16$ dB and $0.9462\pm0.0241$ for the female, and $41.13\pm1.62$ dB and $0.9526\pm0.0188$ for the male. The corresponding numbers for RecycleGAN are $41.71\pm2.07$ dB and $0.9523\pm0.0224$ for the female, and $42.10\pm1.17$ dB and $0.9600\pm0.0108$ for the male. RecycleGAN achieves not only the greater average values, but also the smaller variances than CycleGAN. The best models were then applied to the test dataset and the PSNR and SSIM results are shown in Table 3.3. The similar findings to the best validation metrics are observed although the number of cases of that RecycleGAN is worse than CycleGAN slightly increases. RecycleGAN outperformances CycleGAN in most cases, except for PSNR of CV1 male and CV8 female (marked as bold blue in Table 3.3). The PSNR and SSIM for CycleGAN are $40.36\pm2.23$ dB and $0.9431\pm0.0250$ for the female test data, and

40.91±2.16 dB and 0.9501±0.0208 for the male test data. The corresponding numbers for RecycleGAN are 40.84±2.05 dB and 0.9512±0.0215 for the female test data, and 41.43±2.11 dB and 0.9572±0.0178 for the male test data. The test results demonstrated again that RecycleGAN can achieve better denoising performance than CycleGAN.
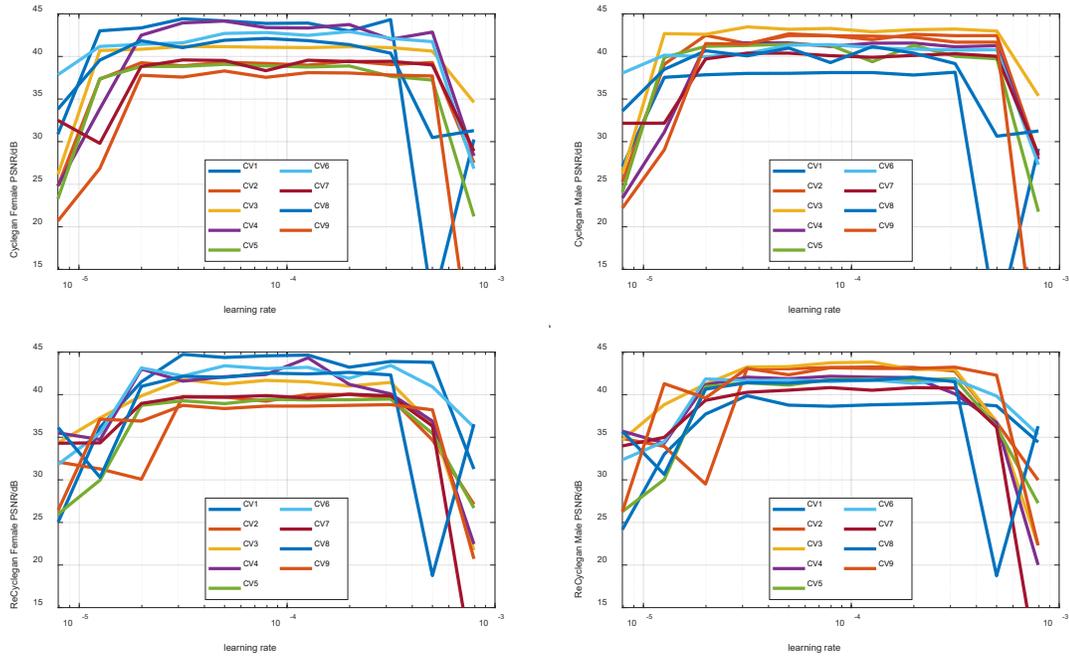


Figure 3.3 The validate PSNR changes for CycleGAN and RecycleGAN with different learning rates. (Top row: CycleGAN; Bottom row: RecycleGAN. Left column: Female; Right column: male)

Table 3.2 The best validation metrics for CycleGAN and RecycleGAN

| Cross-Validation | CycleGAN | | | | RecycleGAN | | | |
|---|---|---|---|---|---|---|---|---|
| | Female PSNR | Female SSIM | Male PSNR | Male SSIM | Female PSNR | Female SSIM | Male PSNR | Male SSIM |
| 1 | 44.42 | 0.9761 | 38.01 | 0.9071 | 44.73 | 0.9814 | 39.88 | 0.9394 |
| 2 | 39.32 | 0.9284 | 42.46 | 0.9658 | 40.02 | 0.9367 | 43.25 | 0.9710 |
| 3 | 41.17 | 0.9494 | 43.47 | 0.9729 | 41.75 | 0.9575 | 43.83 | 0.9745 |
| 4 | 44.16 | 0.9762 | 41.59 | 0.9547 | 44.32 | 0.9778 | 42.05 | 0.9592 |
| 5 | 39.08 | 0.9258 | 41.45 | 0.9552 | 39.48 | 0.9303 | 41.87 | 0.9578 |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| 6 | 42.92 | 0.9679 | 40.86 | 0.9549 | 43.44 | 0.9723 | 41.77 | 0.9603 |
| 7 | 39.59 | 0.9265 | 40.38 | 0.9609 | 40.06 | 0.9335 | 40.83 | 0.9458 |
| 8 | 42.11 | 0.9600 | 39.28 | 0.9359 | 42.62 | 0.9640 | 42.04 | 0.9611 |
| 9 | 38.30 | 0.9060 | 42.66 | 0.9654 | 38.83 | 0.9153 | 43.23 | 0.9696 |
| Average | 41.23 | 0.9462 | 41.13 | 0.9526 | 41.71 | 0.9523 | 42.10 | 0.9600 |
| Standard Deviation | 2.16 | 0.0241 | 1.62 | 0.0188 | 2.07 | 0.0224 | 1.17 | 0.0108 |

Table 3.3 Quantitative metrics for the test data for CycleGAN and RecycleGAN

| | CycleGAN | | | | RecycleGAN | | | |
|---|---|---|---|---|---|---|---|---|
| Cross-Validation | Female PSNR | Female SSIM | Male PSNR | Male SSIM | Female PSNR | Female SSIM | Male PSNR | Male SSIM |
| 1 | 42.18 | 0.9601 | 41.31 | 0.9562 | 42.33 | 0.9665 | 41.28 | 0.9620 |
| 2 | 41.03 | 0.9467 | 43.26 | 0.9706 | 41.70 | 0.9533 | 43.83 | 0.9749 |
| 3 | 39.04 | 0.9245 | 41.44 | 0.9543 | 40.40 | 0.9474 | 42.67 | 0.9677 |
| 4 | 39.54 | 0.9261 | 40.34 | 0.9408 | 40.00 | 0.9321 | 40.71 | 0.9452 |
| 5 | 37.92 | 0.8993 | 42.64 | 0.9645 | 38.74 | 0.9116 | 42.93 | 0.9687 |
| 6 | 37.40 | 0.9636 | 37.16 | 0.9396 | 37.55 | 0.9668 | 37.60 | 0.9495 |
| 7 | 43.16 | 0.9680 | 41.36 | 0.9547 | 43.62 | 0.9711 | 41.96 | 0.9594 |
| 8 | 43.57 | 0.9721 | 37.77 | 0.9026 | 43.29 | 0.9761 | 38.71 | 0.9174 |
| 9 | 39.44 | 0.9274 | 42.90 | 0.9672 | 39.94 | 0.9358 | 43.19 | 0.9702 |
| Average | 40.36 | 0.9431 | 40.91 | 0.9501 | 40.84 | 0.9512 | 41.43 | 0.9572 |
| Standard Deviation | 2.23 | 0.0250 | 2.16 | 0.0208 | 2.05 | 0.0215 | 2.11 | 0.0178 |

We show an image (Phase 5) of the test female and male data denoised by CycleGAN and RecycleGAN for CV1, CV2 and CV3 in Figure 3.4 (female) and Figure 3.5 (male), respectively. The full-dose and low-dose images are also shown as reference. Both CycleGAN and RecycleGAN effectively remove the noise in the images. RecycleGAN has less noise and is closer to the full-dose images than CycleGAN as shown in Figure 3.4 and Figure 3.5.
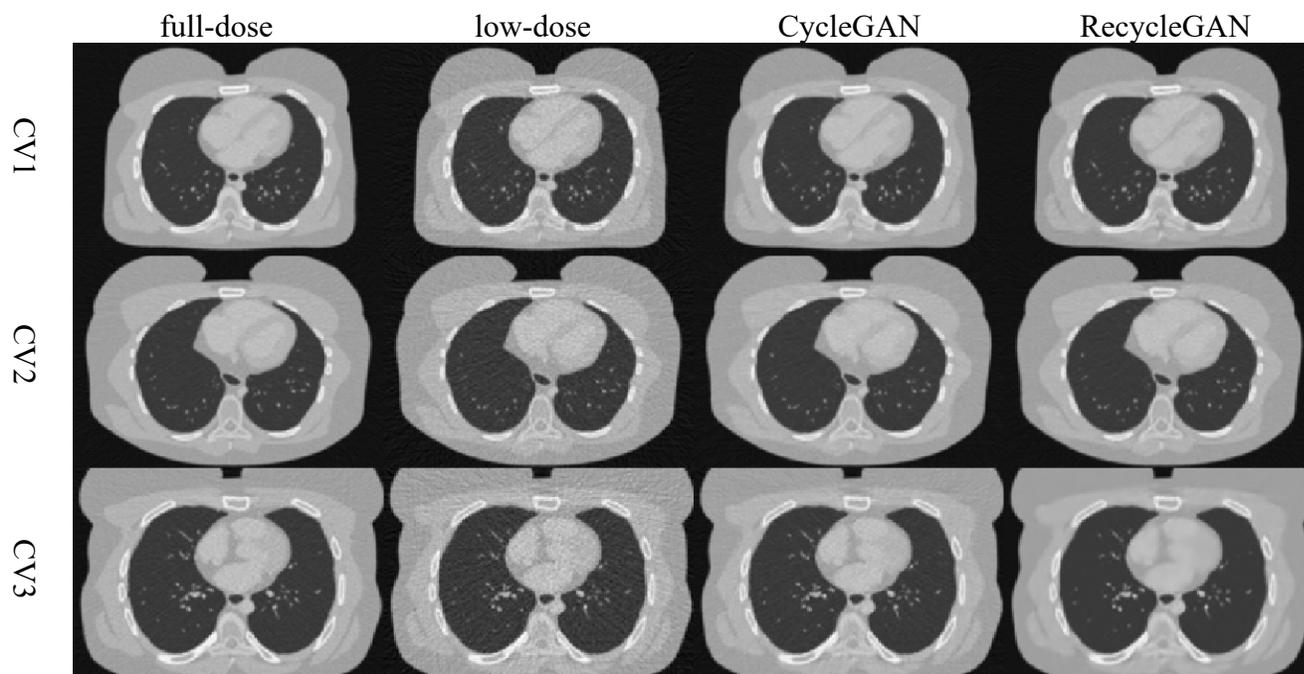
Figure 3.4 The transverse slice of phase 5 in the testing dataset for female (From top to bottom: CV1, CV2, and CV3; from left to right: full-dose, low-dose, CycleGAN, and RecycleGAN).
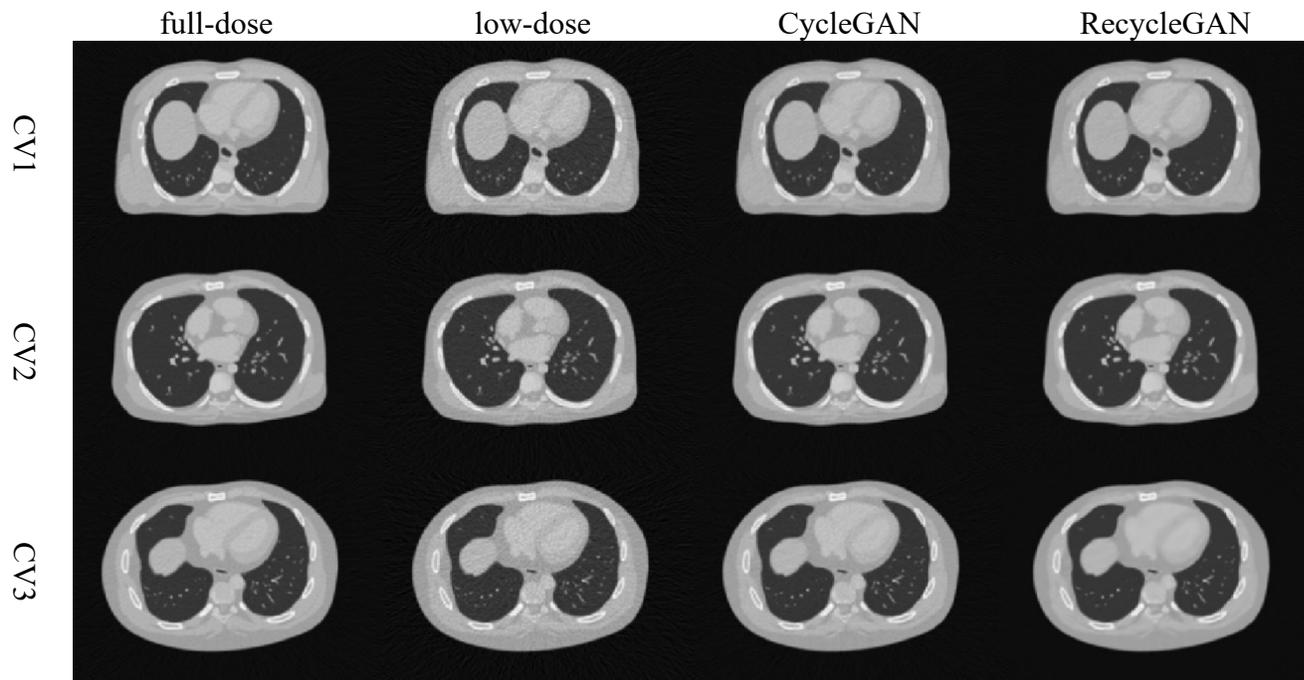


Figure 3.5 The transverse slice of phase 5 in the testing dataset for male (From top to bottom: CV1, CV2, and CV3; from left to right: full-dose, low-dose, CycleGAN, and RecycleGAN).
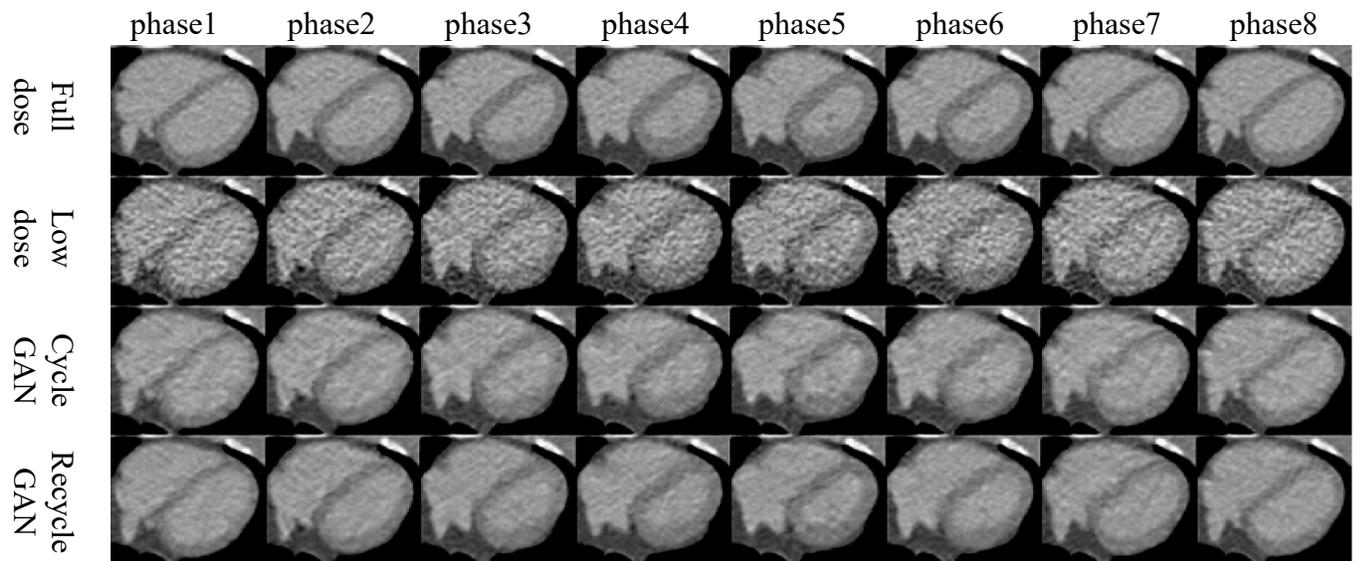
Figure 3.6  Eight phases of the heart region for CV2 female test data (From top to bottom: full-dose, low-dose, CycleGAN, and RecycleGAN; from left to right: phase 1 to phase 8. Display Window [-215 335]HU).

In Figure 3.6 and Figure 3.7, the eight phases of the heart region are shown for different methods along with the full-dose and low-dose references. Again, both CycleGAN and RecycleGAN effectively suppress the noise. RecycleGAN does a better job to further remove the noise than CycleGAN in the myocardium and the blood pool. RecycleGAN also achieves better contrast and structure preservation than CycleGAN.
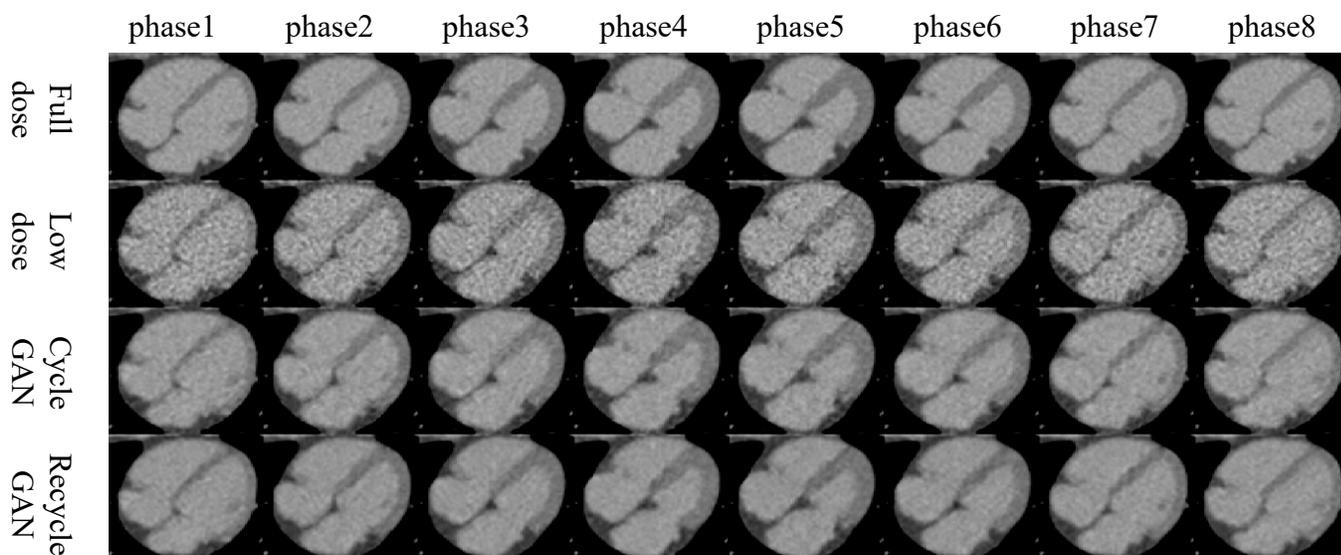
Figure 3.7 Eight phases of the heart region for CV2 male test data (From top to bottom: full-dose, low-dose, CycleGAN, and RecycleGAN; from left to right: phase 1 to phase 8. Display Window [-215 335]HU).

## 3.4.2 Patient results

For the patient CTA data, some phases are with full-dose (at 100% MTC) and some with low-dose (at 20% MTC or transition between 100% MTC to 20% MTC). Twenty phases of the test patient are shown in Figure 3.8. Phase 8-13 in this test patient should be in the 100% MTC window (full dose), while others should be in the 20% MTC window (low dose) or the transition window. For simplicity, we selected eight consecutive phases for each category (while excluding four transition phases): 1) phases 1-6, 19 and 20 for low-dose with high noise; and 2) phases 8-15 for full-dose with low noise. Phases 4- 6 show less noise than other low-dose phases as the MTC was ramped up. The black box in the aorta is used as region of interest (ROI) to calculate the standard deviation (STD) of the intensity to represent the noise level and the magnified views of ROI are shown in Figure 3.9. The noise texture can be seen more clearly, and the top row (low-dose

images) are much noisier than the bottom row (full-dose images). The STD values in HU are listed in Table 3.4, where the low-dose STD values are greater than 45 HU and the full-dose STD values are less than 40. This is also different from the thresholds for the selection of low-dose and full-dose training data (full dose < 39 HU and low dose > 59 HU). The purpose is to see how effective RecycleGAN can denoise these relatively high noise slices.
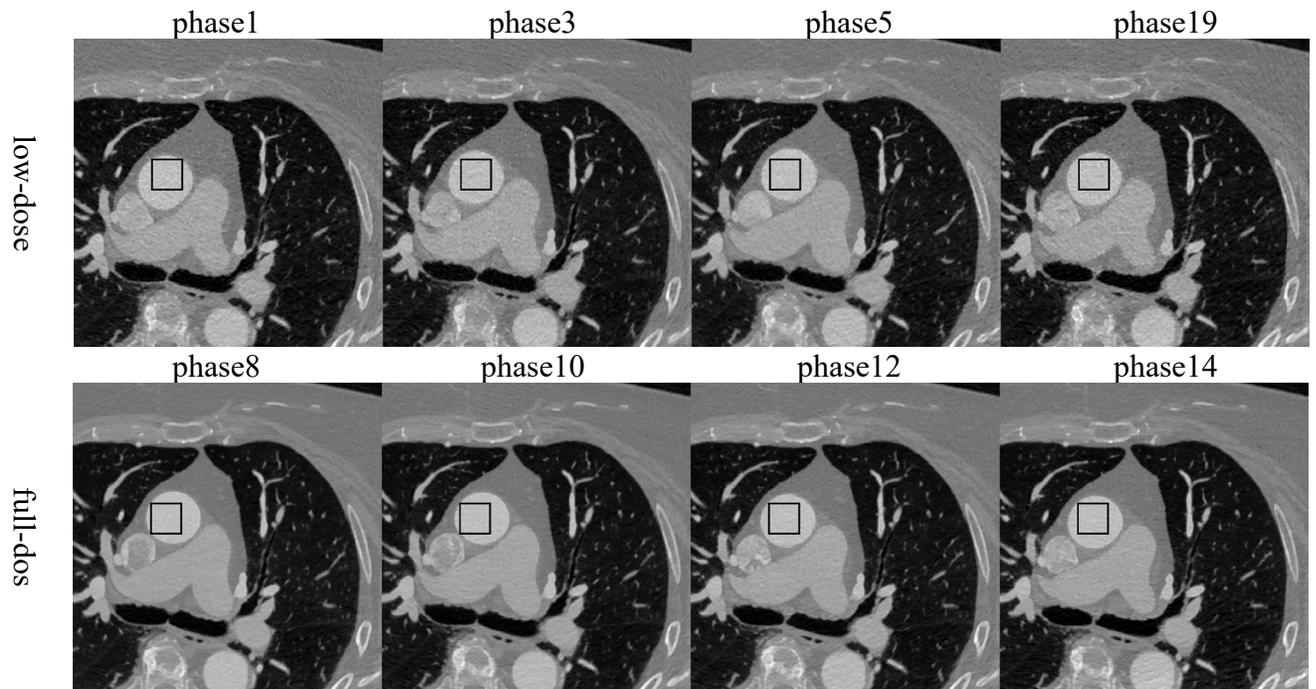


Figure 3.8 Twenty phases of the test patient CTA images (Top row: low-dose images of phase 1-6, 19, and 20; bottom row: full-dose images of phase 8-15). The black box in the aorta is used as a region of interest (ROI) to calculate the standard deviation (STD) of the intensity to represent the noise level. Display Window [-1000 950]HU.
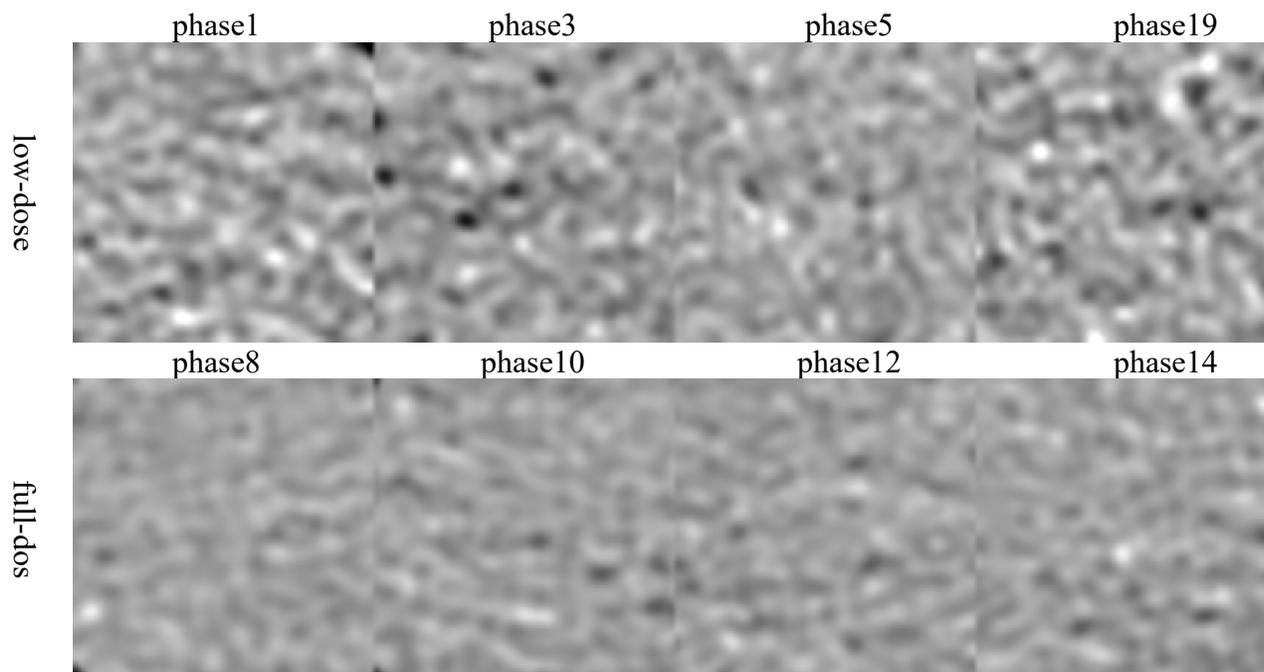
Figure 3.9 Twenty phases of the ROI (black box in Figure 3.10) (Top row: low-dose images of phase 1-6, 19, and 20; bottom row: full-dose images of phase 8-15). Display Window is [76 676].

Table 3.4 The standard deviation (STD) values in HU of low-dose and full-dose ROI

| Phase | 1 | 2 | 3 | 4 | 5 | 6 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|
| Low-dose | 60.12 | 62.35 | 58.17 | 49.34 | 45.06 | 51.09 | 65.61 | 66.56 |
| phase | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 |
| Full-dose | 31.09 | 29.01 | 36.84 | 34.17 | 35.27 | 35.33 | 38.39 | 39.59 |

In Figure 3.10, we compared the low-dose CTA images (phases 1-6, 19 and 20) of the test patient with CycleGAN and RecycleGAN denoised images. Similar to the findings in the phantom results, both CycleGAN and RecycleGAN can effectively suppress the noise, while RecycleGAN keeps the image details much better than CycleGAN. CycleGAN also suffers from some intensity artifacts as

marked by the yellow arrows in Figure 3.10, which are consistent with those reported in the previous study [76]. The ROI images are shown in Figure 3.11, CycleGAN and RecycleGAN yield less noisy looking compared to the original low-dose images. Furthermore, RecycleGAN images are least noisy and more consistent across all phases, while CycleGAN suffers from some noise bumps for phase 5 and 6. This is likely due to the recurrent loss used RecycleGAN, which takes the temporal correlation into the denoising mechanism. The quantitative measures of STD of ROI are shown in Table 3.5. CycleGAN does a good job for most phases (bringing down the noise from 50~60 HU to 30~40 HU) except for phase 6. RecycleGAN further suppresses the noise to the range of 16~26 HU.
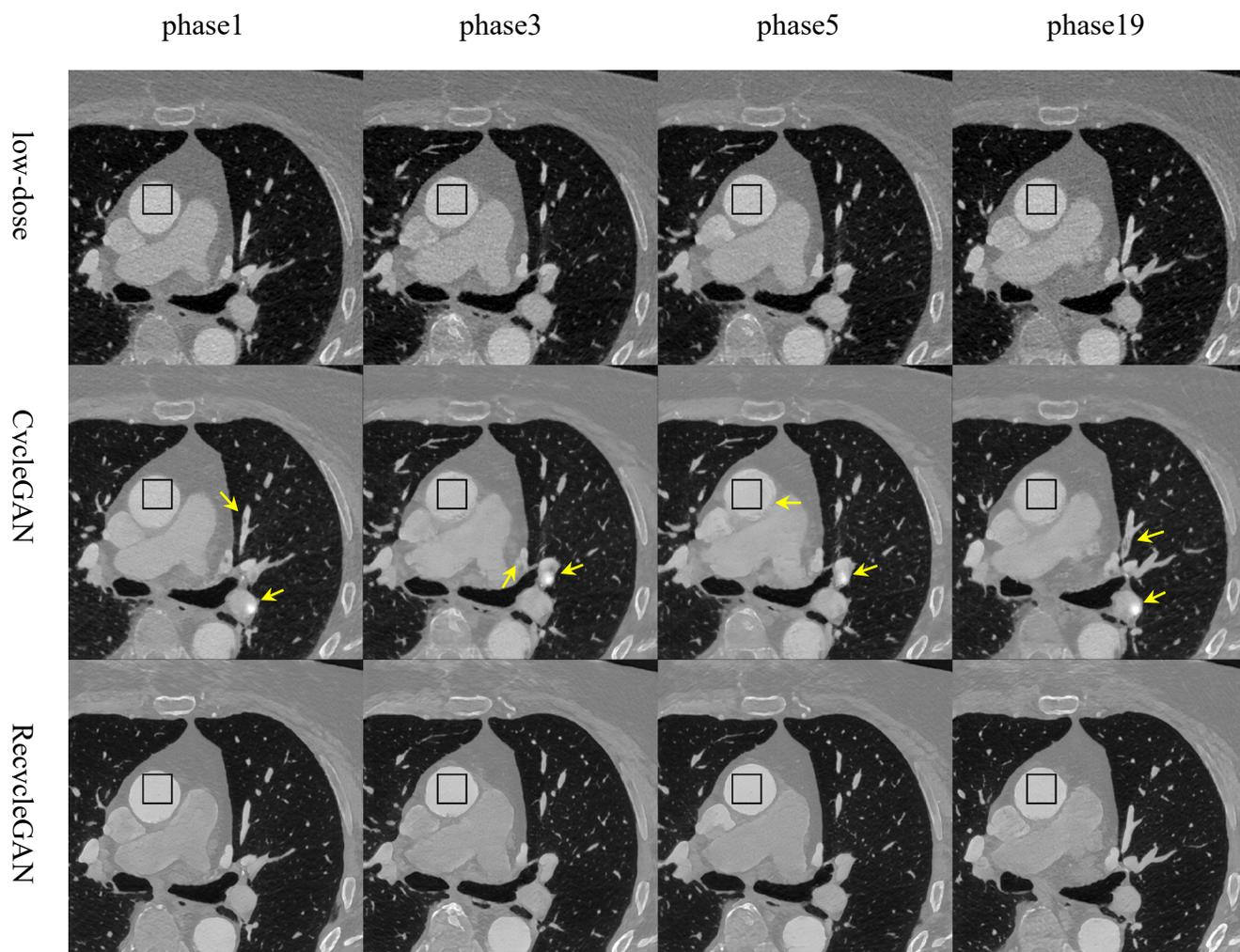
Figure 3.10 Low-dose phases of the test patient denoised by different methods. (Top row: original low-dose images; middle row: CycleGAN; bottom row: RecycleGAN) Display Window [-1000 950]HU.
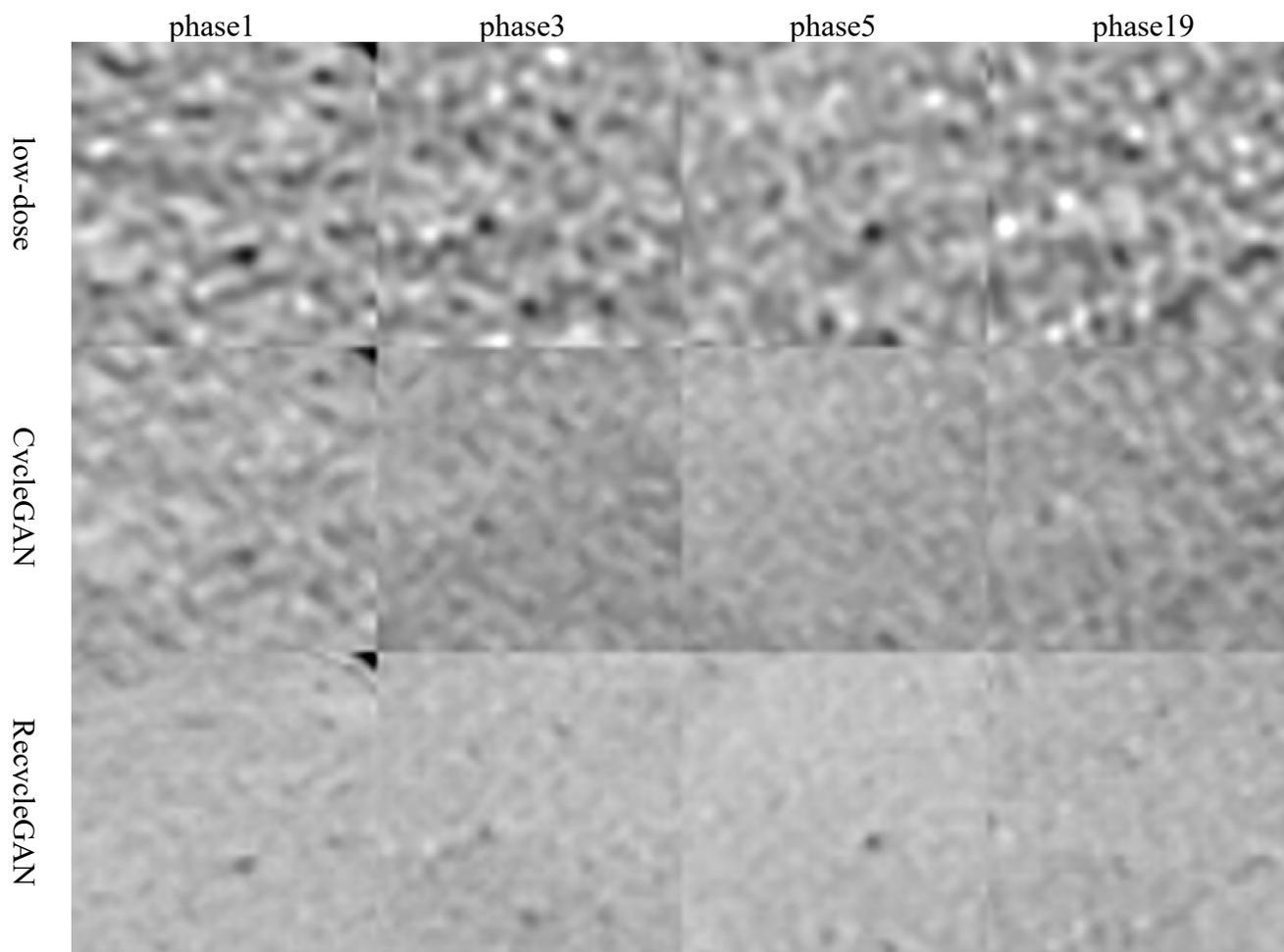
Figure 3.11 The ROI images of the test patient denoised by different methods. (Top row: original low-dose images; middle row: CycleGAN; bottom row: RecycleGAN) Display Window [76 676]HU.

Table 3.5 The standard deviation (STD) values in HU in ROI for CycleGAN and RecycleGAN

| Phase | 1 | 2 | 3 | 4 | 5 | 6 | 19 | 20 |
|---|---|---|---|---|---|---|---|---|
| Low-dose | 60.12 | 62.35 | 58.17 | 49.34 | 45.06 | 51.09 | 65.61 | 66.56 |
| CycleGAN | 40.46 | 36.48 | 34.50 | 30.43 | 38.87 | 54.59 | 53.14 | 46.46 |
| RecycleGAN | 26.01 | 26.70 | 23.75 | 19.82 | 17.16 | 23.30 | 20.48 | 25.32 |

## 3.5   Discussion

RecycleGAN is more effective than CycleGAN for denoising CT image sequence as it uses a recurrent loss to enforce the temporal consistence. In essence, it treats 2D image series as a 3D signal (2D space + 1D time) and denoises in 3D instead of 2D. This leads to more effective noise suppression and structure preservation. In the future, the whole 3D volume image plus time may be treated as a 4D signal to see if further improvement could be achieved. Right now, the training of RecycleGAN is more time consuming (37 hours for RecycleGAN Vs 18 hours for CycleGAN). The computational burden moving from 3D to 4D may be alleviated by multiple GPU parallelism.

In this work, we focus on comparing RecycleGAN and CycleGAN with extensive phantom and patient studies (with 9-fold cross-validation for the phantom study and 50 patients for the patient study). We used CycleGAN as a baseline, which was extensively compared with other state-of-the-art denoising methods [74, 75]. Although the direct comparison between RecycleGAN and other methods may be lack in this work, their relative performance can be deduced from the comparison between RecycleGAN and CycleGAN.

MP-CTA can offer more diagnostic information than SP-CTA. However, the full radiation dose is a major hurdle to adopt MP-CTA broadly for CAD diagnosis. Therefore, to lower MP-CTA dose level to comparable to SP-CTA will be clinically significant. RecycleGAN is an important development moving toward to this goal. First, RecycleGAN is a software-based method and does not require the aligned

low-dose and full-dose images. Although the hardware difference may demand further tuning of the RecycleGAN model trained on a certain type of scanner (e.g. Siemens Force in this work), as the nature of CT images is the same, a comprehensive model could be built using data from multi-scanners and multi-centers. Secondly, RecycleGAN showed superior performance on suppressing noise and preserving the structure details and contrast for CTA image sequences compared to CycleGAN. If a constant 20% MTC could be used for MP-CTA, the radiation dose could be lowered by ~55% (assuming 6 phase 100%MTC pulse window for a total of 20 phases). Although this dose level is still higher than SP-CTA, further reduction, such as sparse sampling, could be exploit. Use of advance deep learning or reconstruction methods to explore the lower bound of MP-CTA dose level without compromising the diagnostic outcomes is worth further investigation.

For the patient MP-CTA cases used in this study, an ECG-gated tube current modulation was turned on with the pulsing window between 40% and 70% of the cardiac phases. The tube current reduction outside the pulsing window was 20% of the full tube current. Therefore, this study focused on reducing noise of low-dose images acquired outside the pulsing window. One previous study has investigated CycleGAN denoising of extreme low-dose (high-noise) CT [76]. At 4% of full dose, although the baseline CycleGAN method [73] introduces some artificial features, CycleGAN denoised images still improved the signal-to-noise ratio (SNR) and the radiologist reading rates over the original LDCT images. To address the performance deterioration of CycleGAN, the wavelet-assisted noise

disentanglement (WAND) [76] was introduced to extract high-frequency sub-band images (including both noise and edge information) before CycleGAN training. Their results showed that WAND were effective to suppress high noise and avoid artifacts. In Figure 3.10, we also discovered similar artifacts in CycleGAN images reported in [76], which were successfully removed in RecycleGAN images. This demonstrated that the spatiotemporal training in RecycleGAN may be an alternative way to correct for the inconsistent translation of CycleGAN. Nevertheless, we believe that WAND can be deployed similarly to RecycleGAN, i.e. adding high-frequency sub-band image extraction before RecycleGAN training, when its denoising performance is significantly degraded due to substantially elevated noise. This will be a topic for future investigation.

In summary, we developed a spatiotemporal deep learning denoising method, RecycleGAN, for cardiac CT image sequences. Compared to the state-of-the-art spatial domain denoising method, CycleGAN, RecycleGAN utilizes the temporal relationship of several consecutive phases through a recurrent loss to further improve the denoising performance. Note that RecycleGAN still enjoys the advantage of CycleGAN without need of aligned low-noise and high-noise images. Both phantom and patient studies show that RecycleGAN outperform CycleGAN in quantitative metrics and image quality for CT image sequences. It is envisioned that RecycleGAN could be used to significantly lower the MP CTA dose by effectively removing the image noise. More clinically relevant evaluations will be conducted in the future work.

# CHAPTER 4.    TEXTURE TRANSFORMER SUPER-RESOLUTION FOR LOW-DOSE COMPUTED TOMOGRAPHY

Shiwei Zhou[1] , Lifeng Yu[2] and Mingwu Jin[1]

[1] Department of Physics, University of Texas at Arlington, TX 76019, United States of America

[2] Department of Radiology, Mayo Clinic, Rochester, MN 55905, United States of America

## 4.1 Introduction

Computed tomography (CT) is a common technique in modern medicine with millions of exams each year. However, due to the technical limitations of clinical scanners, CT images with typical resolution of millimeter or submillimeter are hard to resolve structures on an order of tens of microns for certain physiological and pathological applications [84], e.g., coronary artery analysis [85]. High-resolution CT (HRCT) can be done through hardware innovation, such as using smaller detector elements and pitches. Not only is this costly, but also the elevated quantum noise will become an issue if the incident X-ray intensity does not increase accordingly. Furthermore, low-dose CT has been actively pursued recently to lower the ionizing radiation to patients. A relatively smoother kernel is typically used in low-dose CT to suppress image noise, which may sacrifice spatial resolution. An alternative approach is to apply noise reduction methods to control image noise in low-dose CT. Many advanced algorithms were developed to alleviate the noise either in the projection domain, or in the image domain, or both. However, these algorithms may still suffer from degradation of spatial resolution, especially at low contrast level. It is desirable to develop algorithms that can simultaneously suppress image noise while enhancing spatial resolution (Super Resolution or SR).

There are three categories of computational methods proposed to suppress noise and enhance spatial resolution for LRCT images: 1) model based iterative reconstruction methods [86, 87]; 2) sparse representation methods [88-90]; and 3) deep learning methods[74, 91-93] . In [93], a generative adversarial network (GAN)

[79] with a perceptual loss function combining both an adversarial loss and a content loss, called "SRGAN", has been proposed to improve natural image resolution, which was used as a comparative method for super-resolution CT (SRCT) in [74]. Recently, GAN-CIRCLE [74], a GAN using residual and cycle-consistent learning [94], was proposed to produce SRCT images. Extensive experiments have been conducted to show the SRCT performance of GAN-CIRCLE superior or comparable to other state-of-the-art SR methods, including SRGAN.

In this work, we propose a texture transformer network to simultaneously reduce image noise and improve spatial resolution in CT images. This network, referred to as Texture Transformer for Super Resolution (TTSR) [95], is a reference-based deep-learning image super-resolution method, which is another GAN-based deep learning method, to achieve SRCT from noisy LRCT. To our best knowledge, this is the first time to apply a transformer for SRCT, particularly for low-dose CT. The performance of TTSR outperforms the state-of-the-art methods, such as SRGAN and GAN-CIRCLE, for both simulated XCAT phantom data and Mayo low-dose CT dataset. We also show that TTSR can be used to allow the large detector size and fewer detectors for high-resolution CT scan and to save computation time.

## 4.2  Methods

### 4.2.1  SRGAN

SRGAN [93] adapts the basic generator and discriminator structure of GAN to build the generator that maps LRCT to HRCT. The generator employs a residual network with skip-connection to force a perceptual loss of high-level feature maps from a VGG network in addition to the adversarial loss. SRGAN is a supervised method requiring the paired LRCT and HRCT images for training. It has shown a decent performance for SRCT [74] and more details can be found in [74, 93].

### 4.2.2 GAN-CIRCLE

The basic structure of GAN-CIRCLE [74] is composed of two GANs: one is to learn the forward mapping from LRCT to HRCT (generator G: x → y and discriminator Dy) and the other is to learn the backward mapping from HRCT to LRCT (generator F: y → x and discriminator Dx). The GAN-CIRCLE principles are shown in Figure 4.1. In order to learn two mappings without paired LR and HR CT images, the cycle consistency loss is used to enforce F(G(x)) = x and G(F(y)) = y, as indicated by the dashed arrows in Figure 4.1. Although CAN-CIRCLE still needs the general labels of LRCT and HRCT images for training, the one-to-one correspondence between them is not required. More details of GAN-CIRCLE can be found in [74, 75].
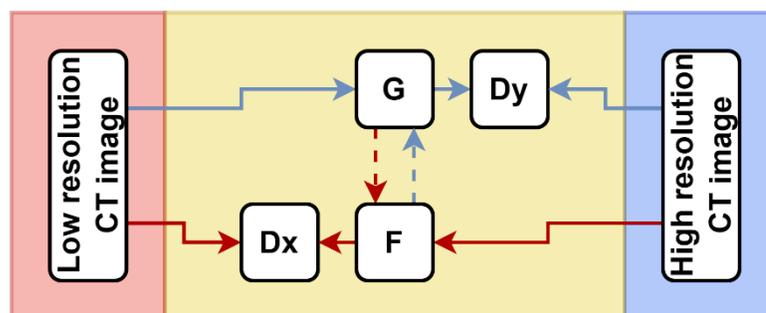


Figure 4.1 The principles of GAN-CIRCLE. G: generator from LRCT (x) to HRCT (y); F: generator from HRCT (y) to LRCT (x); Dy: discriminator for real or fake

HRCT images; Dx discriminator for real or fake LRCT images. The dashed arrows enforce the cycle-consistency loss.

### 4.2.3 TTSR with attention mechanism

The TTSR [95] is a reference-based image super-resolution method through a GAN mechanism (the blue branch of the forward mapping from LRCT to HRCT in Figure 4.1). A complex transformer network is used for the generator to translate the LRCT images to the HRCT images. The LRCT images and HRCT images are severed as queries and keys in a transformer, respectively. The image translation is optimized through deep neural network (DNN) texture extraction, relevance embedding, and attention-based texture transfer and synthesis to enable joint feature learning between LRCT and HRCT images. The generator of TTSR is shown in Figure 4.2. Q, K and V are the texture features extracted from an up-sampled LR image, a sequentially down/up-sampled reference image, and an original reference image, respectively. F is the LR extracted features and is further fused with the transferred texture features T to generate the SR output. ↑ represents upsampling. ↓ represents downsampling. The discriminator is used to distinguish HRCT and SRCT. The detailed network structures of the generator and discriminator can be found in Appendix. The parameters of TTSR network are optimized using the loss function that is composed of three parts: 1) $L_1$ difference between the HRCT and HRCT images; 2) the GAN loss; and 3) the perceptual loss of feature maps. More details of TTSR can be found in [95].
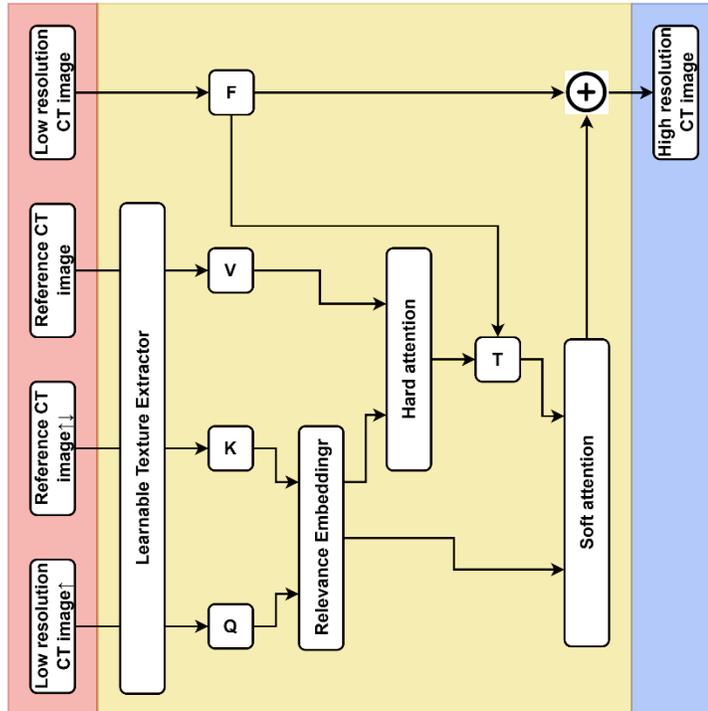
Figure 4.2 The generator of the texture transformer super resolution (TTSR) network. Q: texture features extracted from an up-sampled LRCT image; K: texture features extracted from a sequentially down/up-sampled reference HRCT image; V: texture features from an original HRCT reference image; F: features from a LRCT image; T: transferred texture features; ↑: upsampling; ↓: downsampling.

## 4.3 Experiment Setup

To evaluate the super-resolution performance of TTSR, we performed both a phantom study by using the XCAT phantom, and a real patient data study by using a real clinical dataset from the 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge [96]. Cubic spline interpolation, SRGAN [93], and GAN-CIRCLE [74] were used for comparison.

### 4.3.1 Evaluation metrics

We use two metrics to quantitatively evaluate different methods: peak signal-to-noise ratio (PSNR) [81] and structural similarity index measure (SSIM) [82, 83]. PSNR and SSIM are defined below,

$$PSNR = -10 \cdot \log_{10}\left(\frac{MSE}{MAX_y^2}\right) \qquad (4.1)$$

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

where x denotes the SRCT image, y is the corresponding HRCT image, $MSE$ is mean squared error between x and y, $MAX_y$ denotes the maximum intensity value in the HRCT image y, $\mu_x$ and $\mu_y$ are the averages of the SRCT image $x$ and the HRCT image $y$, $\sigma_x^2$, $\sigma_y^2$ and $\sigma_{xy}$ are the corresponding variance and covariance, and $c_1$ and $c_2$ are two constants to stabilize the division operation. Additionally, we calculate Feature Similarity Index (FSIM) [97], which is a structure-based image quality assessment using the phase congruency (PC) and the image gradient magnitude (GM). The local quality maps, i.e. $PC_{SRCT}$, $PC_{HRCT}$ and $GM_{SRCT}$, $GM_{HRCT}$, are used to calculated the similarity measures $S_{PC}$ and $S_{GM}$. The combined similarity is defined as $S_L = S_{PC} \cdot S_{GM}$, where • is pixel-wise multiplication. Finally, the FSIM index is defined as follows:

$$FSIM = \frac{\sum_\Omega S_L \cdot \max(PC_{SRCT}, PC_{HRCT})}{\sum_\Omega \max(PC_{SRCT}, PC_{HRCT})} \qquad (4.2)$$

where $\Sigma_{\Omega}$ represents the sum over the image spatial domain $\Omega$.

The PSNR measures the super-resolution performance by calculating overall intensity difference between the SRCT image (translated from the LRCT image) and the HRCT image. The SSIM measures the perceptual similarity between the SRCT and HRCT images while pooling all locations to be the same importance. The FSIM is devised to measure the low-level feature sets between SRCT and HRCT images. The higher PSNR, SSIM, and FSIM values, the better the SR performance.

### 4.3.2 Phantom data experiments

The 4D XCAT phantom program [80] based on 18 patients' CT data was used to produce the projection data through the ray tracing algorithm [64, 98]. The photon noise at $1\times10^5$ and $2\times10^4$ per ray and the electron noise of 10 were added respectively for full dose and low dose projection [99]. The benchmark HRCT images ($1\times10^5$ photon counts/ray) were reconstructed as 512x512 slices with 1 mm x 1 mm pixel size. For LRCT images, the photon counts were reduced to 20% of HRCT, i.e. lower radiation dose ($2\times10^4$ photon counts/ray), and 128x128 slices with 4 mm x 4 mm pixel size were reconstructed. 17 patients' data were used as training data, while the remaining patient' data were used as test data. The TTSR model was trained on CUFED dataset (i.e. only nature images without CT images). The pre-trained model was further tuned with 10 epochs using the simulated CT images with learning rate of $1\times10^{-4}$. The reference image to obtain the SRCT image for the test LRCT image can be: 1) a HRCT image totally different from the LRCT

image ("random"), e.g. the different organ areas from the different patients (in the training set); 2) an HRCT image similar to the LRCT image ("partially aligned"), e.g. the similar organ area from different patients (in the training set); and 3) an HRCT image identical to the scene of the LRCT image ("fully aligned"), e.g. the same organ area from the same patient (in the test set), as shown in Figure 4.3 (green lines for random, red lines for partially aligned, and blue lines for fully aligned). Note that the first two types of reference images are not from the test set and represent a realistic achievable performance and the third one is trivial (as the corresponding HRCT image exists), but provides the best benchmark. For comparison, the cubic spline interpolation, SRGAN, and GAN-CIRCLE were used to obtain SRCT images from LRCT images as well. SRGAN was tuned to the learning rate of $1x10^{-4}$, and GAN-CIRCLE was tuned using lambda of 10 for the consistency loss and learning rate of $1x10^{-4}$.
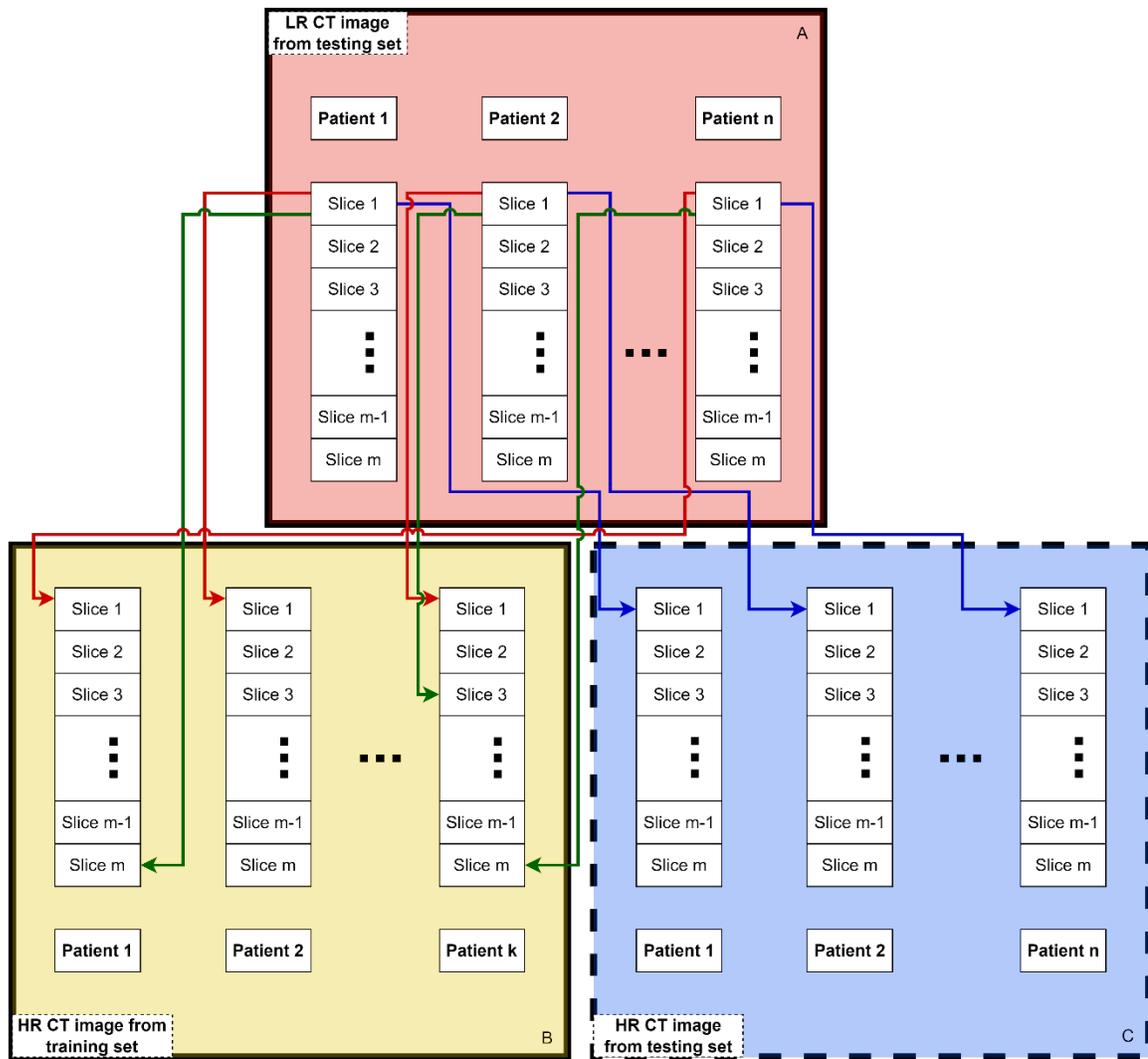
Figure 4.3 Three types of reference images of TTSR to obtain SRCT images for the test LRCT images: a) Green lines, random; b) Red lines, partially aligned; c) Blue lines, fully aligned, trivial in a real application. k: the number of patients in the training set; n: the number of patients in the test set (n=1 in this study).

### 4.3.3 Patient data experiments

The 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge dataset contains 10 anonymous patients' full-dose and corresponding low-dose (1/4 of the full dose) CT projection data. First, we used the conjugate gradient least squares (CGLS) iterative method [100] to reconstruct LRCT and HRCT images. HRCT images were reconstructed from the full-dose projection data (736x64 detector matrix, detector

size 1.2858 mm x 1.0947 mm per pixel, helical pitch 0.6) with the image size of 512x512. To obtain LRCT images, the low-dose projection data were binned from 736x64 to 184x64, i.e. 4x binning in the in-plane direction. Consequently, LRCT images were reconstructed with an image size of 128x128. As the low-dose projection data used only 1/4 of the full dose, the 4x binning was roughly achieved the similar noise level as the full dose. For GAN-CIRCLE and TTSR methods, 9 patients' data were used as training data, while the remaining patient' data were used as test data. To train TTSR for the patient data, we found that the pre-trained model using CUFED dataset ("pre-trained") did not provide any faster convergence for good SRCT performance (see "Discussion") in contrast to the phantom data. Thus, we randomized the network parameters of TTSR and trained it for the patient data from scratch, where a learning rate of $8\times10^{-5}$ and 150 epochs were used. For SRGAN, a learning rate of $5\times10^{-4}$ and 200 epochs were used. For GAN-CICLE, a learning rate of $2\times10^{-4}$ and 100 epochs were used. For the reference image of the test LRCT image of TTSR, we used the similar slice in the training HRCT images to represent a realistic performance. The evaluation of SR performance of different methods in this patient data study is shown in Figure 4.4.
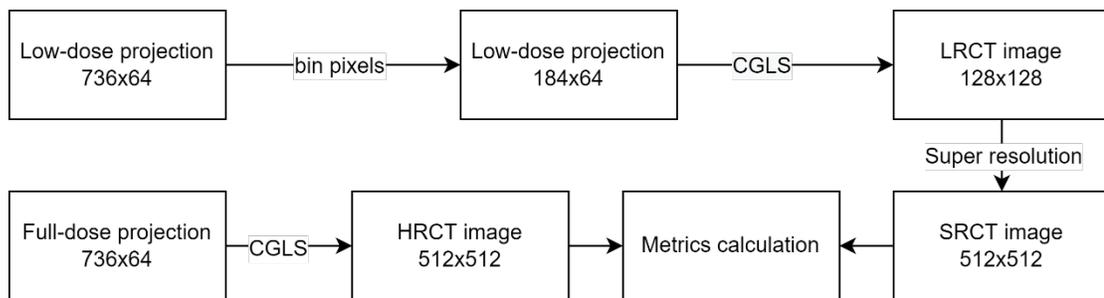


Figure 4.4 Evaluation of SRCT for the real patient data study. (CGLS: conjugate gradient least squares)

Finally, we conducted a study to compare the aforementioned TTSR method (that translated LRCT to HRCT) to denoising of high-resolution low-dose CT using block-matching and 3D filtering (BM3D) [101] and GAN-CIRCLE [74]. We use GAN-CIRCLE instead of SRGAN because GAN-CIRCLE has been investigated for low-dose CT denoising [75], while SRGAN has not. For denoising methods, the high-resolution 512x512 low-dose CT images were first reconstructed from low-dose high-resolution projections (736x64 projection matrix). Then, SRCT images from TTSR were compared with these high-resolution low-dose CT images denoised by BM3D GAN-CIRCLE. Note that both methods used the same radiation dose. However, TTSR has two advantages over high-resolution low-dose CT denoising: 1) the hardware requirement is less demanding, as 184 (for TTSR) vs 736 (for high-resolution low-dose CT denoising) in-slice detector elements; and 2) the reconstruction time of 128x128 images is much less than that of 512x512 images. Once TTSR and GAN-CIRCLE were trained, the times for applying the models were in the same order (see Table 4.4).

## 4.4 Results

### 4.4.1 Phantom data results

The SRCT images from different methods are shown in Figure 4.5 along with the original HRCT image (Figure 4.5a). Note that the original LRCT images (128x128) are too small to show here. The cubic spline interpolation method is one of the simplest SR methods. However, as shown in Figure 4.5b, the interpolated SR image is blurry and suffers some noise. SRGAN (Figure 4.5c) greatly improves the

clarity of fine structures. GAN-CIRCLE (Figure 4.5d) yields a SR image with better defined edges and suppressed noise than the interpolation method, but seems to still oversmooth the image. TTSR (Figure 4.5e-g) shows more details than the other two methods although some artifacts seem to be present in SR images (e.g. in the heart). The difference among TTSR using different reference images is small. The magnified view of the red box in Figure 4.5a is shown in Figure 4.6. Both interpolation and GAN-CIRCLE are unable to recover the lung nodules clearly and cause blurred edges, while SRGAN and TTSR provides much better resolution recovery. Although SRGAN maintains the sharp edges well, it produces piece-wise smoothness and is not able to keep the small structures inside the lung as well as TTSR. Full-aligned TTSR preserves the nodule shape better than the other TTSR methods. The quantitative results averaged over 128 slices of the test patient are listed in Table 4.1. TTSR achieves the best PSNR and SRGAN yields the best SSIM. Both TTSR and SRGAN ties for the best FSIM. As SRGAN and TTSR use paired training data, they seem to achieve better SRCT performance than GAN-CIRCLE, which is an un-paired training. Some artifacts are noticeable in TTSR images inside the heart (Figure 4.5), which may have caused inferior SSIM of TTSR to that of SRGAN and GAN-CIRCLE. Nevertheless, these results suggest that previously acquired HRCT images can be used as reference images to improve newly acquired LRCT images with much reduced radiation dose without the requirement of alignments (for TTSR-random and TTSR-partially aligned).
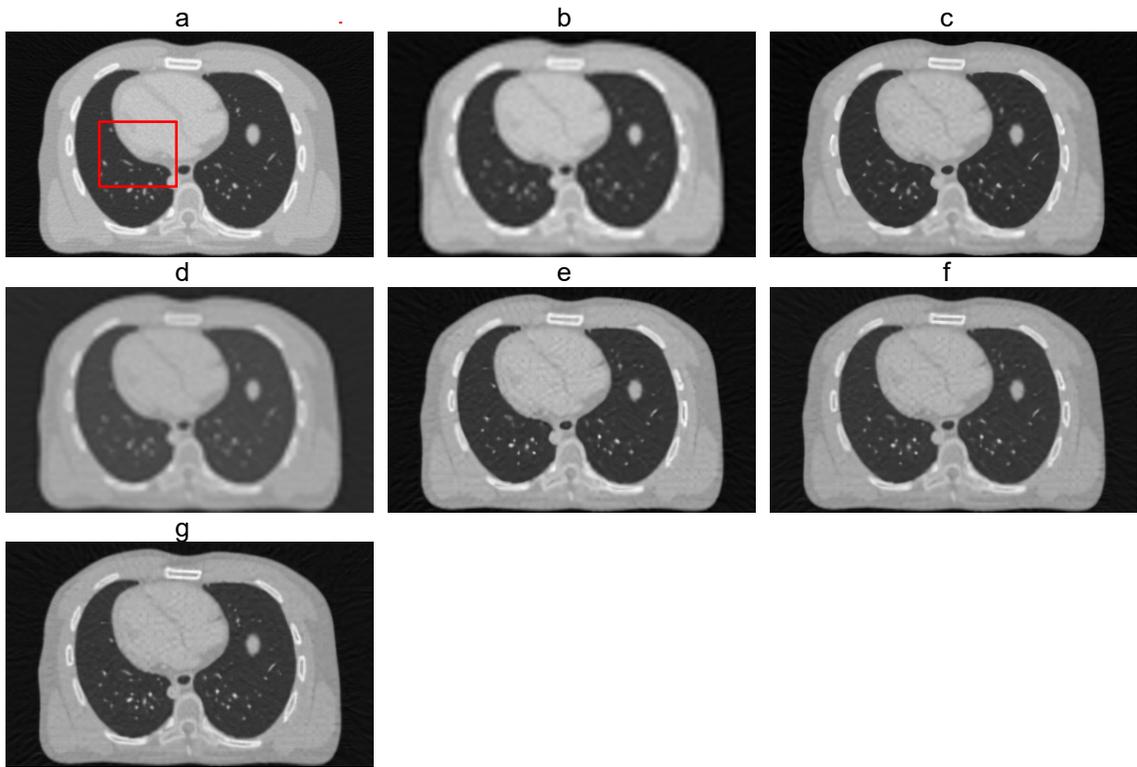
Figure 4.5 HRCT image and SRCT images from different methods (a: HRCT; b: cubic spline interpolation; c: SRGAN; d: GAN-CIRCLE; e: TTSR random; f: TTSR partially aligned; g: TTSR fully aligned). (Display window: [-1345 782]HU)

Table 4.1 Quantitative results for different methods for the phantom data

| Methods | PSNR (dB) | SSIM | FSIM |
|---|---|---|---|
| Cubic spline | 31.57±0.89 | 0.74±0.02 | 0.87±0.01 |
| SRGAN | 35.84±0.41 | **0.89±0.02** | **0.95±0.01** |
| GAN-CIRCLE | 34.85±0.32 | 0.88±0.02 | 0.94±0.01 |
| TTSR/random | 37.23±0.42 | 0.84±0.02 | 0.93±0.01 |
| TTSR/partially aligned | 37.23±0.42 | 0.84±0.02 | 0.93±0.01 |
| TTSR/fully aligned | **37.99±0.49** | 0.86±0.02 | **0.95±0.01** |
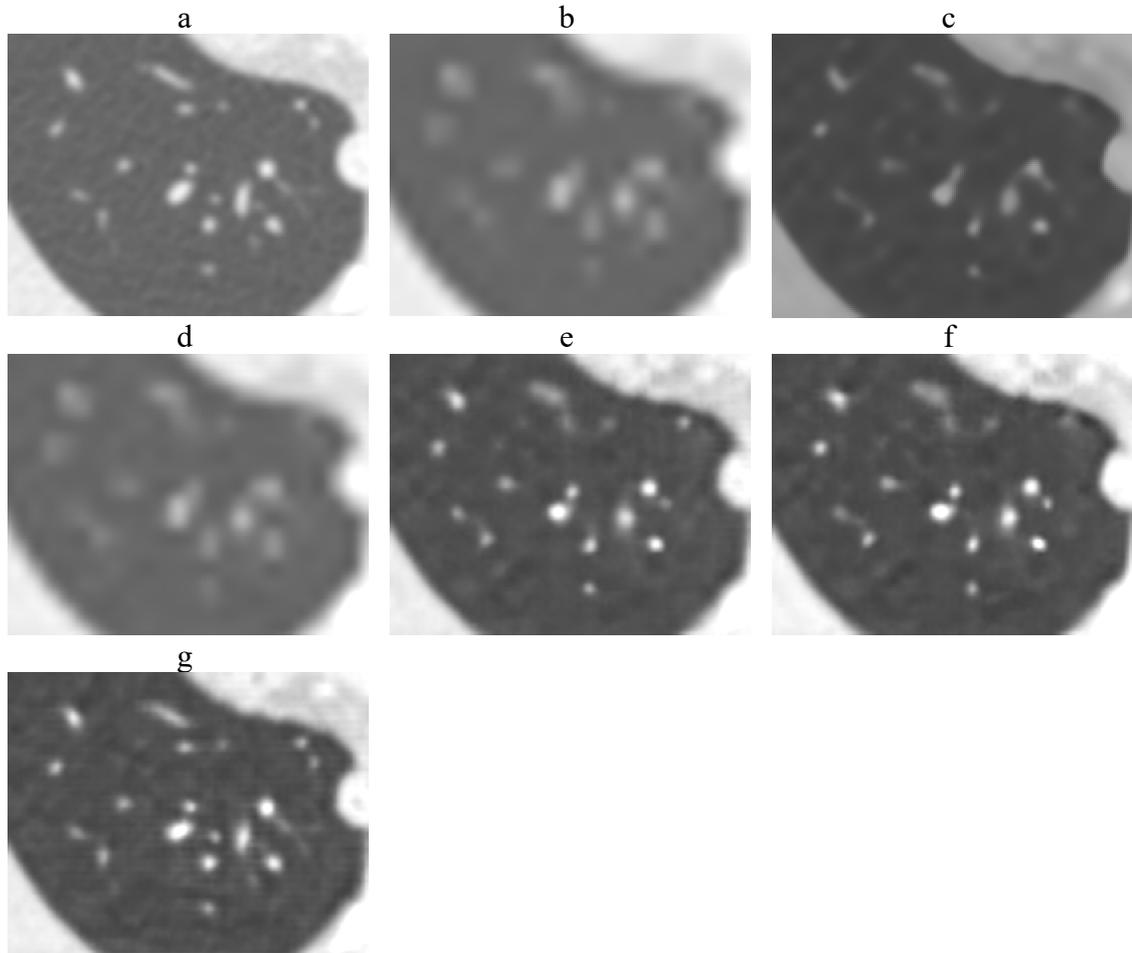
Figure 4.6 HRCT image and SRCT images from different methods for the red box area in Figure 4.5a. (a: HRCT; b: cubic spline interpolation; c: SRGAN; d: GAN-CIRCLE; e: TTSR random; f: TTSR partially aligned; g: TTSR fully aligned). (Display window: [-1345 150]HU)

### 4.4.2 Patient data results

The images from different methods for real patient data are shown in Figure 4.7 (for the thighs) and Figure 4.8 (for the thorax). Again, the original LRCT images (128x128) are too small to show here. The interpolation (Figure 4.7b, Figure 4.8b) and GAN-CIRCLE (Figure 4.7d, Figure 4.8d) SR images are kind of blurry, although GAN-CIRCLE performs slightly better. SRGAN (Figure 4.7c, Figure 4.8c) suffers less blurry, but the SR images seem to show different patterns from the

HRCT images, e.g. in the thighs (Figure 4.7c) and in the lungs (Figure 4.8c). Note that the TTSR method uses the similar HRCT slice in the training set as the reference image, corresponding to TTSR-partially aligned in the simulation study. The TTSR images (Figure 4.7e, Figure 4.8e) successfully recover image resolution without notable artifacts and yield most similar images to the original HRCT images (Figure 4.7a, Figure 4.8a). The quantitative results in Table 4.2also show the superior SR performance of the TTSR model, followed by SRGAN, over other methods. As can be seen in the zoomed region of the part of the lung and heart (red box in Figure 4.8a) in Figure 4.9, GAN-CIRCLE (d) has less noise than the interpolation method (b), but is still lack of details. SRGAN generates a sharp SRCT image (c) at the cost of eliminating the fine structures in the lung and the texture in the heart. The TTSR method shows better structure details, e.g. red arrows, and better preserved edge, e.g. blue arrows. The TTSR image (e) resembles the HRCT image (a) most although the texture inside the heart in the TTSR image (with less streak-artifacts) seems to be different from and more visually appealing than the original HRCT. It suggests that TTSR may not benefit from the pre-trained network with totally unrelated data for better CT SR performance [102]. The quantitative results for the region shown in Figure 4.9 are listed in Table 4.3. The TTSR model beats all other methods with a large margin in all three metrics. The performance of GAN-CIRCLE becomes comparable to SRGAN as SRGAN suffers loss of a lot of details in this region full of fine structures and textures.
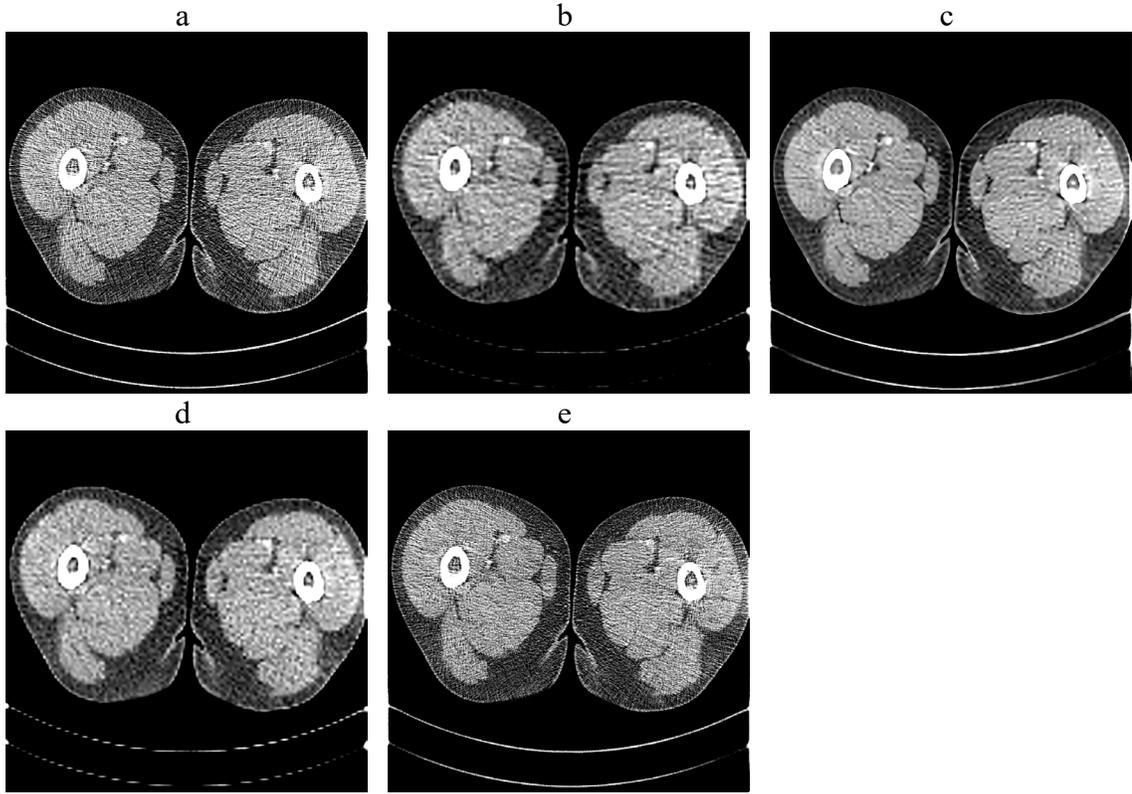
Figure 4.7 HRCT image and SRCT images of the legs from different methods (a: HRCT; b: cubic spline interpolation; c: SRGAN; d: GAN-CIRLCE; e: TTSR). (Display window: [-160 240]HU)
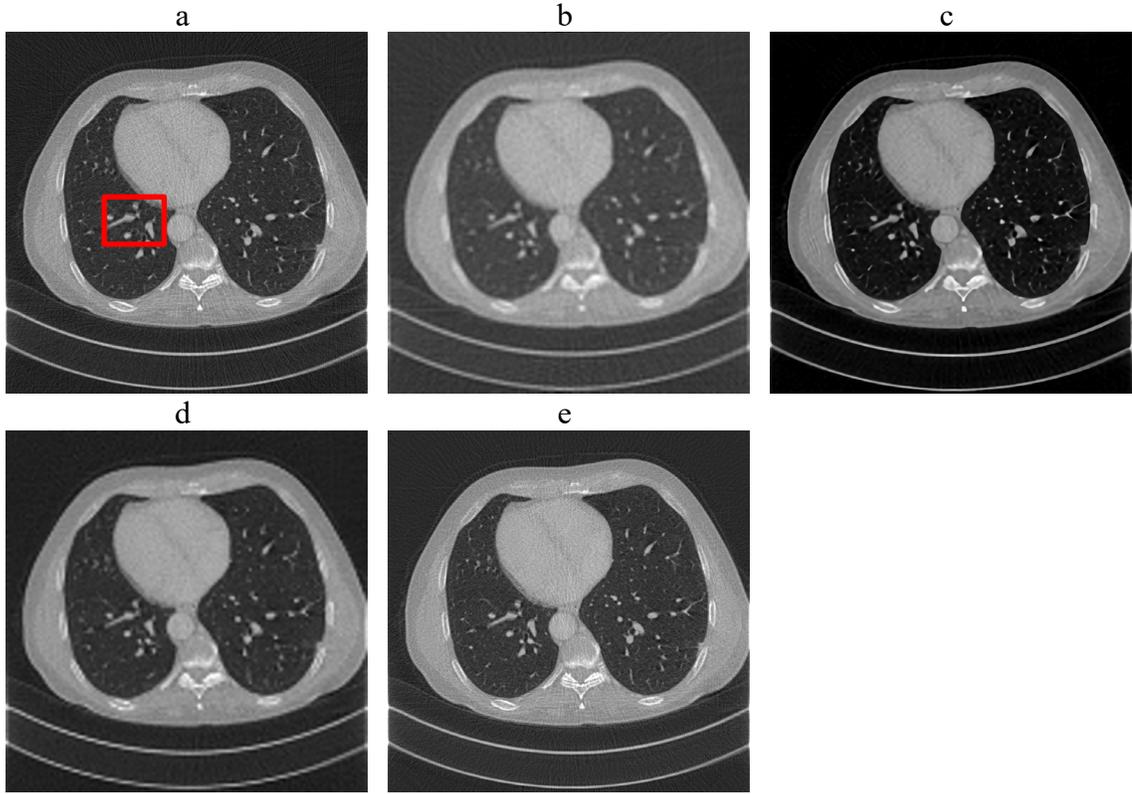
Figure 4.8 HRCT image and SRCT images of the chest from different methods
(a: HRCT; b: cubic spline interpolation; c: SRGAN; d: GAN-CIRLCE e: TTSR).
(Display window: [-1556, 1043]HU)

Table 4.2 Quantitative results for different methods for the patient data

| Methods | PSNR (dB) | SSIM | FSIM |
|---|---|---|---|
| Cubic spline SR | 27.25±0.49 | 0.53±0.04 | 0.89±0.01 |
| SRGAN SR | 29.59±0.70 | 0.61±0.04 | 0.94±0.01 |
| GAN-CIRCLE SR | 27.75±0.41 | 0.54±0.05 | 0.89±0.01 |
| TTSR | 31.16±1.38 | 0.73±0.06 | 0.97±0.01 |
| BM3D denoising | 28.18±0.52 | 0.54±0.04 | 0.88±0.01 |
| GAN-CIRCLE denoising | 29.39±0.51 | 0.58±0.03 | 0.91±0.01 |
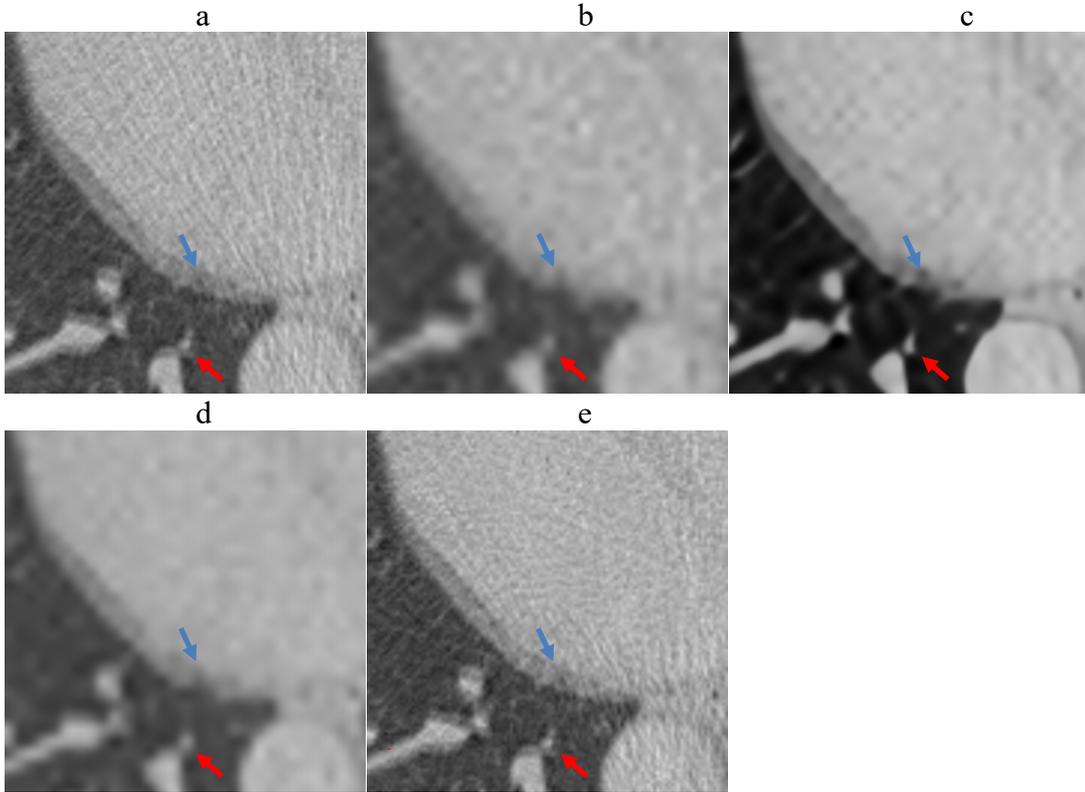
Figure 4.9 Zoomed-in HRCT image and SRCT images of the chest from different methods for the red box area in Figure 4.8a (a: HRCT; b: cubic spline interpolation; c: SRGAN; d: GAN-CIRLCE e: TTSR). (Display window: [-1556, 1043]HU)

Table 4.3 Quantitative results for different methods for the patient data in the zoomed-in range

| Methods | PSNR (dB) | SSIM | FSIM |
|---|---|---|---|
| Cubic spline SR | 27.12 | 0.51 | 0.79 |
| SRGAN | 28.77 | 0.59 | 0.82 |
| GAN-CIRCLE denoising | 28.03 | 0.57 | 0.82 |
| TTSR | **31.39** | **0.81** | **0.98** |

To compare the denoised high-resolution low-dose CT images with TTSR, the corresponding images are shown in Figure 4.10 and Figure 4.11. The BM3D

85

method largely suppresses the noise, but leads to over-smoothing as shown particularly in the zoomed-in view of Figure 4.11b. Although GAN-CIRCLE successfully addresses the blurring issue, the apparency of denoised image is different from the original HRCT, e.g. the contrast and the edge of the heart as indicated by the red arrow in Figure 4.11. Again, TTSR SRCT images are closest to the original image compared to the two denoising methods. The quantitative results for BM3D and GAN-CIRCLE denoising of low-dose CT were listed in the bottom two rows of Table 4.3. Both denoising methods of high-resolution low-dose CT images are outperformed by TTSR, although GAN-CIRCLE denoising yields better PSNR, SSIM and FSIM values than GAN-CIRCLE SR.
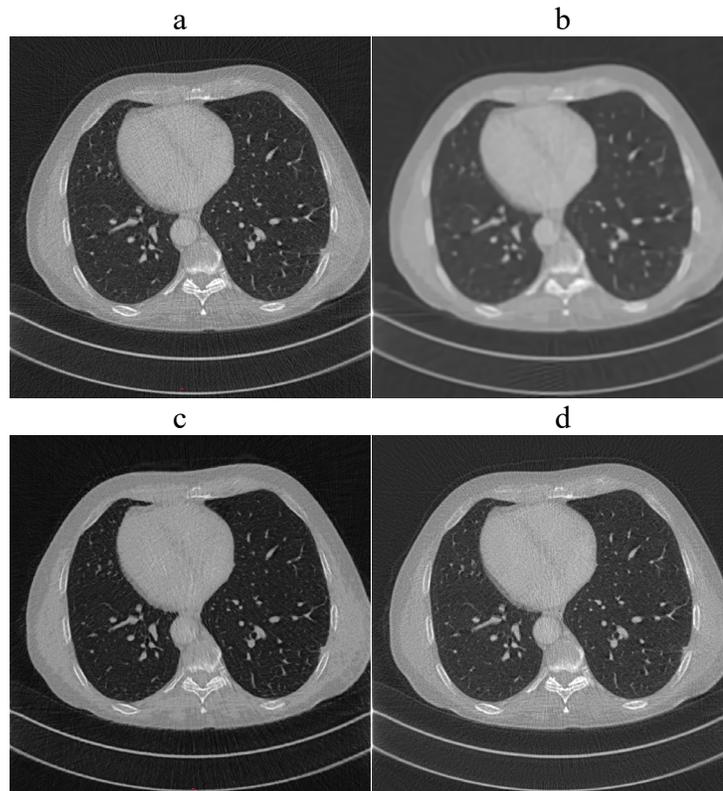


Figure 4.10 Comparison of denoised high-resolution low-dose CT and TTSR SRCT images: a) HRCT (512x512 full dose); b) BM3D denoised low-dose CT; c)

Figure 4.11 Comparison of denoised high-resolution low-dose CT and TTSR SRCT images - zoomed-in views for the red box area in Figure 4.8a: a) HRCT (512x512 full dose); b) BM3D denoised low-dose CT; c) GAN-CIRCLE denoised low-dose CT; d) TTSR SRCT. (Display window: [-1556, 1043]HU)

### 4.4.3 Computational efficiency

To compare the computational efficiency of different methods, we listed the computation time for CGLS reconstruction of a 3D image for one set of patient data (471 slices), denoising, SRGAN, and TTSR in Table 4.4. As can be seen from the Table 4.4, TTSR combined with low-resolution reconstruction (CGLS of 128x128 with 4x projection binning) is much less time consuming than the high-resolution

low-dose CT denoising (CGLS of 512x512 + either BM3D , GAN-CIRCLE, or SRGAN), yet provides much better image quality as shown in previous results.

Table 4.4 Time consuming for generating testing patient images (471 slices)

| Method | Time(second) |
| --- | --- |
| CGLS of 128x128 with 4x projection binning | 163 |
| CGLS of 512x512 | 1396 |
| BM3D | 3499 |
| SRGAN | 70 |
| GAN-CIRCLE | 29 |
| TTSR | 80 |

## 4.5   Discussion

The training for TTSR is substantially slower than SRGAN and GAN-CIRCLE. It takes six hours for 10 epochs of TTSR training on an NVIDIA A6000 GPU card, while 16 minutes for SRGAN and two hours for GAN-CIRCLE for the same number of epochs. The high computational efficiency of SRGAN is due to its straightforward generator and discriminator structure. The complexity of TTSR is also much higher than that of SRGAN and GAN-CIRCLE. There are more than seven million network parameters for TTSR, while 1.6 million for SRGAN and less than 200 thousand for GAN-CIRCLE.

In this work, we mainly tuned the learning rate and the number of epochs for different models empirically for the best PSNR using the test data, while keeping other hyperparameters the same as previous publications. This is mainly due to the heavy computational cost for a systematic investigation of the optimal hyperparameter set and the main purpose of this work is to demonstrate the

feasibility of TTSR for SRCT. For the phantom data, we found that the TTSR model pre-trained on the natural images could facilitate the training of SRCT (10 epochs). However, this is not the case for real patient data. The PSNR of the test patient data using TTSR trained from the pre-trained model and from scratch (learning rate = $8\text{x}10^{-5}$) are plotted in Figure 4.12. As can be seen, the pre-trained model does not provide a faster convergence than the from-scratch model, but seems to be less stable at the later epochs. This is likely due to that the difference between real patient data and natural images is bigger than that between phantom data and natural images. We also compared the quantitative metrics and images for two TTSR models. The from-scratch model is slightly better than the pre-trained model.

We also decided not to use a separated validation set as the datasets, particularly for the patient data, are relatively small. Although the performance may be over-estimated due to the involvement of the test data in hyperparameter tuning, the relative ranking of the different deep learning methods shall not change. Furthermore, due to the computational hurdle and the limited number of patients, the SR image translation was limited in 2D in this work. Although this may not be an issue for helical CT where 2D slices are reconstructed, the translation in 3D space could further utilize the features in the axial direction and lead to better performance.
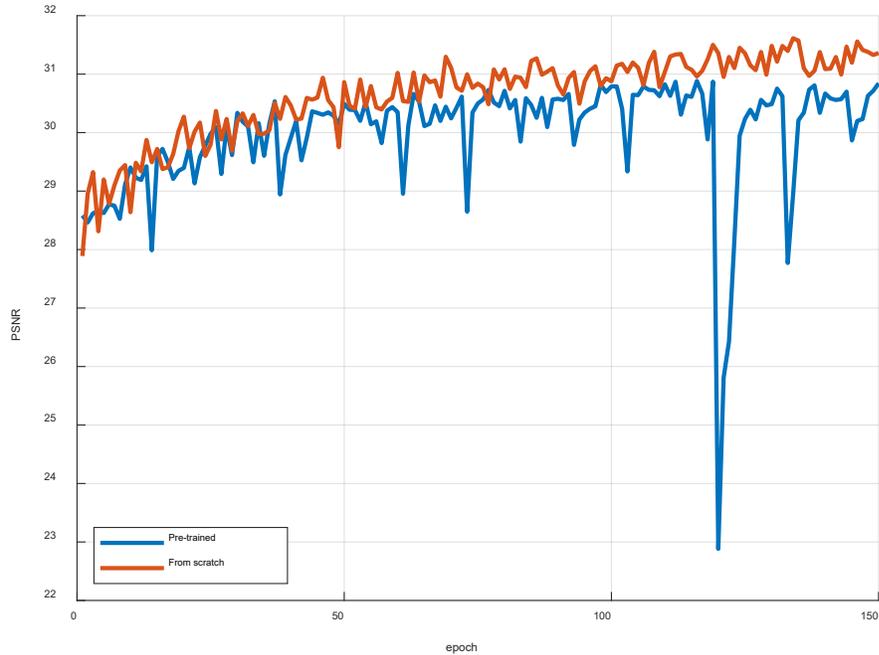
Figure 4.12 PSNR of the test patient set for pre-trained and from-scratch TTSR. (Learning rate = $8 \times 10^{-5}$)

To evaluate the SR performance, we used the engineering metrics, such as PSNR, SSIM and FSIM, as a preliminary demonstration of the effectiveness of TTSR. TTSR outperforms both SRGAN and GAN-CIRCLE for the patient data, especially for the regions with fine structures and textures (Table 4.3). The results in Table 4.2 should be used in caution as large blank regions outside the body may favor SRGAN, which tends to produce over-cleaned images. It is also worth mentioning that GAN-CIRCLE used in this work is corresponding to un-supervised GAN-CIRCLE ("GAN-CIRCLE[u]") in [74], i.e. no paired LRCT and HRCT were used in training. Our quantitative results (PSNR and SSIM) on the Mayo low-dose CT data seem not to be consistent with that in Table 4.1 of [74] (except for the PSNR relationship, i.e. SRGAN has a higher PSNR than GAN-CIRCLE[u]). This inconsistency may be due to several reasons: 1) the split of training and test data

is not the same (unclear in[74]); 2) the super-resolution is different (1X in each dimension in[74] and 4X in this work); 3) the reconstruction methods are different (FBP in [74] and CGLS in this work); and 4) the implementation of SRGAN and GAN-CIRCLE may be different although we downloaded the original code and tuned the hyperparameters based on PSNR. Furthermore, the caveat should be kept in mind that these quantitative metrics may not reflect the real impact of image quality on the clinical decision making. In future studies, the task-based evaluation should be used for the relevant clinical task, such as liver lesion detection or artery plaque quantification etc.

TTSR has been developed originally for SR of natural images [95]. A detailed ablation study was conducted to show that all three loss functions together achieved the best visual results. In this work, we adapted the optimal structure suggested by the original TTSR and tuned the learning rate for SRCT. In addition, we compared two other deep-learning based methods, SRGAN and GAN-CIRCLE, as they were extensively studied and compared for SRCT[74]. As our primary goal is to demonstrate that the current implementation of TTSR could achieve SRCT comparable to or better than other state-of-the-art methods, such as SRGAN and GAN-CIRCLE, we leave the detailed ablation study in future work, which may determine the effectiveness of different components of TTSR for improved computational efficiency and/or SRCT performance.

In summary, we proposed a TTSR method for low-dose CT super-resolution in this work. TTSR based on texture transformer and attention mechanism is effective to improve the spatial resolution and to suppress the noise of low-dose CT images

for both phantom and patient data. The high-quality super-resolution CT images can be obtained through TTSR even with much reduced dose (1/4) and fewer projection data (1/4), which can lower ionizing radiation and computation time. This development could not only contribute to conventional CT super-resolution, but also improve image quality using less expensive CT detectors.

# CHAPTER 5.    SUMMARY

This work is primarily centered on improving the image quality of computed tomography (CT) scans. The goal has always been to minimize potential harm to patients while delivering superior imaging quality. Our approach consists of two distinct categories: conventional numerical methods and deep learning techniques.

For numerical methods, we developed G-SMEIR for CT, which includes motion estimation in both projection and image domains and iterative reconstruction techniques. With G-SMEIR, we made use of all available information across all phases to enhance image quality under low-dose conditions. Initially, we reconstructed a series of noisy 3D images using isolated respiratory phases. Afterward, we estimated motion among different phases and incorporated the information into the iterative reconstruction process. Given the potential inaccuracies of motion estimation from noisy starting images, we estimated the motion in both projection and image domains and looped motion estimation in the iterative reconstruction process to greatly improve the motion estimation, thus the final reconstruction images.

In the deep learning part, we advanced the CycleGAN/ReCycleGAN for low-dose CT denoising, and TTSR for low-dose CT super-resolution. These techniques are two avenues towards image quality improvement. In the denoising process, since the original projection data are unattainable, we proposed to denoise CT images directly. To maintain generality, we implemented the cycle-consistent loss, which

enables unpaired training, i.e. not one-to-one correspondence between a full-dose image and a low-dose image. This is very important as the labeling the corresponding full-dose and low-dose images is not only time consuming and costly, but also unethical to scan patients twice before enough evidence is collected to show the improved benefit-to-risk ratio of low-dose CT. For CT super-resolution, as original projection data is available, we combined projection pixels to lessen noise even before reconstruction at the cost of loss of image resolution. Then, TTSR kicked in to achieve clean and high-resolution CT images. This approach can also be applied to low-cost low-resolution CT detectors to obtain high-resolution CT images.

Testing our methods initially involved computer simulated phantoms for several reasons. Since phantoms are known entities, we can juxtapose results with the ground truth, allowing for early detection and correction of any discrepancies in our methods or code. Furthermore, accessing patient data in a clinical setting can be challenging. Initiating testing with patient data might lead to model overfitting to a specific patient or device, given the scarcity of data sources. Nevertheless, our methods developed on phantoms were all tested on real patient data for their clinical viability.

In conclusion, the crux of our work is to continuously improve the quality of CT images and reduce the radiation dose to a negligible level for patients. As the motion estimation and image reconstruction techniques as well as deep learning methods are continuously evolving, we are committed to a path where patients receive enhanced diagnostic clarity with minimal potential harm. The methods

discussed in this work are distinct yet complimentary approaches towards this objective.

# APPENDIX A. PSEUDO-CODE OF G-SMEIR

The G-SMEIR pseudo-code is listed below along with SMEIR, TV minimization [36] and the image domain Demons registration [103, 104].

## A.1 Pseudocode for G-SMEIR algorithm

Input and Parameter Setup:

$v = \{v^{r \rightarrow t}\}$ - all DVF pairs

$p = \{p^r\}$ - all projection data

$\mu^r$ - initialization of image at the reference phase $r$ (from 1 to $P$) from 3D TV reconstruction

$M_{Iteration} = 2$ - number of maximum iterations in image domain DVF updating step (demons)

$N_{Iteration} = 11$ - number of maximum iterations in image updating step (SMEIR)

$P = 10$ - number of phases

for $i = 1: M_{Iteration}$

    for $r = 1: P$

for $t = 1:P$

      update $v^{r \rightarrow t}$ using Demons registration of $\mu^r$ and $\mu^t$

    end for

end for

for $j = 1:N_{Iteration}$

    for $r = 1:P$

      update $\mu^r$ using SMEIR for the reference phase $r$, DVF $\boldsymbol{v}$ and all

      projection data $\boldsymbol{p}$

    end for

end for

end for

## A.2 Pseudocode for SMEIR algorithm

Input and Parameter Setup:

$\boldsymbol{p} = \{p^r\}$ - all projection data

$\mu_j^{r,(0)}$ - initialization of image at the reference phase $r$, obtained by 3D TV reconstruction or previous SMEIR iteration

$v^{r \to t,(0)}$ - initialization of DVF (from the reference phase $r$ to the target phase $t$), estimated from the DVF obtained from the initial

$I_{Iteration} = 20$ - number of maximum iterations in image updating step

$M_{Iteration} = 10$ - number of maximum iterations in DVF updating step

$P = 10$ - number of phases

$\varepsilon = 1 \times 10^{-4}$; $\alpha_{red} = 0.5$ - two parameters in backtracking line search for updating DVF in projection domain.

$\lambda = 0.1$.

$\lambda_{red} = 1$.

Iteration:

for $k = 0 : I_{interation}$

   for $j = 1 : J$ do - image update

$$\mu_j^{r,(k+1)} = \mu_j^{r,(k)} + \lambda \lambda_{red}^k \frac{\sum_{t,n} d_{jn}^{t \to r} \sum_i \left[ a_{in} \frac{p_i^t - \sum_n a_{in} \mu_n^{t,(k)}}{\sum_{n=1}^J a_{in}} \right]}{\sum_{t,n} d_{jn}^{t \to r} \sum_i a_{in}}$$

$$\mu_n^{t,(k)} = \sum_j d_{jn}^{r \to t} \mu_j^{r,(k)}$$

end for

Perform TV minimization on $\mu_j^{r,(k+1)}$

for $t = \{1:P\}\backslash r$

$\quad q^{r\to t,(0)} = \nabla f_1\big(v^{r\to t,(0)}\big); h^{r\to t,(0)} = -q^{r\to t,(0)};$

$\quad$ for $m = 1:M_{interation}$ do - DVF update by minimizing $f_1$ and $f_2$ alternately

$\quad\quad \alpha = 1.0$ - initial step size for minimizing $f_1$

$\quad\quad$ while $f_1\big(v^{r\to t,(m)} + \alpha h^{r\to t,(m)}\big) > f_1\big(v^{r\to t,(m)}\big)$

$\quad\quad +\varepsilon\alpha\nabla f_1\big(v^{r\to t,(m)}\big)h^{r\to t,(m)}$ do

$\quad\quad\quad \alpha = \alpha \times \alpha_{red}$

$\quad\quad$ end while

$\quad\quad v^{r\to t,(m+1)} = v^{r\to t,(m)} + \alpha h^{r\to t,(m)}$

$\quad\quad q^{r\to t,(m+1)} = \nabla f_1\big(v^{r\to t,(m+1)}\big)$

$\quad\quad \rho = \dfrac{[q^{r\to t,(m+1)}]'(q^{r\to t,(m+1)}-q^{r\to t,(m)})}{[q^{r\to t,(m)}]'q^{r\to t,(m)}}$

$\quad\quad h^{r\to t,(m+1)} = -q^{r\to t,(m+1)} + \rho h^{r\to t,(m)}$

$\quad\quad v^{t\to r,(m)} = -v^{r\to t,(m)}$

$$q^{t \to r,(m)} = \nabla f_2\left(v^{t \to r,(m)}\right); \; h^{t \to r,(m)} = -q^{t \to r,(m)};$$

$\alpha = 1.0$ - initial step size for minimizing $f_2$

while $f_2\left(v^{t \to r,(m)} + \alpha h^{t \to r,(m)}\right) > f_2\left(v^{t \to r,(m)}\right)$

$+\varepsilon \alpha \nabla f_2\left(v^{t \to r,(m)}\right) h^{t \to r,(m)}$ do

$$\alpha = \alpha \times \alpha_{red}$$

end while

$$v^{t \to r,(m+1)} = v^{t \to r,(m)} + \alpha h^{t \to r,(m)}$$

$$q^{t \to r,(m+1)} = \nabla f_2\left(v^{t \to r,(m+1)}\right)$$

$$\rho = \frac{\left[q^{t \to r,(m+1)}\right]'\left(q^{t \to r,(m+1)} - q^{t \to r,(m)}\right)}{\left[q^{t \to r,(m)}\right]' q^{t \to r,(m)}}$$

$$h^{t \to r,(m+1)} = -q^{t \to r,(m+1)} + \rho h^{t \to r,(m)}$$

$$v^{r \to t,(m+1)} = -v^{t \to r,(m+1)};$$

end for

end for

end for

## A.3 Pseudocode for TV minimization

Input and Parameter Setup:

$\mu$ - image with J voxels.

$Iteration = 5$ - number of maximum iterations

$\gamma_{red} = 0.8$ - parameter in line search

Iteration:

for $n = 1$ to $Iteration$ do

    $\gamma = 0.8$ - initial step size

    for $j$ = 1: J do

        $\partial g_j = \nabla g(\mu_j)$

        $|\nabla g(\mu)| = \sqrt{\Sigma_j (\partial g_j)^2}.$

    end for

    for $j$ = 1: J do

        $\partial g_j = \partial g_j / |\nabla g(\mu)|$

        $\mu'_j = \mu_j - \gamma \cdot \partial g_j$

if $\mu'_j < 0$

$$\mu'_j = 0$$

end if

end for

while $g(\mu'_j) > g(\mu_j)$ do

$$\gamma = \gamma \times \gamma_{red}$$

for *j = 1: J* do

$$\partial g_j = \partial g_j / |\nabla g(\mu)|$$

$$\mu'_j = \mu_j - \gamma \cdot \partial g_j$$

if $\mu'_j < 0$

$$\mu'_j = 0$$

end if

end for

end while

$$\mu_j = \mu'_j$$

end for

## A.4 Pseudocode for Demons registration

Input and Parameter Setup:

$\mu^r$ - the reference image

$\mu^t$ - the target image

$N = 100$ - number of maximum iterations

$f$ – Gaussian filter kernel

Initialization:

$T_x = 0$ – Transformation vector in x direction

$T_y = 0$ – Transformation vector in y direction

$T_z = 0$ – Transformation vector in z direction

$\boldsymbol{g} = \nabla\mu^r$ – Gradient of the reference image

Iteration:

for $k = 1:N$

$$d = I_{mov} - I_{ref}$$

$$u_x = \frac{d * g_x}{g^2 + d^2}$$

$$u_y = \frac{d * g_y}{g^2 + d^2}$$

$$u_z = \frac{d * g_z}{g^2 + d^2}$$

$$T_x = T_x + u_x$$

$$T_y = T_y + u_y$$

$$T_z = T_z + u_z$$

$$T_x = \sum_i \sum_j \sum_k f(i,j,k) T_x(x - i, y - j, z - k)$$

$$T_y = \sum_i \sum_j \sum_k f(i,j,k) T_y(x - i, y - j, z - k)$$

$$T_z = \sum_i \sum_j \sum_k f(i,j,k) T_z(x - i, y - j, z - k)$$

$$\mu^t = \mu^r + \boldsymbol{T}$$

end for

$$v^{r \to t} = \boldsymbol{T}$$

The parameters shown in the pseudo code were used to reconstruct images in this work, unless otherwise stated. Particularly, to balance the TV constraints and the data fidelity term, the numbers of ART and TV iterations were set as 20 and 5, respectively. The step size parameters, $\lambda$, $\lambda_{red}$, $\gamma$, $\gamma_{red}$, were 0.1, 1, 0.1, and 0.8 for the phantom data, and 0.1, 0.99, 0.3, and 0.9 for the patient data. These parameters were tuned using 3D TV for the best RMSE of phantom reconstruction and visual inspection of patient reconstruction. In addition, G-SMEIR performance is influenced by $M_{Iteration}$ and $N_{Iteration}$ as shown in the results.

# APPENDIX B.  RECYCLEGAN NETWORK STRUCTURE

The generator, predictor, and discriminator of RecycleGAN are shown in Table A.I, Table A.II, and Table A.III, respectively. For the CycleGAN structure, please refer to Fig. 4 and Fig. 5 in [75].

## B.1 The generator structure of RecycleGAN

```
===============================================================

Function blocks (level # - layer #)          Kernel Shape

===============================================================

Generator                      --

├─Sequential: 1-1                 --

│    └─ReflectionPad2d: 2-1          --

│    └─Conv2d: 2-2               [1, 64, 7, 7]

│    └─InstanceNorm2d: 2-3           --

│    └─ReLU: 2-4                --

│    └─Conv2d: 2-5              [64, 128, 3, 3]

│    └─InstanceNorm2d: 2-6           --

│    └─ReLU: 2-7                --
```

| └─Conv2d: 2-8                    [128, 256, 3, 3]

| └─InstanceNorm2d: 2-9                --

| └─ReLU: 2-10                --

| └─ResnetBlock: 2-11                --

| └─ResnetBlock: 2-12                --

| └─ResnetBlock: 2-13                --

| └─ResnetBlock: 2-14                --

| └─ResnetBlock: 2-15                --

| └─ResnetBlock: 2-16                --

| └─ConvTranspose2d: 2-17            [128, 256, 3, 3]

| └─InstanceNorm2d: 2-18                --

| └─ReLU: 2-19                --

| └─ConvTranspose2d: 2-20            [64, 128, 3, 3]

| └─InstanceNorm2d: 2-21                --

| └─ReLU: 2-22                --

| └─ReflectionPad2d: 2-23                --

| └─Conv2d: 2-24                [64, 1, 7, 7]

```
|     └─Tanh: 2-25                              --
```

## B.2 The predictor structure of RecycleGAN

```
===============================================================

Function blocks (level # - layer #)          Kernel Shape

===============================================================

Predictor                                                  --

├─UnetSkipConnectionBlock: 1-1                             --

|     └─Sequential: 2-1                                    --

|     |     └─Conv2d: 3-1                          [2, 64, 4, 4]

|     |     └─UnetSkipConnectionBlock: 3-2                 --

|     |     |     └─Sequential: 4-1                        --

|     |     |     |     └─LeakyReLU: 5-1                    --

|     |     |     |     └─Conv2d: 5-2              [64, 128, 4, 4]

|     |     |     |     └─InstanceNorm2d: 5-3              --

|     |     |     |     └─UnetSkipConnectionBlock: 5-4     -
-
```

| | | | | └─Sequential: 6-1                                    --

| | | | | | └─LeakyReLU: 7-1                                    --

| | | | | | └─Conv2d: 7-2                          [128, 256, 4, 4]

| | | | | | └─InstanceNorm2d: 7-3                                --

| | | | | | └─UnetSkipConnectionBlock: 7-4                       --

| | | | | | | └─Sequential: 8-1                                 --

| | | | | | | | └─LeakyReLU: 9-1                                --

| | | | | | | | └─Conv2d: 9-2                      [256, 512, 4, 4]

| | | | | | | | └─InstanceNorm2d: 9-3                            --

| | | | | | | | └─UnetSkipConnectionBlock: 9-4                  --

| | | | | | | | | └─Sequential: 10-1                            --

| | | | | | | | | | └─LeakyReLU: 11-1                           --

| | | | | | | | | | └─Conv2d: 11-2                 [512, 512, 4, 4]

| | | | | | | | | | └─InstanceNorm2d: 11-3

│ │ │ │ │ │ │ │ │ │ └─UnetSkipConnectionBlock: 11-4

│ │ │ │ │ │ │ │ │ │ │ └─Sequential: 12-1　　　　　　-

│ │ │ │ │ │ │ │ │ │ │ └─LeakyReLU: 13-1

│ │ │ │ │ │ │ │ │ │ │ └─Conv2d: 13-2　[512, 512, 4, 4]

│ │ │ │ │ │ │ │ │ │ │ └─InstanceNorm2d: 13-3

│ │ │ │ │ │ │ │ │ │ │ └─UnetSkipConnectionBlock: 13-4

│ │ │ │ │ │ │ │ │ │ │ │ └─Sequential: 14-1

│ │ │ │ │ │ │ │ │ │ │ │ │ └─LeakyReLU: 15-1

│ │ │ │ │ │ │ │ │ │ │ │ │ │ └─Conv2d: 15-2　[512, 512, 4, 4]

│ │ │ │ │ │ │ │ │ │ │ │ │ │ └─ReLU: 15-3

│ │ │ │ │ │ │ │ │ │ │ │ │ │ └─ConvTranspose2d: 15-4　[512, 512, 4, 4]

│ │ │ │ │ │ │ │ │ │ │ │ │ │ └─InstanceNorm2d: 15-5

│ │ │ │ │ │ │ │ │ │ │ │ │ └─ReLU: 13-5

│ │ │ │ │ │ │ │ │ │ │ │ │ └─ConvTranspose2d: 13-6　[512, 1024, 4, 4]

│ │ │ │ │ │ │ │ │ │ │ │    └─InstanceNorm2d: 13-7

│ │ │ │ │ │ │ │ │ │ │    └─ReLU: 11-5                              --

│ │ │ │ │ │ │ │ │ │ │    └─ConvTranspose2d: 11-6      [512, 1024, 4, 4]

│ │ │ │ │ │ │ │ │ │ │    └─InstanceNorm2d: 11-7

│ │ │ │ │ │ │ │ │ │    └─ReLU: 9-5                              --

│ │ │ │ │ │ │ │ │    └─ConvTranspose2d: 9-6        [256, 1024, 4, 4]

│ │ │ │ │ │ │ │ │    └─InstanceNorm2d: 9-7

│ │ │ │ │ │ │    └─ReLU: 7-5                              --

│ │ │ │ │ │ │    └─ConvTranspose2d: 7-6             [128, 512, 4, 4]

│ │ │ │ │ │ │    └─InstanceNorm2d: 7-7                              --

│ │ │ │    └─ReLU: 5-5                              --

│ │ │ │    └─ConvTranspose2d: 5-6             [64, 256, 4, 4]

│ │ │ │    └─InstanceNorm2d: 5-7                              --

│ │    └─ReLU: 3-3                              --

│ │    └─ConvTranspose2d: 3-4                   [1, 128, 4, 4]

│ │    └─Tanh: 3-5                              --

## B.3 The discriminator structure of RecycleGAN

```
================================================================

Function blocks (level # - layer #)          Kernel Shape

================================================================

Discriminator                    --

├─Sequential: 1-1                    --

│    └─Conv2d: 2-1                [1, 64, 4, 4]

│    └─LeakyReLU: 2-2                --

│    └─Conv2d: 2-3                [64, 128, 4, 4]

│    └─InstanceNorm2d: 2-4           --

│    └─LeakyReLU: 2-5                --

│    └─Conv2d: 2-6                [128, 256, 4, 4]

│    └─InstanceNorm2d: 2-7           --

│    └─LeakyReLU: 2-8                --

│    └─Conv2d: 2-9                [256, 512, 4, 4]

│    └─InstanceNorm2d: 2-10          --
```

|    └─LeakyReLU: 2-11            --

|    └─Conv2d: 2-12            [512, 1, 4, 4]

Definition of fundamental blocks:

Conv2d: 2D convolution

ReLU: rectified linear unit function

LeakyReLU: A type of activation function based on a ReLU, which has a small slope for negative values instead of strict zero

ConvTranspose2d: Applies a 2D transposed convolution operator over an input image composed of several input planes.

Tanh: Applies the Hyperbolic Tangent function element-wise.

ReflectionPad2d: Pads the input tensor using the reflection of the input boundary.

InstanceNorm2d: Applies Instance Normalization over a 4D input

ResnetBlock: Sequence of ReflectionPad2d, Conv2d(shape of [256, 256, 3, 3]), InstanceNorm2d, ReLU, ReflectionPad2d, Conv2d(shape of [256, 256, 3, 3]), and InstanceNorm2d functions

# APPENDIX C.  TTSR NETWORK STRUCTURE

The generator of TTSR is shown in B.1 and the discriminator in B.2. For more details, please refer to the supplementary material in [95].

## C.1 The generator structure of TTSR

```
================================================================

Function blocks (level # - layer #)          Kernel Shape

================================================================

├─Generator: 1-1                    --

|     └─SFE: 2-1                     --

|     |     └─Conv2d: 3-1           [3, 64, 3, 3]

|     |     └─16 x ResBlock: 3-2         --

|     |     └─Conv2d: 3-3           [64, 64, 3, 3]

|     └─Conv2d: 2-2                 [320, 64, 3, 3]

|     └─16 x ResBlock: 2-3              --

|     └─Conv2d: 2-4                 [64, 64, 3, 3]
```

| └─Conv2d: 2-5 [64, 256, 3, 3]

| └─PixelShuffle: 2-6 --

| └─Conv2d: 2-7 [192, 64, 3, 3]

| └─CSFI2: 2-8 --

| | └─Conv2d: 3-20 [64, 64, 1, 1]

| | └─Conv2d: 3-21 [64, 64, 3, 3]

| | └─Conv2d: 3-22 [128, 64, 3, 3]

| | └─Conv2d: 3-23 [128, 64, 3, 3]

| └─8 x ResBlock: 2-9 --

| └─8 x ResBlock: 2-10 --

| └─Conv2d: 2-11 [64, 64, 3, 3]

| └─Conv2d: 2-12 [64, 64, 3, 3]

| └─Conv2d: 2-13 [64, 256, 3, 3]

| └─PixelShuffle: 2-14 --

| └─Conv2d: 2-15 [128, 64, 3, 3]

```
|       └─CSFI3: 2-16                --

|   |       └─Conv2d: 3-40          [64, 64, 1, 1]

|   |       └─Conv2d: 3-41          [64, 64, 1, 1]

|   |       └─Conv2d: 3-42          [64, 64, 3, 3]

|   |       └─Conv2d: 3-43          [64, 64, 1, 1]

|   |       └─Conv2d: 3-44          [64, 64, 3, 3]

|   |       └─Conv2d: 3-45          [64, 64, 3, 3]

|   |       └─Conv2d: 3-46          [64, 64, 3, 3]

|   |       └─Conv2d: 3-47          [192, 64, 3, 3]

|   |       └─Conv2d: 3-48          [192, 64, 3, 3]

|   |       └─Conv2d: 3-49          [192, 64, 3, 3]

|   └─4 x ResBlock: 2-17             --

|   └─4 x ResBlock: 2-18             --

|   └─4 x ResBlock: 2-19             --

|   └─Conv2d: 2-20                  [64, 64, 3, 3]
```

```
|       └─Conv2d: 2-21                [64, 64, 3, 3]

|       └─Conv2d: 2-22                [64, 64, 3, 3]

|       └─MergeTail: 2-23              --

|    |      └─Conv2d: 3-62         [64, 64, 1, 1]

|    |      └─Conv2d: 3-63         [64, 64, 1, 1]

|    |      └─Conv2d: 3-64         [192, 64, 3, 3]

|    |      └─Conv2d: 3-65         [64, 32, 3, 3]

|    |      └─Conv2d: 3-66         [32, 3, 1, 1]

├─LTE: 1-2                        --

|    └─Sequential: 2-24              --

|    |      └─Conv2d: 3-67         [3, 64, 3, 3]

|    |      └─ReLU: 3-68            --

|    └─Sequential: 2-25              --

|    |      └─Conv2d: 3-69         [64, 64, 3, 3]

|    |      └─ReLU: 3-70            --
```

```
|   |       └─MaxPool2d: 3-71           --

|   |       └─Conv2d: 3-72             [64, 128, 3, 3]

|   |       └─ReLU: 3-73               --

|       └─Sequential: 2-26            --

|   |       └─Conv2d: 3-74             [128, 128, 3, 3]

|   |       └─ReLU: 3-75               --

|   |       └─MaxPool2d: 3-76           --

|   |       └─Conv2d: 3-77             [128, 256, 3, 3]

|   |       └─ReLU: 3-78               --

|       └─MeanShift: 2-27             [3, 3, 1, 1]

├─LTE: 1-3 (repeat LET: 1-2)          --

├─SearchTransfer: 1-4                 --
```

## C.2 The discriminator structure of TTSR

```
================================================================
```

Function blocks (level # - layer #)        Kernel Shape

```
============================================================

├─Discriminator: 1-1                --

|     └─Conv2d: 2-1               [3, 32, 3, 3]

|     └─LeakyReLU: 2-2             --

|     └─Conv2d: 2-3              [32, 32, 3, 3]

|     └─LeakyReLU: 2-4             --

|     └─Conv2d: 2-5              [32, 64, 3, 3]

|     └─LeakyReLU: 2-6             --

|     └─Conv2d: 2-7              [64, 64, 3, 3]

|     └─LeakyReLU: 2-8             --

|     └─Conv2d: 2-9              [64, 128, 3, 3]

|     └─LeakyReLU: 2-10            --

|     └─Conv2d: 2-11             [128, 128, 3, 3]

|     └─LeakyReLU: 2-12            --

|     └─Conv2d: 2-13             [128, 256, 3, 3]
```

|       └─LeakyReLU: 2-14             --

|       └─Conv2d: 2-15              [256, 256, 3, 3]

|       └─LeakyReLU: 2-16             --

|       └─Conv2d: 2-17              [256, 512, 3, 3]

|       └─LeakyReLU: 2-18             --

|       └─Conv2d: 2-19              [512, 512, 3, 3]

|       └─LeakyReLU: 2-20             --

|       └─Linear: 2-21              [12800, 1024]

|       └─LeakyReLU: 2-22             --

|       └─Linear: 2-23              [1024, 1]


Definition of fundamental blocks:

Conv2d: 2D convolution

ReLU: rectified linear unit function

MaxPool2d: 2D max pooling

MeanShift: Subtract channel-wise mean from the input

PixelShuffle: Rearranges elements

SearchTransfer: Sequence of unfold, permute, normalize, multiply, max, unfold, fold functions

ResBlock: Sequence of Conv2d and ReLU functions

LeakyReLU: A type of activation function based on a ReLU, which has a small slope for negative values instead of strict zero

Linear: Linear transformation

# APPENDIX D.  PUBLICATIONS & AWARDS

## D.1 Journal

1.      Shiwei Zhou, Jinyu Yang, Krishnateja Koduri, Junzhou Huang, Lifeng Yu and Mingwu Jin. "Spatiotemporal Denoising of Low-dose Cardiac CT Image Sequences using RecycleGAN ." (Submitted)

2.      Shiwei Zhou, Lifeng Yu, and Mingwu Jin. "Texture transformer super-resolution for low-dose computed tomography." Biomedical Physics & Engineering Express 8.6 (2022): 065024.

3.      Shiwei Zhou, Yujie Chi, Jing Wang, and Mingwu Jin. "General simultaneous motion estimation and image reconstruction (G-SMEIR)." Biomedical physics & engineering express 7, no. 5 (2021): 055011.

4.      Lai Youfang, Qian Wang, Shiwei Zhou, Zhaoheng Xie, Jinyi Qi, Simon R. Cherry, Mingwu Jin, Yujie Chi, and Junwei Du. "H2RSPET: a 0.5 mm resolution high-sensitivity small-animal PET scanner, a simulation study." Physics in Medicine & Biology 66.6 (2021): 065016.

5.      Zeheng Li, Shiwei Zhou, Junzhou Huang, Lifeng Yu, and Mingwu Jin. "Investigation of low-dose CT image denoising using unpaired deep learning methods." IEEE transactions on radiation and plasma medical sciences 5, no. 2 (2020): 224-234. (Co-first author)

## D.2 Conference

1.　　Zhou, Shiwei, et al. "Comparison of Different Attenuation Correction Methods for Dual-Gated Cardiac SPECT." 2023 American Association of Physicists in Medicine Conference (AAPM).

2.　　Zhou, Shiwei, et al. "Multi-Phase Cardiac CT Angiography (CTA) Denoising Using a Spatiotemporal Deep Learning Method." 2023 American Association of Physicists in Medicine Conference (AAPM).

3.　　Zhou, Shiwei, et al. "Deep-Learning Based Super-Resolution for Low-Dose CT." 2022  IEEE Nuclear Science Symposium, Medical Imaging Conference and Room Temperature Semiconductor Detector Conference (NSS-MIC-RTSD).

4.　　Zhou, Shiwei, et al. "Texture Transformer Super-Resolution (TTSR) for Patient CT Images." 2022 American Association of Physicists in Medicine Conference (AAPM).

5.　　Zhou, Shiwei, et al. "Texture Transformer Super-Resolution for Computed Tomography." 2021 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC).

6.　　Zhou, Shiwei, et al. "Low-Dose CT Image Enhancement Through a Texture Transformer." 2021 American Association of Physicists in Medicine Conference (AAPM).

7.　　Zhou, Shiwei, et al. "A Spatiotemporal Denoising Method for Low-Dose Cardiac CT Images." 2021 American Association of Physicists in Medicine Conference (AAPM).

8.      Zhou, Shiwei, et al. "Improvement and Evaluation of General Simultaneous Motion Estimation and Image Reconstruction (G-SMEIR)." 2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC).

9.      Zhou, Shiwei, et al. "Supervised and Unsupervised Deep Learning Methods for Low-Dose CT Image Denoising." 2020 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC).

10.     Zhou, Shiwei, et al. "Pursuit of Efficient Image Domain Motion Estimation for G-SMEIR." 2020 American Association of Physicists in Medicine Conference (AAPM).

11.     Zhou, Shiwei, et al. "Development and Evaluation of General Simultaneous Motion Estimation and Image Reconstruction (G-SMEIR)." 2020 American Association of Physicists in Medicine Conference (AAPM).

12.     Zhou, Shiwei, et al. "4D Reconstruction with Projection and Image Domain Motion Estimation." 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC).

13.     Lai, Youfang, Yuncheng Zhong, Ananta Chalise, Shiwei Zhou, Yiping Shao, Mingwu Jin, Xun Jia, and Yujie Chi. "Development of a GPU-Based Mont Carlo Simulation Tool for PET." In 2019 IEEE Nuclear Science Symposium and Medical Imaging Conference (NSS/MIC), pp. 1-3. IEEE, 2019.

14.    Zhou, Shiwei, et al. "General Simultaneous Motion Estimation and Image Reconstruction (G-SMEIR) for CBCT." 2019 American Association of Physicists in Medicine Conference (AAPM).

## D.3 Significant Awards

1.    2022 – 2023 TACC Frontera Computational Science Fellowship (National award)

# APPENDIX E.  COPYRIGHT PERMISSION

Enclosed is the written copyright permission for Chapters 2 and 4. Owing to certain restrictions, only the original versions have been included in the dissertation. Links to the final published versions can be found in the footnotes of Chapters 2 and 4.

**Re: Request for Copyright Permission for PhD Dissertation**

Permissions <permissions@ioppublishing.org>
Mon 2023-08-14 2:58
To:Zhou, Shiwei <shiwei.zhou@mavs.uta.edu>
Cc:Jin, Mingwu <mingwu@uta.edu>
[External]
Dear Shiwei Zhou,

Thank you for your email and for taking the time to seek this permission.

When you transferred the copyright in your articles to IOP, we granted back to you certain rights, including the right to include all or part of the Final Published Version of the articles within any thesis or dissertation provided it is not then shared or deposited online. Full details can be found in our Author Rights Policy.

If you are required by your institution to share your thesis/dissertation publicly (such as in an institutional repository), the Final Published Versions would need to be removed prior to publication online or in print. Your institution should be able to withhold the IOP article sections of your thesis from this version. However, you should still reference the article, include the abstract and provide DOI links to them on IOPscience so that people know that they have been published.

Please include citation details, "© IOP Publishing.  Reproduced with permission.  All rights reserved" and for online use, a link to the Version of Record.

Please note you may need to obtain separate permission for any third party content you included within your articles.

If you do not wish to remove the IOP article sections of your thesis, you also have the following alternative options:

1) Our Author Rights Policy allows authors to post the Accepted Manuscript in certain places **after** the embargo period has elapsed. This includes your institutional repository, subject to certain restrictions (please see the full terms and conditions in our Author Rights Policy linked in the signature of this email.). Therefore, an option may be to include your Accepted Manuscript article version in your thesis or dissertation, as it would allow you to place the unaltered dissertation onto your repository after the embargo period.

**Please note that one of your articles is still within its embargo period**. An Embargo Period is 'a period of 12 months from the Date of Publication'.

2)  A further option would be to include the Author's Original or Preprint version of the article, provided that you do so  in line with our Preprint pre-publication policy.

I wish you the best of luck with the completion of your thesis/dissertation.

Kind regards,


Sophie
**Copyright & Permissions Team**
Sophie Brittain - Rights & Permissions Assistant
Cameron Wood - Legal & Rights Adviser
Contact Details
E-mail: permissions@ioppublishing.org

For further information about copyright and how to request permission:
https://publishingsupport.iopscience.iop.org/copyright-journals/
See also:  https://publishingsupport.iopscience.iop.org/
Please see our Author Rights Policy https://publishingsupport.iopscience.iop.org/author-rights-policies/

**Please note**: We do not provide signed permission forms as a separate attachment. Please print this email and provide it to your publisher as proof of permission. **Please note**: Any statements made by IOP Publishing to the effect that authors do not need to get permission to use any content where IOP Publishing is not the publisher is not intended to constitute any sort of legal advice. Authors must make their own decisions as to the suitability of the content they are using and whether they require permission for it to be published within their article.

---

**From:** Zhou, Shiwei <shiwei.zhou@mavs.uta.edu>
**Sent:** 09 August 2023 18:05
**To:** Permissions <permissions@ioppublishing.org>
**Cc:** Jin, Mingwu <mingwu@uta.edu>
**Subject:** Request for Copyright Permission for PhD Dissertation

Dear Rights and Permissions team,
I hope this email finds you well. My name is Shiwei Zhou, and I am currently working on my Ph.D. Dissertation. As part of my research, I would like to include two published papers for which I am seeking your permission to reproduce. These papers are integral to my work and will significantly contribute to the context and substance of my dissertation.

The papers I wish to include are as follows:
1. Shiwei Zhou, Lifeng Yu, and Mingwu Jin. "Texture transformer super-resolution for low-dose computed tomography." Biomedical Physics & Engineering Express 8.6 (2022): 065024.
2. Shiwei Zhou, Yujie Chi, Jing Wang, and Mingwu Jin. "General simultaneous motion estimation and image reconstruction (G-SMEIR)." Biomedical physics & engineering express 7.5 (2021): 055011.

Understanding the time-sensitive nature of this request, I kindly ask for your approval by August 14th, as my dissertation submission deadline is August 15th. Your prompt response will greatly assist in maintaining my timeline.

Should you require any additional information or documentation, please do not hesitate to contact me. I appreciate your consideration of this request, and I look forward to hearing from you soon.

Thank you for your time and support.

Sincerely,
Shiwei Zhou

---

# REFERENCE

1.  Kalender, W.A., *X-ray computed tomography.* Physics in Medicine & Biology, 2006. **51**(13): p. R29.

2.  Hsieh, J., *Computed tomography: principles, design, artifacts, and recent advances.* 2003.

3.  Flohr, T.G., et al., *Multi–Detector Row CT Systems and Image-Reconstruction Techniques.* Radiology, 2005. **235**(3): p. 756-773.

4.  Kalender, W.A., et al., *Spiral volumetric CT with single-breath-hold technique, continuous transport, and continuous scanner rotation.* Radiology, 1990. **176**(1): p. 181-183.

5.  Pan, X., et al., *Anniversary Paper: Development of x-ray computed tomography: The role of Medical Physics and AAPM from the 1970s to present.* Medical Physics, 2008. **35**(8): p. 3728-3739.

6.  Wernick, M.N. and J.N. Aarsvold, *Emission tomography: the fundamentals of PET and SPECT*. 2004: Elsevier.

7.  Durrani, T.S. and D. Bisset, *The Radon transform and its properties.* GEOPHYSICS, 1984. **49**(8): p. 1180-1187.

8.  Fritz, G. and U.G. Oppel. *Generalized Radon Transformations*. in *Mathematical Aspects of Computerized Tomography*. 1981. Berlin, Heidelberg: Springer Berlin Heidelberg.

9.  Vedam, S.S., et al., *Quantifying the predictability of diaphragm motion during respiration with a noninvasive external marker.* Medical Physics, 2003. **30**(4): p. 505-513.

10. Rit, S., et al., *Quantification of the Variability of Diaphragm Motion and Implications for Treatment Margin Construction.* International Journal of Radiation Oncology, Biology, Physics, 2012. **82**(3): p. e399-e407.

11. Pan, T., et al., *4D-CT imaging of a volume influenced by respiratory motion on multi-slice CT.* Medical Physics, 2004. **31**(2): p. 333-340.

12. Keall, P.J., et al., *The management of respiratory motion in radiation oncology report of AAPM Task Group 76a).* Medical Physics, 2006. **33**(10): p. 3874-3900.

13. Yamamoto, T., et al., *Retrospective Analysis of Artifacts in Four-Dimensional CT Images of 50 Abdominal and Thoracic Radiotherapy*

*Patients.* International Journal of Radiation Oncology, Biology, Physics, 2008. **72**(4): p. 1250-1258.

14. Zhang, Q., et al., *A patient-specific respiratory model of anatomical motion for radiation treatment planning.* Medical Physics, 2007. **34**(12): p. 4772-4781.

15. Berbeco, R.I., et al., *Residual motion of lung tumours in gated radiotherapy with external respiratory surrogates.* Physics in Medicine & Biology, 2005. **50**(16): p. 3655.

16. Chun, S.Y. and J.A. Fessler, *Noise properties of motion-compensated tomographic image reconstruction methods.* IEEE transactions on medical imaging, 2012. **32**(2): p. 141-152.

17. Bai, W. and M. Brady, *Regularized B-spline deformable registration for respiratory motion correction in PET images.* Physics in Medicine & Biology, 2009. **54**(9): p. 2719.

18. Dawood, M., et al., *Lung motion correction on respiratory gated 3-D PET/CT images.* IEEE transactions on medical imaging, 2006. **25**(4): p. 476-485.

19. Klein, G., B. Reutter, and R. Huesman, *Non-rigid summing of gated PET via optical flow.* IEEE Transactions on nuclear science, 1997. **44**(4): p. 1509-1512.

20. Gravier, E., Y. Yang, and M. Jin, *Tomographic reconstruction of dynamic cardiac image sequences.* IEEE transactions on image processing, 2007. **16**(4): p. 932-942.

21. Jin, M., et al., *4D reconstruction for low-dose cardiac gated SPECT.* Medical physics, 2013. **40**(2): p. 022501.

22. Jin, M., Y. Yang, and M.A. King, *Reconstruction of dynamic gated cardiac SPECT.* Medical physics, 2006. **33**(11): p. 4384-4394.

23. Jin, M., et al., *A quantitative evaluation study of four-dimensional gated cardiac SPECT reconstruction.* Physics in Medicine & Biology, 2009. **54**(18): p. 5643.

24. Mair, B.A., D.R. Gilland, and J. Sun, *Estimation of images and nonrigid deformations in gated emission CT.* IEEE transactions on Medical Imaging, 2006. **25**(9): p. 1130-1144.

25. Blume, M., et al., *Joint reconstruction of image and motion in gated positron emission tomography.* IEEE transactions on medical imaging, 2010. **29**(11): p. 1892-1906.

26. Chun, S.Y. and J.A. Fessler. *Joint image reconstruction and nonrigid motion estimation with a simple penalty that encourages local invertibility*. in *Medical Imaging 2009: Physics of Medical Imaging*. 2009. International Society for Optics and Photonics.

27. Jacobson, M. and J.A. Fessler. *Joint estimation of image and deformation parameters in motion-corrected PET*. in *2003 IEEE Nuclear Science Symposium. Conference Record (IEEE Cat. No. 03CH37515)*. 2003. IEEE.

28. Lamare, F., et al., *List-mode-based reconstruction for respiratory motion correction in PET using non-rigid body transformations.* Physics in Medicine & Biology, 2007. **52**(17): p. 5187.

29. Li, T., et al., *Model-based image reconstruction for four-dimensional PET.* Medical physics, 2006. **33**(5): p. 1288-1298.

30. Nehmeh, S.A., et al., *Effect of respiratory gating on quantifying PET images of lung cancer.* Journal of nuclear medicine, 2002. **43**(7): p. 876-881.

31. Qiao, F., et al., *A motion-incorporated reconstruction method for gated PET studies.* Physics in Medicine & Biology, 2006. **51**(15): p. 3769.

32. Ritchie, C.J., et al., *Predictive respiratory gating: a new method to reduce motion artifacts on CT scans.* Radiology, 1994. **190**(3): p. 847-852.

33. Taguchi, K., et al. *Image-domain motion compensated time resolved 4D cardiac CT*. in *Medical Imaging 2007: Physics of Medical Imaging*. 2007. International Society for Optics and Photonics.

34. Kalantari, F., et al., *Respiratory motion correction in 4D-PET by simultaneous motion estimation and image reconstruction (SMEIR).* Physics in Medicine and Biology, 2016. **61**(15): p. 5639-61.

35. Van Stevendaal, U., et al., *A motion-compensated scheme for helical cone-beam reconstruction in cardiac CT angiography.* Medical physics, 2008. **35**(7Part1): p. 3239-3251.

36. Wang, J. and X. Gu, *Simultaneous motion estimation and image reconstruction (SMEIR) for 4D cone-beam CT.* Medical Physics, 2013. **40**(10): p. 101912-101912.

37. Zhao, C., et al., *4D cone-beam computed tomography (CBCT) using a moving blocker for simultaneous radiation dose reduction and scatter correction.* Physics in Medicine & Biology, 2018. **63**(11): p. 115007.

38. Chun, S.Y. and J.A. Fessler, *Spatial resolution properties of motion-compensated tomographic image reconstruction methods.* IEEE transactions on medical imaging, 2012. **31**(7): p. 1413-1425.

39. Sonke, J.J., et al., *Respiratory correlated cone beam CT.* Medical physics, 2005. **32**(4): p. 1176-1186.

40. Mory, C., G. Janssens, and S. Rit, *Motion-aware temporal regularization for improved 4D cone-beam computed tomography.* Physics in Medicine & Biology, 2016. **61**(18): p. 6856.

41. Zhang, H., et al., *Applications of nonlocal means algorithm in low-dose X-ray CT image processing and reconstruction: A review.* Medical physics, 2017. **44**(3): p. 1168-1185.

42. Sweeney, R.A., et al., *Accuracy and inter-observer variability of 3D versus 4D cone-beam CT based image-guidance in SBRT for lung tumors.* Radiation Oncology, 2012. **7**(1): p. 81.

43. Kumar, P., et al., *Four-dimensional Cone Beam CT (4D-CBCT) Significantly Improves Localization of Lung Tumors in Comparison to X-Ray 6D Image Guidance in the Delivery of Stereotactic Body Radiation Therapy (SBRT).* International Journal of Radiation Oncology, Biology, Physics, 2012. **84**(3): p. S743.

44. Shrestha, D., et al., *Dosimetric evaluation of 4D-CBCT reconstructed by Simultaneous Motion Estimation and Image Reconstruction (SMEIR) for carbon ion therapy of lung cancer.* Medical Physics, 2019. **46**(9): p. 4087-4094.

45. Qin, A., et al., *A clinical 3D/4D CBCT-based treatment dose monitoring system.* Journal of Applied Clinical Medical Physics, 2018. **19**(6): p. 166-176.

46. Morton, L.M., et al., *The rising incidence of second cancers: patterns of occurrence and identification of risk factors for children and adults.* American Society of Clinical Oncology Educational Book, 2014. **34**(1): p. e57-e67.

47. Zhao, C., et al., *MODIFIED SIMULTANEOUS MOTION ESTIMATION AND IMAGE RECONSTRUCTION (M-SMEIR) FOR 4D-CBCT*, in *International Symposium of Biomedical Imaging*. 2017: Washington DC. p. To appear.

48. Gu, X., et al., *GPU-based ultra-fast dose calculation using a finite size pencil beam model.* Physics in Medicine & Biology, 2009. **54**(20): p. 6287.

49. Thirion, J.-P., *Image matching as a diffusion process: an analogy with Maxwell's demons.* Medical image analysis, 1998. **2**(3): p. 243-260.

50. Segars, W.P., et al., *4D XCAT phantom for multimodality imaging research.* Medical Physics, 2010. **37**(9): p. 4902-15.

51. Han, G., Z. Liang, and J. You. *A fast ray-tracing technique for TCT and ECT studies*. in *1999 IEEE Nuclear Science Symposium. Conference Record. 1999 Nuclear Science Symposium and Medical Imaging Conference (Cat. No. 99CH37019)*. 1999. IEEE.

52. Carrigan, T.P., et al., *Prognostic utility of 64-slice computed tomography in patients with suspected but no documented coronary artery disease.* European Heart Journal, 2009. **30**(3): p. 362-371.

53. Cademartiri, F., et al., *Diagnostic accuracy of computed tomography coronary angiography in patients with a zero calcium score.* European Radiology, 2010. **20**(1): p. 81-87.

54. Nieman, K., et al., *Coronary angiography with multi-slice computed tomography.* The Lancet, 2001. **357**(9256): p. 599-603.

55. Motoyama, S., et al., *Computed Tomographic Angiography Characteristics of Atherosclerotic Plaques Subsequently Resulting in Acute Coronary Syndrome.* Journal of the American College of Cardiology, 2009. **54**(1): p. 49-57.

56. Sun, Z., et al., *CT virtual intravascular endoscopy assessment of coronary artery plaques: A preliminary study.* European Journal of Radiology, 2010. **75**(1): p. e112-e119.

57. Miszalski-Jamka, T., et al., *The composition and extent of coronary artery plaque detected by multislice computed tomographic angiography provides incremental prognostic value in patients with suspected coronary artery disease.* The International Journal of Cardiovascular Imaging, 2012. **28**(3): p. 621-631.

58. Yu, L., et al., *Radiation dose reduction in computed tomography: techniques and future perspective.* Imaging in medicine, 2009. **1**(1): p. 65.

59. Desjardins, B. and E.A. Kazerooni, *ECG-Gated Cardiac CT.* American Journal of Roentgenology, 2004. **182**(4): p. 993-1010.

60. May, M.S., et al., *Radiation dose considerations by intra-individual Monte Carlo simulations in dual source spiral coronary computed tomography angiography with electrocardiogram-triggered tube current modulation and adaptive pitch.* European Radiology, 2012. **22**(3): p. 569-578.

61. Weustink, A.C., et al., *Optimal Electrocardiographic Pulsing Windows and Heart Rate: Effect on Image Quality and Radiation Exposure at Dual-Source Coronary CT Angiography.* Radiology, 2008. **248**(3): p. 792-798.

62. La Rivière, P.J., *Penalized-likelihood sinogram smoothing for low-dose CT.* Medical Physics, 2005. **32**(6Part1): p. 1676-1683.

63. Wang, J., et al., *Sinogram noise reduction for low-dose CT by statistics-based nonlinear filters*. Medical Imaging. Vol. 5747. 2005: SPIE.

64. Zhou, S., et al., *General simultaneous motion estimation and image reconstruction (G-SMEIR).* Biomedical Physics &amp; Engineering Express, 2021. **7**(5): p. 055011.

65. Zhao, C., et al. *Modified simultaneous motion estimation and image reconstruction (m-SMEIR) for 4D-CBCT*. in *2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018)*. 2018.

66. Tian, Z., et al., *Low-dose CT reconstruction via edge-preserving total variation regularization.* Physics in Medicine & Biology, 2011. **56**(18): p. 5949.

67. Wu, D., et al., *Iterative Low-Dose CT Reconstruction With Priors Trained by Artificial Neural Network.* IEEE Transactions on Medical Imaging, 2017. **36**(12): p. 2479-2486.

68. He, J., et al., *Optimizing a Parameterized Plug-and-Play ADMM for Iterative Low-Dose CT Reconstruction.* IEEE Transactions on Medical Imaging, 2019. **38**(2): p. 371-382.

69. Wu, D., K. Kim, and Q. Li, *Low-dose CT reconstruction with Noise2Noise network and testing-time fine-tuning.* Medical Physics, 2021. **48**(12): p. 7657-7672.

70. Chen, H., et al., *Low-Dose CT With a Residual Encoder-Decoder Convolutional Neural Network.* IEEE Transactions on Medical Imaging, 2017. **36**(12): p. 2524-2535.

71. Kang, E., J. Min, and J.C. Ye, *A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction.* Medical Physics, 2017. **44**(10): p. e360-e375.

72. Wolterink, J.M., et al., *Generative Adversarial Networks for Noise Reduction in Low-Dose CT.* IEEE Transactions on Medical Imaging, 2017. **36**(12): p. 2536-2545.

73. Kang, E., et al., *Cycle-consistent adversarial denoising network for multiphase coronary CT angiography.* Medical Physics, 2019. **46**(2): p. 550-562.

74. You, C., et al., *CT Super-Resolution GAN Constrained by the Identical, Residual, and Cycle Learning Ensemble (GAN-CIRCLE).* IEEE Transactions on Medical Imaging, 2020. **39**(1): p. 188-203.

75. Li, Z., et al., *Investigation of Low-Dose CT Image Denoising Using Unpaired Deep Learning Methods.* IEEE Transactions on Radiation and Plasma Medical Sciences, 2021. **5**(2): p. 224-234.

76. Gu, J., et al., *CycleGAN denoising of extreme low-dose cardiac CT using wavelet-assisted noise disentanglement.* Medical Image Analysis, 2021. **74**: p. 102209.

77. Bansal, A., et al. *Recycle-gan: Unsupervised video retargeting.* in *Proceedings of the European conference on computer vision (ECCV).* 2018.

78. Zhu, J., et al. *Unpaired Image-to-Image Translation Using Cycle-Consistent Adversarial Networks.* in *2017 IEEE International Conference on Computer Vision (ICCV).* 2017.

79. Goodfellow, I.J., et al., *Generative adversarial nets*, in *Proceedings of the 27th International Conference on Neural Information Processing Systems - Volume 2.* 2014, MIT Press: Montreal, Canada. p. 2672–2680.

80. Segars, W.P., et al., *4D XCAT phantom for multimodality imaging research.* Medical Physics, 2010. **37**(9): p. 4902-4915.

81. Huynh-Thu, Q. and M. Ghanbari, *Scope of validity of PSNR in image/video quality assessment.* Electronics letters, 2008. **44**(13): p. 800-801.

82. Horé, A. and D. Ziou. *Image Quality Metrics: PSNR vs. SSIM.* in *2010 20th International Conference on Pattern Recognition.* 2010.

83. Zhou, W., et al., *Image quality assessment: from error visibility to structural similarity.* IEEE Transactions on Image Processing, 2004. **13**(4): p. 600-612.

84. Greenspan, H., *Super-Resolution in Medical Imaging.* The Computer Journal, 2008. **52**(1): p. 43-63.

85. Hassan, A., S.A. Nazir, and H. Alkadhi, *Technical challenges of coronary CT angiography: Today and tomorrow.* European Journal of Radiology, 2011. **79**(2): p. 161-171.

86. Smith, E.A., et al., *Model-based iterative reconstruction: effect on patient radiation dose and image quality in pediatric body CT.* Radiology, 2014. **270**(2): p. 526-34.

87. Yasaka, K., et al., *High-resolution CT with new model-based iterative reconstruction with resolution preference algorithm in evaluations of lung nodules: Comparison with conventional model-based iterative*

*reconstruction and adaptive statistical iterative reconstruction.* European Journal of Radiology, 2016. **85**(3): p. 599-606.

88. Yang, J., et al., *Image Super-Resolution Via Sparse Representation.* IEEE Transactions on Image Processing, 2010. **19**(11): p. 2861-2873.

89. Dong, W., et al., *Image Deblurring and Super-Resolution by Adaptive Sparse Domain Selection and Adaptive Regularization.* IEEE Transactions on Image Processing, 2011. **20**(7): p. 1838-1857.

90. Zhang, Y., et al. *Reconstruction of super-resolution lung 4D-CT using patch-based sparse representation.* in *2012 IEEE Conference on Computer Vision and Pattern Recognition.* 2012.

91. Yu, H., et al. *Computed tomography super-resolution using convolutional neural networks.* in *2017 IEEE International Conference on Image Processing (ICIP).* 2017.

92. Park, J., et al., *Computed tomography super-resolution using deep convolutional neural network.* Physics in Medicine & Biology, 2018. **63**(14): p. 145011.

93. Ledig, C., 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.

94. Isola, P., et al. *Image-to-Image Translation with Conditional Adversarial Networks.* in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR).* 2017.

95. Yang, F., et al. *Learning Texture Transformer Network for Image Super-Resolution.* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 2020.

96. McCollough, C., *TU-FG-207A-04: Overview of the Low Dose CT Grand Challenge.* Medical Physics, 2016. **43**(6Part35): p. 3759-3760.

97. Zhang, L., et al., *FSIM: A Feature Similarity Index for Image Quality Assessment.* IEEE Transactions on Image Processing, 2011. **20**(8): p. 2378-2386.

98. Han, G., Z. Liang, and J. You. *A fast ray-tracing technique for TCT and ECT studies.* in *1999 IEEE Nuclear Science Symposium. Conference Record. 1999 Nuclear Science Symposium and Medical Imaging Conference (Cat. No.99CH37019).* 1999.

99. Wang, J. and X. Gu, *Simultaneous motion estimation and image reconstruction (SMEIR) for 4D cone-beam CT.* Medical Physics, 2013. **40**(10): p. 101912.

100. Qiu, W., D. Titley-Peloquin, and M. Soleimani, *Blockwise conjugate gradient methods for image reconstruction in volumetric CT.* Computer Methods and Programs in Biomedicine, 2012. **108**(2): p. 669-678.

101. Dabov, K., et al., *Image denoising with block-matching and 3D filtering.* Electronic Imaging 2006. Vol. 6064. 2006: SPIE.

102. Ginart, A., et al., *Making ai forget you: Data deletion in machine learning.* Advances in Neural Information Processing Systems, 2019. **32**.

103. Jin, S., et al., *Registration of PET and CT images based on multiresolution gradient of mutual information demons algorithm for positioning esophageal cancer patients.* Journal of applied clinical medical physics, 2013. **14**(1): p. 50-61.

104. Vercauteren, T., et al., *Diffeomorphic demons: Efficient non-parametric image registration.* NeuroImage, 2009. **45**(1): p. S61-S72.