# Cone Beam Computed Tomography Image Quality Augmentation using Novel Deep Learning Networks

by

Yao Zhao

Duke Kunshan and Duke University
Medical Physics Graduate Program

Date: _____
Approved:

_____
Lei Ren, Supervisor

_____
James Bowsher

_____
Zheng (Jim) Chang

Thesis submitted in partial fulfillment of
the requirements for the degree of
Master of Science in the Graduate Program in
Medical Physics in the Graduate School
of Duke University

2019

ABSTRACT

Cone Beam Computed Tomography Image Quality Augmentation using Novel Deep Learning Networks

by

Yao Zhao

Duke Kunshan and Duke University
Medical Physics Graduate Program

Date: _____
Approved:

_____
Lei Ren, Supervisor

_____
James Bowsher

_____
Zheng (Jim) Chang

An abstract of a thesis submitted in partial
fulfillment of the requirements for the degree of
Master of Science in the Graduate Program in
Medical Physics in the Graduate School
Duke University

2019

# Abstract

**Purpose**: Cone beam computed tomography (CBCT) plays an important role in image guidance for interventional radiology and radiation therapy by providing 3D volumetric images of the patient. However, CBCT suffers from relatively low image quality with severe image artifacts due to the nature of the image acquisition and reconstruction process. This work investigated the feasibility of using deep learning networks to substantially augment the image quality of CBCT by learning a direct mapping from the original CBCT images to their corresponding ground truth CT images. The possibility of using deep learning for scatter correction in CBCT projections was also investigated.

**Methods**: Two deep learning networks, i.e. a symmetric residual convolutional neural network (SR-CNN) and a U-net convolutional network, were trained to use the input CBCT images to produce high-quality CBCT images that match with the corresponding ground truth CT images. Both clinical and Monte Carlo simulated datasets were included for model training. In order to eliminate the misalignments between CBCT and the corresponding CT, rigid registration was applied to clinical database. The binary masks achieved by Otsu auto-thresholding method were applied to for Monte Carlo simulate data to avoid the negative impact of non-anatomical structures on images. After model training, a new set of CBCT images were fed into the trained

network to obtain augmented CBCT images, and the performances were evaluated and compared both qualitatively and quantitatively. The augmented CBCT images were quantitatively compared to CT using the peak-signal-to-noise ratio (PSNR) and the structural similarity index measure (SSIM).

Regarding the study for using deep learning for the scatter correction in CBCT, the scatter signal for each projection was acquired by Monte Carlo simulation. U-net model was trained to predict the scatter signals based on the original CBCT projections. Then the predicted scatter components were subtracted from the original CBCT projections to obtain scatter-corrected projections. CBCT image reconstructed by the scatter-corrected projections were quantitatively compared with that reconstructed by original projections.

**Results**: The augmented CBCT images by both SR-CNN and U-net models showed substantial improvement in image quality. Compared to original CBCT, the augmented CBCT images also achieve much higher PSNR and SSIM in quantitative evaluation. U-net demonstrated better performance than SR-CNN in quantitative evaluation and computational speed for CBCT image quality augmentation.

With the scatter correction in CBCT projections predicted by U-net, the scatter-corrected CBCT images demonstrated substantial improvement of the image contrast and anatomical details compared to the original CBCT images.

**Conclusion**: The proposed deep learning models can effectively augment CBCT image quality by correcting artifacts and reducing scatter. Given their relatively fast computational speeds and great performance, they can potentially become valuable tools to substantially enhance the quality of CBCT to improve its precision for target localization and adaptive radiotherapy.

# Contents

# List of Tables

# List of Figures

# Acknowledgements

I would like to thank my thesis advisor Dr. Lei Ren of the department of radiation oncology at Duke University, for his generous guidance and encouragement in all steps of this project. In addition, I want to thank the medical physics graduate program at Duke Kunshan University. It provided me with the great opportunity to get involved in the advanced research. I am grateful to my committee members, Dr. Zheng (Jim) Chang and Dr. James Bowsher, for being on my thesis committee and for providing valuable suggestions to this project. Moreover, I also want to thank all the staff and my classmates, for their help with my project and encouragements during my research.

# 1. Introduction

## *1.1 Introduction of Computed Tomography*

Computed Tomography (CT) is an imaging modality which utilizes x-ray projection radiographs from several view angles to generate 3D images.

### 1.1.1 Development of Computed Tomography

In 1972, the first head CT scanner was developed as a non-invasive imaging technology to view the interior of patients[1]. Since then, CT has been widely used for diagnostic, dose calculation, and image guidance purpose in clinical application.

A CT scanner is required to produce one-dimensional transmission measurements between an x-ray source and a detector. Seven generations of CT have been developed so far. The first generation of CT scanner utilized pencil beam systems with a single x-ray detector, requiring the constant distance between the x-ray tube and detectors. The transmission measurements were acquired by translating the source and detectors in one direction to get one projection and then rotating them to get more projections from multiple scan angles. The scan for each slice typically took about 270 seconds, which was fairly long compared to the later generations of CT. The second generation CT subsequently replaced the first generation by increasing the size of the detector and the x-ray source beam angle. Although it efficiently improved data acquisition, the second generation CT with fan beam geometry still needed translation of the x-ray source and the detectors for transmission measurements at each scan angle.

The third generation CT was developed to avoid translation during data acquisition. Thus, a wide fan beam was utilized, allowing to complete the measurement only by rotating the tube and detectors synchronously. The fourth generation CT utilized a stationary 360-degree ring of detectors to replace the row of detectors, and a rotating x-ray tube was also included to achieve faster acquisition. The electron beam steering with the stationary x-ray tube and ring detectors were used in the fifth generation CT to accelerate the scanning process. By applying slip-ring gantry technology, the sixth generation CT is helical CT which allows continuous acquisition when the couch is moving. Then, multiple arrays of the detector in the axial dimension and cone beam were introduced in helical CT to develop the seventh generation CT, which further improves the overall performance of CT to achieve much faster 3-D scan[1].

Multiple-row detector CT (MDCT) has been widely used in radiology and radiation oncology with the advantage of fast data acquisition. The whole body scan can be achieved within 1 minute by MDCT, which is very practical for clinical application. The increase of the x-ray beam angle and the detectors' size has been one main trend of CT development recently, aiming to accelerate the scan speed. However, the application of MDCT for image guidance in radiation therapy is limited because of its high cost and large size.

### 1.1.2 Cone Beam CT (CBCT) for Image Guided Radiation Therapy

In present state-of-art, the on-board Cone Beam CT (CBCT) system has been developed for patient positioning and verification in Image-Guided Radiation Therapy (IGRT) due to its high resolution and relatively low radiation dose to patients.[2] In the on-board CBCT system, an on-board imager and an x-ray source are attached to linear accelerator (Linac) gantry. Three-dimensional images of patients can be reconstructed by the cone-beam x-ray projections acquired through a 200-degree or 360-degree gantry rotation[3]. However, CBCT images have very limited image quality due to the various image artifacts caused by the image acquisition and reconstruction process.

## *1.2 Common Artifacts in CBCT*

An imaging artifact can be defined as a structure visualized in the reconstructed image which does not actually exist in the object. [4]There are many imaging artifacts that are inherent in CBCT resulted from reconstruction using scatter contaminated and truncated CBCT projections, contributing to the degradation of image quality. Some artifacts are more pronounced in CBCT images than in CT images due to the larger field size in CBCT and different processes of data acquisition. The common imaging artifacts in CBCT are discussed below.

### 1.2.1 Noise

Noise in CBCT images is an unwanted and randomly distributed disturbance of a signal[5]. Although noise is commonly not considered as an artifact, it is an important factor which blurs the image content and reduces the contrast resolution. Two types of noise are included in the reconstructed images: electronic noise and quantum noise. The electronic noise is caused by analog electronic circuits in the detection systems. [6] The quantum noise (photon-count noise) is a variation in image density due to the statistical fluctuation of photon fluence in the projection, which is expected to follow a Poisson distribution. In the CT imaging system, the quantum noise is inversely proportional to the number of photons acquired in the detector, which is correlated to the number of photons emitted from the x-ray source as determined by the mAs of the projection acquisition. Since lower mAs is typically used in CBCT machines for dose reduction reasons, a high noise level is expected in CBCT images, which results in the erroneous attenuation coefficient values in reconstructed images when backprojecting these CBCT projections with high noise signals[7].

### 1.2.2 Scatter

Scatter is also an important factor which leads to imaging artifacts in CBCT. In the diagnostic energy, ranging from 80-120 kVp, Compton scattering is the dominant interaction between photons and matter inside patients. In Compton interaction, the photons collide with the outer shell electrons of atoms, lose some of their energy, and

are diffracted from their original path after an interaction. In the image reconstruction process, the attenuation coefficient of the material is measured based on Beer's law, accounting for the attenuation of the primary photon beam without considering the scatter components. However, in practice, the photons which are scattered inside the patients may also be detected by the detectors, resulting in higher intensity of measured photons. The integration of attenuation coefficients along a ray inside patients will be falsely decreased because of the overestimation of measured intensities which corresponds to an underestimation of absorption. The scatter signals affect CBCT machines more than conventional CT machines[8]. This is because large cone shape imaging volume is used in CBCT acquisition, which leads to large amount of scatter generated in the imaging process. Scatter can cause cupping and streak artifacts in CBCT, as well as reduce the image contrast[9].

### 1.2.3 Beam Hardening Artifacts

Beam hardening is one of the most prominent artifacts in CBCT. Since all materials attenuate low-energy x-rays more strongly than high-energy ones due to photoelectric interaction, the lower energetic photons in the polychromatic spectrum emitted by the x-ray source will be significantly absorbed when traversing patients. As a result, a polychromatic beam become richer in high-energy photons after passing through patients, resulting in a 'harder' beam. Therefore, the photons recorded by the detectors may less relatively lower energetic rays than those emitted from the x-ray source,

leading to dark streaks or bands in reconstruction. The dark streaks or bands typically occur beside dense parts such as bones due to the more massive beam hardening for the higher atomic materials. Cupping artifacts from beam hardening effect are observed when x-rays traversing through the center of an object becomes harder than those traversing through the edges[4].

### 1.2.4 Partial Volume Artifacts

This effect appears when the entire volume of the object is not covered by the detector. It is assumed that the object is completely covered by the detectors at all view angles and the attenuation of the x-rays only results from the interaction with the object when reconstructing the images from acquired data. The truncated-view artifacts can be observed in reconstructed images if this condition cannot be meted. The partial volume artifacts do not occur in conventional CTs because the entire object is always included within the field of view (FOV). However, this artifact influences CBCT images due to their limited FOV. It is because, in CBCT, some of the x-rays penetrating through the patients are not detected due to the insufficient size of the detector, which contributes to the visualization of some shading artifacts. Furthermore, it also leads to an erroneous calculation of the linear attenuation coefficients[7].

### 1.2.5 Motion Artifacts

Motion artifacts caused by patient movement during the scanning process can result in misregistration artifacts within the image. Since the acquisition time of state-of-the-art

CBCT machines is relatively long, briefly ranging between 30s and 60s, it is not uncommon to see patient movement during the scan. However, the reconstruction process does not account for the motion during scanning process since it assumes a completely stationary geometry. Consequently, image blurring or even ghost images can be observed in CBCT images.[10] Because the resolutions of CBCT images are relatively high, ranging from 0.08mm-0.4mm, even small motions can obviously degrade the imaging quality. Typically, proper patient stabilization during the scanning process is utilized to limit the movement for the patients to reduce the motion artifacts[7].

### 1.2.6 Ring Artifacts

Ring artifacts are visualized as concentric bright or dark rings centered around the isocenter of the scan in axial slices of CBCT images. This type of artifact is more commonly viewed when homogeneous objects are imaged. They are mainly resulted from defective or miscalibrated detector elements. Because of the circular trajectory and the discrete sampling process, these defective detector elements lead to the appearance of rings. Usually, remedy just requires recalibration of the detector, but occasionally the detector elements need replacement which can be costly[5].

## *1.3 Scatter Correction for CBCT*

As discussed in the previous section, scatter, noise and artifacts all play significant roles in the degradation of the CBCT images quality by degrading image contrast and causing inaccurate HU numbers. The scatter reduction and artifacts

correction methods for CBCT have been extensively researched so far. Generally, various methods are being conducted for scatter reduction and scatter correction.

To reduce scattered photons from the projection images, some attempts were made by physically using collimation, an anti-scatter grid, air gap, a bow-tie filter or changing the irradiating arrangement[11]. Collimation such as slot-scan is commonly applied in the clinic to reduce the scatter by reducing the size of the x-ray beam[12]. Increasing the air gap (object-to-detector distance) can result in fewer scattered photons acquired by the detector. On the other hand, it will sacrifice the field of view of the reconstructed CBCT images. Thus, a fixed isocenter-to-detector distance of 50cm is used in CBCT machines to achieve the balance between scatter reduction and the size of the field of view. Wiegert et al [13] quantitatively evaluated the scatter reduction by the anti-scatter grid and air gap methods. The results showed that the performance of an air gap or an anti-scatter grid is limited because they cannot completely eliminate the scattered photons.

Thus, some more effective methods were conducted by performing post-scan scatter correction. In CBCT, a lattice-shaped lead beam stop array was designed in front of the X-ray tube, resulting in a few partially blocked areas in CBCT projections[14]. Then the scatter distribution can be measured in the blocked regions on the acquired projection image, and the scatter in the conventional unblocked CBCT projections can be corrected based on this measurement. However, this method has the disadvantage of longer

imaging time and higher dose to patients because of the need for an additional CBCT scan with beam blockers to estimate the scatter component. Collimator shadow was also applied to measure the scatter component for correction[15]. In this method, a collimator was utilized to measure the scatter photons outside the field of view by interpolating from the boundary data. Apart from modifying the hardware in CBCT machines, several analytical or Monte Carlo modeling methods have also been proposed for scatter correction, which are widely used in clinical CBCT systems. In the analytical method, the direct transmission measurement is convolved with a point spread function to acquire the detected images. The direct transmission measurement is then restored by the deconvolution process, resulting in scatter correction[16]. A more effective and accurate method was also proposed by iterative calculation[17]. In this method, the deconvolutional process was repeated to achieve better scatter correction. Although the analytical or Monte Carlo methods have better performance for scatter correction than experimental methods, they require a relatively long computational process.

With the application of the scatter correction methods mentioned above, the CBCT image quality may still be further improved[18]. Many studies have been conducted to use CT images as prior information to further augment the CBCT image due to its relatively higher image quality[19]. However, since there typically exist some differences in the anatomical structures between CT and CBCT images due to acquiring at different days, this method for scatter correction may produce false structures in the CBCT

images. Thus, low spatial-frequency scatter correction was introduced by Niu T et al[19]. to solve this problem since the anatomical structures typically had high-frequency components. But some contrast deterioration and tissue alteration were still caused in scatter-corrected CBCT images by this method. Shi L et al.[20] proposed to apply sparsely sampled low-frequency components to eliminate the false structures in the CBCT images. However, there still exist the high-frequency artifacts such as streaks which cannot be corrected.

## *1.4 Introduction of the Convolutional Neural Network*

In recent years, deep learning based methods using convolutional neural networks (CNN) have shown great potential in many image related tasks. The typical architecture of a CNN is shown in Figure 1, which contains six fundamental layers in most CNNs: input layer, convolutional layer, activation layer, pooling layer, fully connected layer, and output layer. Their functions are further discussed in the following session.

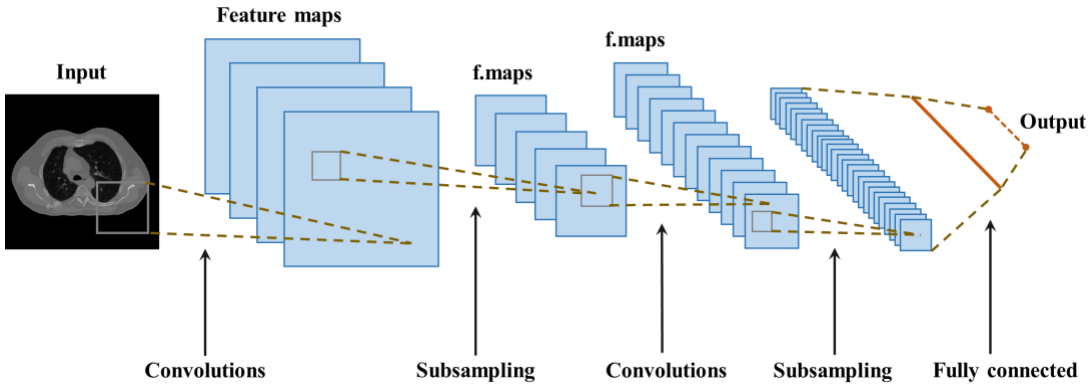



**Figure 1: The Architecture of a typical CNN.**

### 1.4.1 Input Layer

The input layer imports the input images as initial data into the network for further processing by subsequent layers. It can receive both gray level image and RGB image. For the RGB images, three channels on the third dimension are included in the input layer to cover the red, green, and blue layers in the RGB images. The input layer is the very beginning of the entire network.

### 1.4.2 Convolutional Layer

The convolutional layer is the core for a CNN, which extracts the features from the input images. In each convolutional layer, there are multiple filters (or called kernels), which are convolved with all the local receptive fields in the input to generate the feature maps. Each filter can be considered as a feature identifier, detecting a specific feature on the whole input image. Its function is demonstrated in Figure 2, the filter will scan through the input volume and convolve with all the local receptive fields, then other different filters are used to repeat the process to detect different features in the input image.

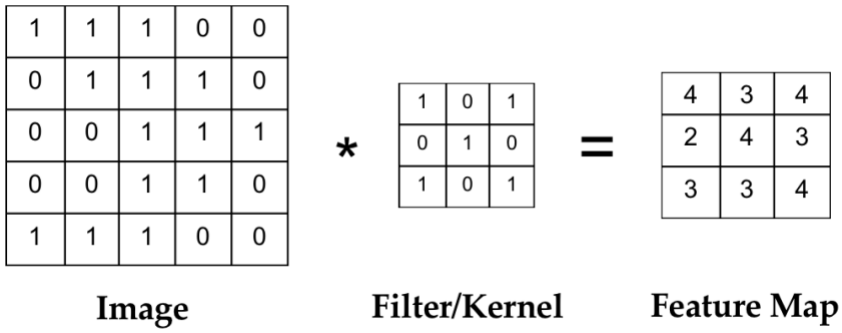



**Figure 2: The function of the convolutional layer.**

By combing the output together, the feature map is acquired, which stores the information about the different features occurring in the image. The number of channels in the feature map is determined by the number of filters as each channel records the feature extracted by a particular filter.

The size of the filters can be varied in the deep learning network. In 2012, Krizhevsky et al. proposed the filter with a size of 11 by 11 in the first layer in AlexNet[21]. Zeiler et al. used the filter with a size of 11 by 11 in ZF-Net[22]. In VGG Net proposed by Simonyan et al. in 2014[23], the filter with a size of 3 by 3 was used, which is the smallest size to contain the center, left/right, and up/down. In this study, the filter size of 3 by 3 was applied.

When the filters convolving with the input image, the filter can move through the input image one pixel by one pixel, and it can also jump several pixels at each step to convolve with the next receptive field. The stride specific the number of the pixels the filter will skip at each step, which may result in reducing the size of the output.

In order to keep the output has the same size as input in width and height dimensions, padding is usually applied. Pixels are added to the original image to enlarge the image. As shown in Figure 3, zeroes padding is most commonly used, but the padding can be any value.

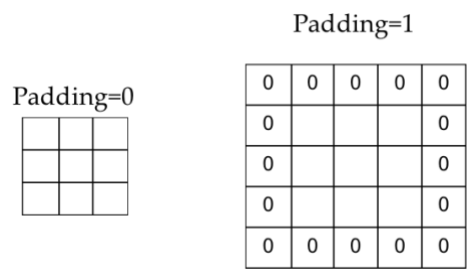



**Figure 3: The demonstration of padding with zeroes.**

### 1.4.3 Activation Layer

In CNN, activation function is applied to introduce non-linearity response to the network. It serves an important function in CNN because all the layers between the input and the output would just be the same as a single matrix without a non-linear activation function.

Sigmoid and tanh are two activation function commonly used in deep learning networks, which are defined in Equation (1) and (2). Their figure is shown in Figure 4 below.

$$\text{Sigmoid(x)} = \frac{1}{1+e^{-x}} \tag{1}$$

$$\tanh(x) = \frac{e^{x}-e^{-x}}{e^{x}+e^{-x}} \tag{2}$$

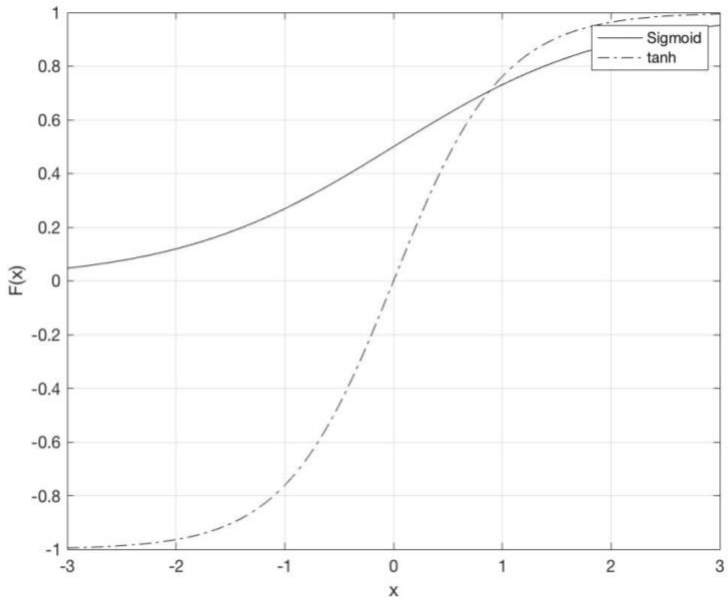



**Figure 4: Sigmoid and tanh function.**

However, if the input is either small or large, the gradients of Sigmoid and tanh are too small resulting in the difficult convergence. In 2011, a new activation function, rectified linear unit (ReLU) function was introduced by Xavier Glorot et al[24], which has a faster and better training performance in deep convolutional neural networks. The function returns 0 if the pixel value is negative and keeps the pixel value if it is positive. The ReLU function is defined in equation (3) and shown in Figure 5.

$$\mathrm{ReLU}(x) = \max(x, 0) \tag{3}$$

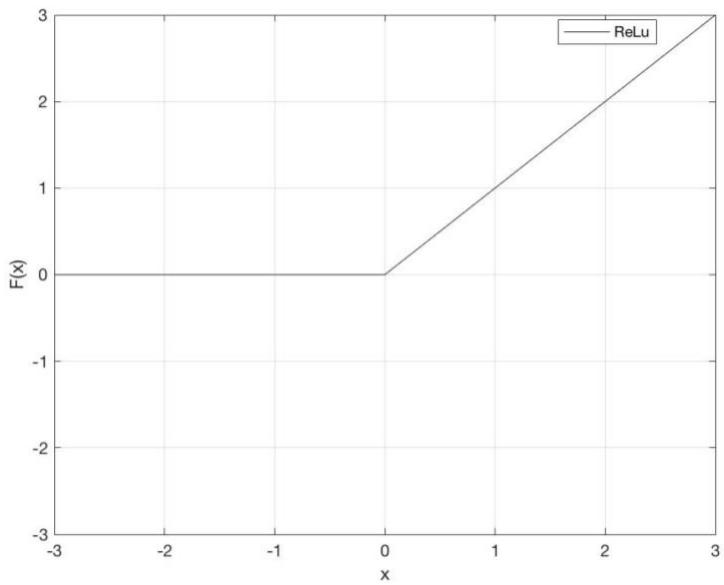



**Figure 5: ReLU function.**

Other activation functions such as Leaky ReLU were developed to further improve the performance the network[25]. In this work, PReLU was applied which is further discussed in Methods and Material section.

### 1.4.4 Pooling Layer

Pooling Layer is also important for CNNs, which reduces the size of feature maps by introducing non-linear down-sampling. This can be achieved by using filters to extract the average or maximum value in small regions. Figure 6 shows a max pooling layer, which is the most common type of pooling. And an average pooling layer is shown in Figure 7. Both filters have the size of 2 by 2 and the stride is also 2.

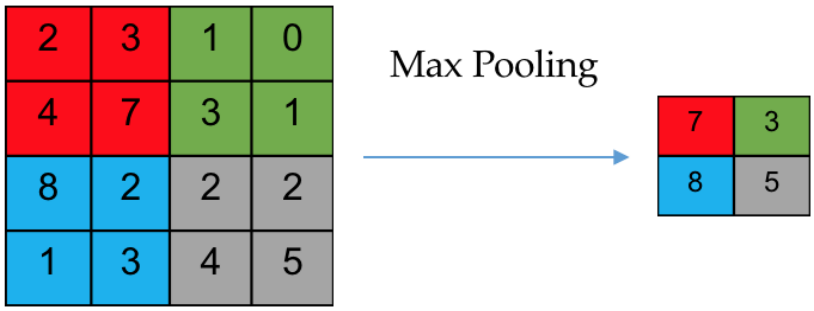



**Figure 6: Max pooling layer, with a filter size of 2 and a stride of 2.**

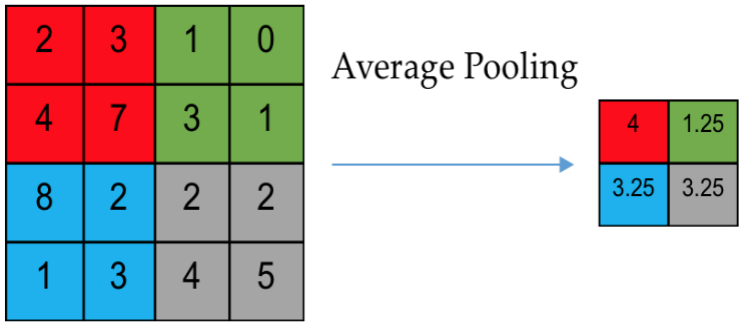



**Figure 7: Average pooling layer, with a filter size of 2 and a stride of 2.**

By applying pooling layers, the parameters in the network can be reduced and the computation load can be decreased.

### 1.4.5 Fully Connected Layer

The extracted features, acquired from several convolutional and pooling layers, can be combined with fully connected layers to the output layer, making the model end-to-end trainable. The neurons in this layer are connected to all activations in the previous layer which contains the features of input images such as shape, contrast, etc. The fully

connected layer flattens and connects the feature map to the output layer. The fully connected layer is commonly introduced in networks trained for the classification task. However, in this study, no fully connected layer is applied.

## 1.5 CNN Application in Medical Imaging

In recent years, deep learning based methods have been proposed in the field of medical image processing, and have achieved tremendous success in many image related tasks such as image segmentation[26], classification [27], and lesion detection[28], etc. For example, Milletari F et al. presented a fully convolutional neural network to achieve volumetric segmentation of MRI prostate volumes, which was fast and accurate[26]. Deep neural networks were also developed Andre Esteva et al.[27] for dermatologist-level classification of skin cancer.

Recently, convolutional neural network (CNN) has been introduced to image reconstruction. A deep cascading CNN was introduced by Schlemper et al.[29] to achieve a fast magnetic resonance image (MRI) reconstruction. Yoseob Han et al.[30] proposed a novel CNN network for sparse-view CT reconstruction. And a deep learning based algorithm was also developed by Chen et al.[31] to achieve statistical iterative CBCT reconstruction using fully-sampled projections.

Furthermore, CNN has also gained overwhelming attention in image restoration tasks. XJ Mao et al.[32] presented a fully convolutional encoding-decoding CNN for image restoration, which had considerably good performance in image denoising and

super-resolution. A residual encoder-decoder CNN (RED-CNN) was proposed by Chen et al.[33] for low dose CT restoration by learning a direct mapping from low dose CT to its corresponding normal CT.

## 1.6 A GPU Tool for Simulation of Cone Beam CT Projections

In this work a GPU tool, called gDRR, was used for Monte Carlo simulation to achieve CBCT projections[34]. Primary signal, scatter signal and noise in a projection are all included and simulate based on this package.

A tri-linear ray-racing algorithm is utilized for primary signal computation, then the Monte Carlo simulation is involved to generate the primary signal and scatter signal with noise. This package can achieve accurate and realistic simulation of CBCT projections due to the inclusion of a number of realistic features such as the detector size, the poly-energetic spectrum and the usage of bowtie filter. The calculation time to generate a typical CBCT projection takes about 30-100 seconds while using a poly-energetic spectrum.

Basically, by importing a planning CT images, this GPU tool can efficiently simulate the realistic CBCT projections at different view angles. And both primary signal, scatter signal and noise in each projection can be acquired separately, which is significant for our study on the scatter correction in CBCT projection.

## 1.7 Research Aim

Considering the disadvantages of the CBCT scatter correction methods mentioned above and advantages of deep learning in medical image processing, the aim of this project is to investigate the feasibility of using CNNs to augment CBCT image quality. Both a symmetric residual convolutional neural network (SR-CNN) and a U-net convolutional network were trained to produce high quality CBCT images by learning the correlation between the input CBCT images and the corresponding ground truth CT images. The performances of the two networks were evaluated and compared both qualitatively and quantitatively, demonstrating the effectiveness of the method in CBCT image quality improvement.

Furthermore, the deep learning method was also applied in this study for scatter correction in CBCT projections. The scatter signal for each projection was predicted based on the CNN network, which was used to correct scatter component. This was a preliminary optimization step to augment the CBCT image quality based on corrections in the CBCT projections.

# 2. Methods and Materials

## *2.1 Problem Formulation and Overall Workflow*

In this study, the CBCT image quality augmentation problem is reducible to the following formulation. Assuming that $\boldsymbol{X} \in \mathbf{R}^{M \times N}$ represent the CBCT image composed of M×N pixels, and $\boldsymbol{Y} \in \mathbf{R}^{M \times N}$ represent the corresponding ground truth image which is CT image. Then the problem can be transformed to find the relation equation $f$ between CBCT image($\boldsymbol{X}$) and CT image($\boldsymbol{Y}$):

$$\arg\min_{f} \; || f(x) - y ||_2^2 \tag{4}$$

A deep learning method is proposed to improve the CBCT image quality. A CNN is applied to learn a direct mapping from the original CBCT images to their corresponding CT images. The general workflow is shown in Figure 8 below.

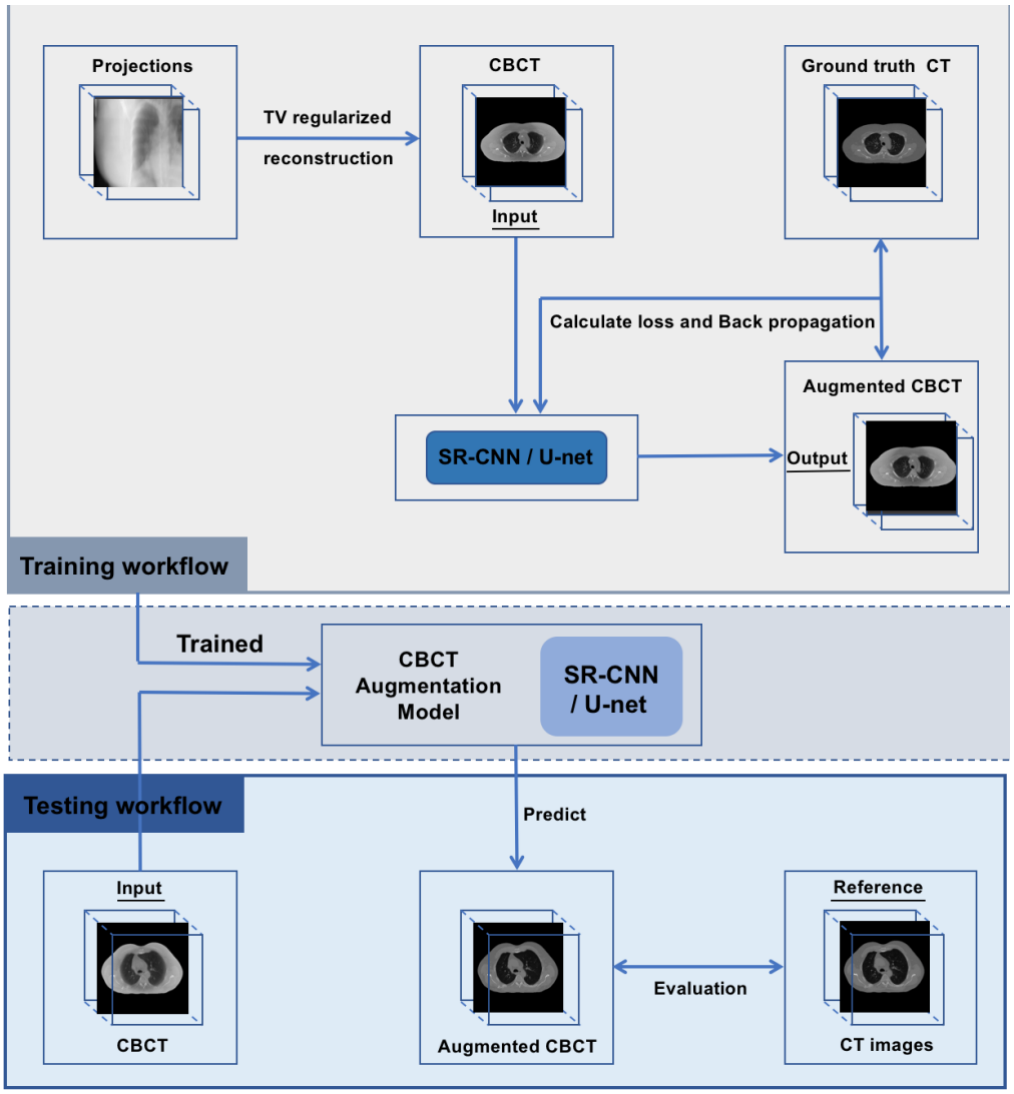



**Figure 8: General workflow of the convolution neural networks (SR-CNN and U-net). The grey box denotes training workflow. The blue box denotes testing workflow.**

In the training workflow, the CBCT images were fed into the convolutional neural network as input, and the corresponding CT images were used as the ground truth images. CBCT projections were acquired from the clinic or generated by Monte Carlo Simulation by a GPU tool developed by Jia et al.[34] CBCT images were then

reconstructed from the projections by an in-house developed iterative algorithm with TV regularization based on Sidky and Pan's method[35], which is implemented and conducted in MATLAB 2018a. Both a symmetric residual convolutional neural network (SR-CNN) and a U-net convolutional network were applied in this study. The details of the SR-CNN and U-net are described in section 2.2 and 2.3.

After the network was trained, testing CBCT images were imported to the testing workflow to evaluate the capability of the trained network to augment CBCT image quality. Briefly, testing CBCT images from both clinical and simulated datasets were fed into the trained network, which produced augmented CBCT images. Then the augmented CBCT images were compared to the ground truth CT images to evaluate the performance of the trained network.

## *2.2 Symmetric Residual Convolutional Neural Network (SR-CNN)*

In this study, a chain of fully-connected convolutional layers and deconvolutional layers are used in symmetry. The convolutional layers are used to extract and preserve the essential feature information of the input images as encoder, and deconvolutional layers are applied to restore the images from extracted features as decoder. Unlike the traditional CNN networks, pooling layer, which is usually used after a convolutional layer, is abandoned in SR-CNN because the it may discard some anatomical details and only preserve the main structures. In order to match the size of input and output, the

convolutional and deconvolutional layers must be arranged symmetrically, which requires the same kernel size, stride, and same number of layers and kernels.

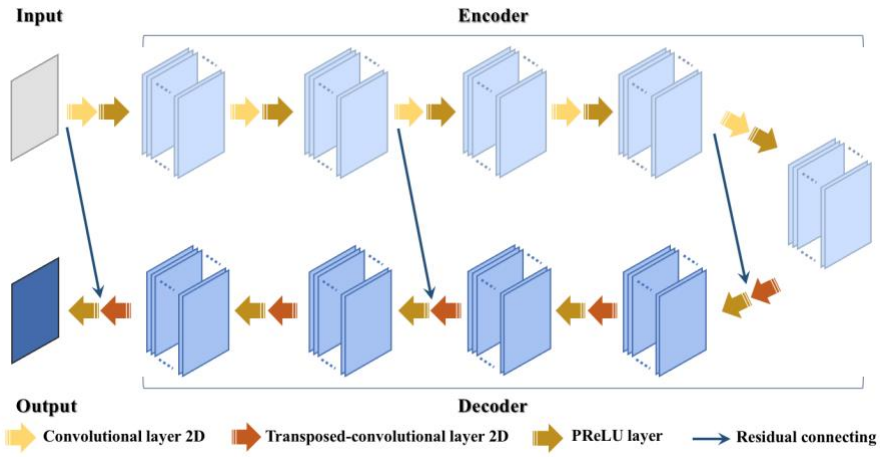



**Figure 9: Overall architecture of symmetric residual convolution neural network (SR-CNN).**

The overall architecture of SR-CNN is shown in Figure 9. Each convolutional or deconvolutional layer is followed by a parametric rectified linear unit (PReLU) layer used as an activation function for the network[36]. Apart from the symmetric convolutional and deconvolutional layers in encoder and decoder, one convolutional layer follows the decoder output to keep the dimension of the output image consistent with the ground truth images. Although deconvolutional layers can restore some structural details, the network will still lose some nonnegligible anatomical details with the increase of the depth of the network. Furthermore, since the network grew deeper, it became susceptible to the degradation problem which could cause negative effects on the network training process[37]. Therefore, residual connections were introduced between convolutional layers and deconvolutional layers with the same size in our

network. These shortcuts were used to preserve more structural details and improve the network's performance with the increase of depth. The network could be simplified to the following formulations:

For the staked encoders:

$$C_i = \text{PReLU}\ (W_{i,j} * C_{i-1} + b_{i,j})$$

$$i = 1, 2, \ldots, N, \text{ and } j = 1, 2, \ldots, M$$

For the staked decoders:

$$D_m = \begin{cases} \text{PReLU}\left(W'_{m,n} \otimes (D_{m-1} + C_{N-m+1}) + b'_{m,n}\right), & if\ with\ shortcut \\ \text{PReLU}\left(W'_{m,n} \otimes D_{m-1} + b'_{m,n}\right) & , if\ without\ shortcut \end{cases}$$

$$m = 1, 2, \ldots, \text{N}, \text{and } n = 1, 2, \ldots, \text{M}$$

For decoder output to the ground truth:

$$O = \text{PReLU}\ (W_o * TC_N + b_o)$$

PReLU:

$$PReLU(x) = \begin{cases} x, & if\ x > 0 \\ a_i x, & if\ x \leq 0 \end{cases}$$

Where $N$ represent the number of convolutional or deconvolutional layers, $M$ denotes the filter number for each convolutional or deconvolutional layers, $W$, $b$, $W'$ and $b'$are the weights and biases of each filter respectively, $C_i$ is the output of each convolutional layer combined with a PReLU layer ( except $C_0$ which is the input for the whole network), $D_m$ represents the output of each deconvolutional layer combined with a PReLU layer (except $D_0$ which is the output of the encoder), * and ⊗ denote the convolution operator and deconvolution operator respectively, and $a_i$ represents a

learnable parameter which control the slope of the negative part, $x$ is the input to the activation function.

### 2.2.1 Parameter Setting

In this study, the network was composed of 5 convolutional layers and 5 deconvolutional layers with 3 shortcut connections across the matching layers at encoder and decoder part. The filter number was 96 for each layer except for the last layer which only had one filter. The size of kernels was set to be 5 by 5, and the kernels were initialized with random Gaussian distribution with mean of 0 and 0.01 standard deviation. The stride for convolution and deconvolution was set to 1 with zero padding.

### 2.2.2 Patch extraction

A large number of training samples is required for deep learning network training. Insufficient training samples may lead to overfitting problems. However, it is difficult to accumulate large number of clinical CT and CBCT images for training. In SR-CNN study, the method of extracting overlapped patches from CBCT and corresponding CT was applied to acquire sufficient training pairs[33]. The patch size was chosen to be $50 \times 50$ with a stride of 5. Resulting from this strategy, the number of training pairs all reached $10^6$ for different datasets.

## *2.3 U-net Convolutional Neural Network*

Another convolutional network architecture applied in this study was U-net. The modified network is shown in Figure 10. This network contains two parts: a contracting

part (left part) and an expansive part (right part). The contracting part are applied to acquire and preserve the feature map of the input images, and the expansive part are used to recover the images form extracted features[38].

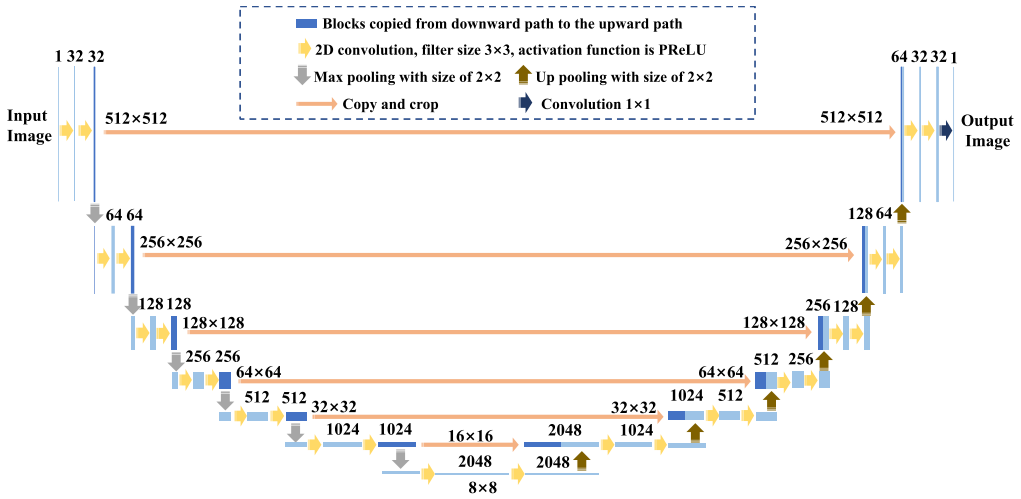



**Figure 10: The architecture of U-net convolution neural network and parameter settings.**

### 2.3.1 Parameter Setting

The architecture of the contracting part was same as a typical convolutional network, where 3×3 convolutional layer followed by a parametric rectified linear unit (PReLU) and a 2 × 2 max pooling layer with stride of 2 for downsampling were repeatedly used. The number of filters in each convolutional layer was doubled with each downsampling step. For the expansive part, 3×3 convolutional layer followed by a parametric rectified linear unit (PReLU) and a 2 × 2 up-convolutional layer with stride of 2 for upsampling was used in each step. The number of filters in each convolutional layer was reduced to

half with each upsampling step. Furthermore, concatenations were applied between matching layers at contracting part and expansive part. The concatenations were used to preserve more anatomical information since the border pixels were lost in each downsampling step which led to loss of some essential feature map. The final layer was a 1× 1 convolution which was applied to keep the dimension of the output image consistent with the ground truth images. The details of image size and the channels of feature map at each layer were shown in Figure 10.

### 2.3.2 Data Augmentation

Since it is difficult to accumulate large number of clinical CT and CBCT slices for training, data augmentation is necessary to ensure the invariance and robustness properties of the trained network. In this study, the ImageDataGenerator, an image augmentation API provided by Keras, was applied to acquire sufficient training samples. ImageDataGenerator can produce batches of real-time augmented image data. The training pairs (CBCT and corresponding CT images) were imported into ImageDataGenerator by specifying parameters like amount of rotation, amount of shift, direction of flip, and range for random zoom. After the image augmentation, the augmented images (CBCT and CT slices) were resized to be 512× 512 because of the usage of max pooling layer with stride of 2.

## 2.4 Network Compilation

The two networks used in our study are trained to learn a direct mapping from the imported CBCT image to its corresponding ground truth CT images. The weighting parameters ($\Theta$) in SR-CNN and U-net convolutional neural network were estimated by minimizing the loss error between the predicted high-quality CBCT images and the corresponding ground truth CT images. The training sample is composed of a set of paired images $P = \{(x_1, y_1), (x_2, y_2), \ldots, (x_M, y_M)\}$, where $x_i$ and $y_i$ are patch of CBCT and its corresponding CT patch respectively for SR-CNN. For U-net, $x_i$ and $y_i$ are slice of CBCT and its corresponding CT slice respectively. $M$ is the total number of training samples, and structural dissimilarity (DSSIM) was utilized as the loss function in this study which is defined as follows:

$$F(D; \Theta) = \frac{1}{N}\sum_{i=1}^{N}\frac{1}{2}(1 - SSIM_i)$$

$$SSIM = \frac{(2\mu_x\mu_y + C1)(2\sigma_{xy} + C2)}{(\mu_x^2 + \mu_y^2 + C1)(\sigma_x^2 + \sigma_y^2 + C2)}$$

Where $SSIM$ represents structural similarity index measure[39], $D$ is the training sample, $\Theta$ represents the network weighting parameters, $N$ is the batch size which represents a subset size of training pairs for each iteration of parameters adjustment, $\mu_x$ and $\mu_y$ are the average of $x$ and $y$ respectively, $\sigma_x$ and $\sigma_y$ are the covariance of $x$ and $y$ respectively, $C1$ equals to $(0.01 * MAX)^2$ ,and $C2$ equals to $(0.03 * MAX)^2$ where $MAX$ is the maximum intensity of the ground truth image. In this study, $N$ equals to 16 for SR-CNN network and 10 for U-net network due to the limitation of the GPU memory. $SSIM$ is calculated

on window sizes of 5× 5. The loss function was optimized utilizing the Adam stochastic optimization method. The learning rate was set to $10^{-4}$, and the exponential decay rates were set to 0.9 and 0.999 which are the default settings[40].

## *2.5 Data Sources and Processing*

In this study, we aimed to reduce scatter, correct artifacts and produce high-quality CBCT images by the trained SR-CNN or U-net network. In order to achieve the training purpose, ideally the anatomical geometry of the input CBCT images and the ground truth CT images should be the same so that the training will focus on the image quality augmentation. Both real clinical database from our own institution and one simulated database by Monte Carlo simulation were selected for this study. Geometrical alignment was achieved before the training of the augmentation network.

### 2.5.1 Clinical Data (Brain Patients)

A set of CBCT and planning CT images of 11 brain patients scanned in our institution was utilized in our study. Planning CT images served as the ground truth. CT and CBCT images all had a matrix size of 512 by 512 on the axial plane but with different pixel sizes, and the thickness of each slice was also different. In order to achieve the balance between computing resource and image resolution, the pixel size of all CT and CBCT images were set to 0.742mm by 0.742mm, and the thickness off each slice was set to 1mm. Because there existed some misalignments between the CBCT and planning CT

images, rigid image registration was applied to achieve three-dimensional alignments using Velocity™ oncology imaging informatics system.

For this dataset, the network was trained using a subset of 10 patients' data, and the image pairs of the rest patient was used for testing the trained network. Cross-validation was applied in this dataset to validate the robustness of the proposed models. The CBCT images of each patient were chosen for testing while the images of the other 10 patients were used for training the network.

### 2.5.2 Simulated Data (Lung Patients)

In this study, CT Images of 10 patients with lung cancers were randomly elected form the AAPM "*Sparse-view Reconstruction Challenge for Four-dimensional Cone-beam CT (4D CBCT)*" (SPARE Challenge) for network training. The provided CT images served as ground truth, and the corresponding CBCT images were reconstructed with TV regularization from 340 projections which were scanned uniformly around 360°. The projections of CBCT images were generated by Monte Carlo Simulation which were also provided by SPARE Challenge. The projection matrix size was 512 by 384 with pixel size of 0.776mm by 0.776mm, and the detector shift was set to 14.8cm for half-fan projection acquisition.

For the simulated data, since the CBCT projections were simulated from the CT images, the anatomical geometry of the CBCT matched with that of the CT without any misalignment. However, the negative impact of non-anatomical structures such as couch

on CBCT and CT images need to be avoided. Based on each paired CBCT and CT images, the binary mask was created by Otsu auto-thresholding method[41]. Then, the binary mask was applied to each CBCT and CT image to separate the patient body from the non-anatomical structures, and the voxel values outside the masked region were set to a Hounsfield Unit (HU) of -1024. The workflow of applying the binary mask to create paired CT and CBCT images is shown in Figure 12.

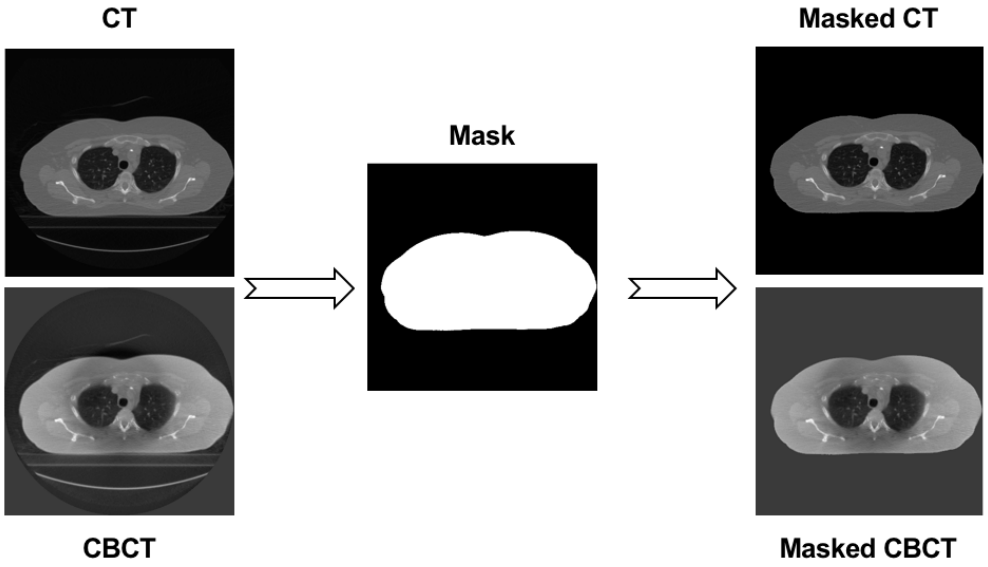



**Figure 11: Workflow to create masked CT and CBCT images.**

For simulated dataset, the CBCT of one patient was chosen for testing the trained network, and the image pairs of the rest patients were used for training.

## *2.6 Evaluation Metrics*

The augmented CBCT images predicted by SR-CNN or U-net network were qualitatively and quantitatively evaluated by comparing to the ground truth CT images.

Visual analysis was concentrated on the artifacts and the blurred anatomical structures, and the quantitative evaluation was taken by *SSIM* and peak-signal-to-noise-ratio (*PSNR*)[42] within the regions of interest (ROIs). *SSIM*, which is defined in section 2.4, could evaluate the similarity between two images by considering different attributes such as luminance, structure and contrast. *PSNR* could measure the reduction of noise and is defined as follows:

$$PSNR = 20.log_{10}(\frac{MAX}{\sqrt{MSE}})$$

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

Where *MAX* is the maximum pixel intensity in the ground truth CT image, *MSE* denotes Mean Squared Error, *m* and *n* represent the number of rows and columns of pixels in the image respectively, *i* and *j* are the corresponding index, *x* represents the intensity of pixel in CBCT or augmented CBCT image, and *y* represents the intensity of pixel in ground truth CT image.

## 2.7 Scatter Correction in CBCT Projections

In this study, the deep learning method was also developed to predict the amount of scatter in the cone beam projections directly for scatter correction in CBCT. As introduced in the section 1-6, by using the GPU tool developed by Jia et al.[34], the CBCT projections and corresponding scatter signals for each projection were acquired. In the training workflow, the CBCT projections were fed into the U-net network as

input, and the corresponding scatter signals were used as ground truth images. After the network was trained, the testing CBCT projections were imported to the workflow to evaluate the capability of the trained network to predict the scatter signals for each projection. Then the predicted scatter signals were subtracted from the CBCT projections, and CBCT images were reconstructed from the scatter-corrected projections by the TV algorithm. The scatter-corrected CBCT images were compared to the CBCT images reconstructed from the original projections to evaluate the augmentation by this deep-learning based scatter correction. Apart from visual inspection, SSIM was applied to quantitatively evaluate the performance.

The database included in this study was the CT images of 11 patients with liver cancers, which were scanned in our institution. The network was trained using a subset of 10 patients' data, and the image pairs of the rest patient was used for testing the trained network. In the simulation process, the projection matrix size was chosen as 512 by 384 with pixel size of 0.776mm by 0.776mm, matching with the actual clinical condition. The source-to-isocenter distance and source to detector distance were set to 100cm and 150cm respectively, and the detector shift was set to 9.8cm for half-fan projection acquisition. 360 projections were scanned uniformly around 360° for each patient.

# 3. Results

## *3.1 Brain Patients Study*

The results of SR-CNN and U-net augmentation for CBCT images in the brain patient study are shown in Figure 12. Four representative slices of axial, sagittal and coronal views from the testing dataset are included to evaluate the performances of SR-CNN and U-net model. In this figure, the first and second rows show two slices of axial view, and the third and fourth row are the sagittal and coronal views. Columns (a), (b), (c) and (d) are original CBCT images, augmented CBCT images by SR-CNN network, augmented CBCT images by U-net network, and the ground truth CT images, respectively. The red arrows are included in each view to indicate anatomy details for visual inspection. The yellow squares were selected as ROIs for PSNR and SSIM, which were shown on the axial views in column (d).

It can be seen that there were high image noise and streaking artifacts adjunct to bones, in which the attenuation coefficients are high, in the original CBCT images marked by the red arrows. With the augmentation of SR-CNN or U-net, these artifacts were corrected, and the noise was suppressed in the augmented CBCT images. The overall image quality of CBCT was improved by both SR-CNN and U-net model, and the image quality was comparable to that of the ground truth CT images.

The histograms of the axial slices are compared in Figure 13. In this figure, rows are the histogram of original CBCT, augmented CBCT images by SR-CNN network,

augmented CBCT images by U-net network, and the ground truth CT images, respectively. The histogram of the augmented CBCT image is much closer to that of the ground truth CT image, compared to the original CBCT image. The Hounsfield units (HU) numbers are corrected in the augmented CBCT images, which is observed from the histogram range and the position of the peaks.

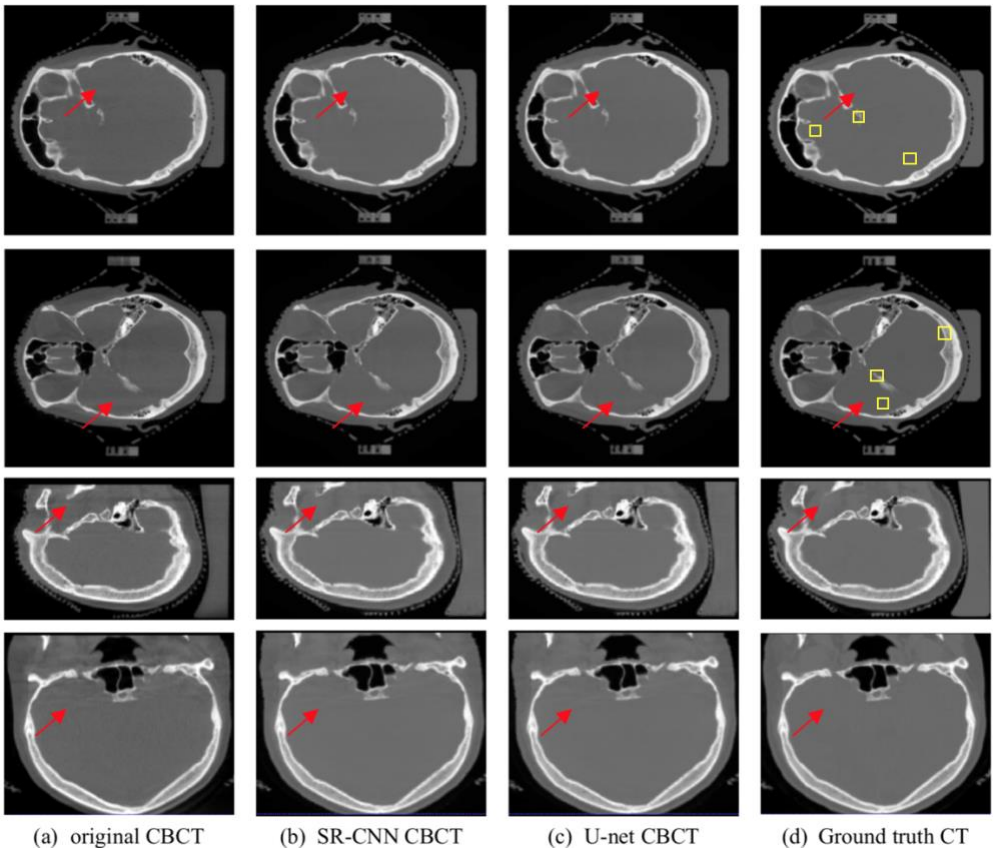



**Figure 12: Qualitative results of brain patients study.**

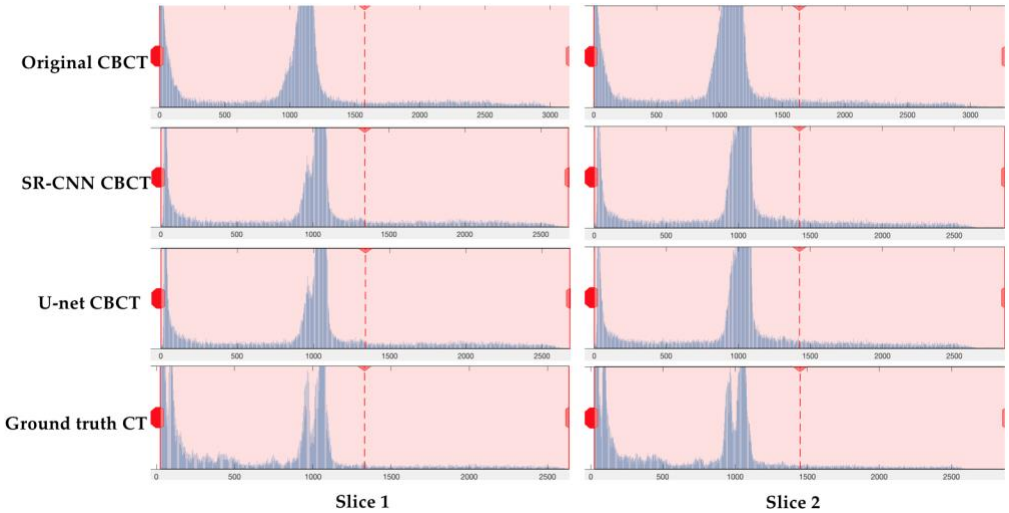



**Figure 13: Histogram comparisons of brain paitents study. The first and second columns correspond to the two axial slices shown in Firgure 12.**

The subtraction images to highlight the improvement of the CBCT image quality are shown in Figure 14. Column(a) shows original CBCT minus ground truth CT. Column(b) shows augmented CBCT images by SR-CNN minus ground truth CT. Column(c) shows augmented CBCT images by U-net minus ground truth CT. From the difference images, the differences between original CBCT image and CT images shown in column (a) are almost eliminated in the augmented CBCT images. The image quality of the augmented CBCT images was comparable to that of the ground truth CT images.

However, only from visual inspection, the augmented CBCT images by SR-CNN and U-net model are almost the same. Therefore, the comparison between the performances of SR-CNN and U-net model could not be evaluated by the qualitative results.

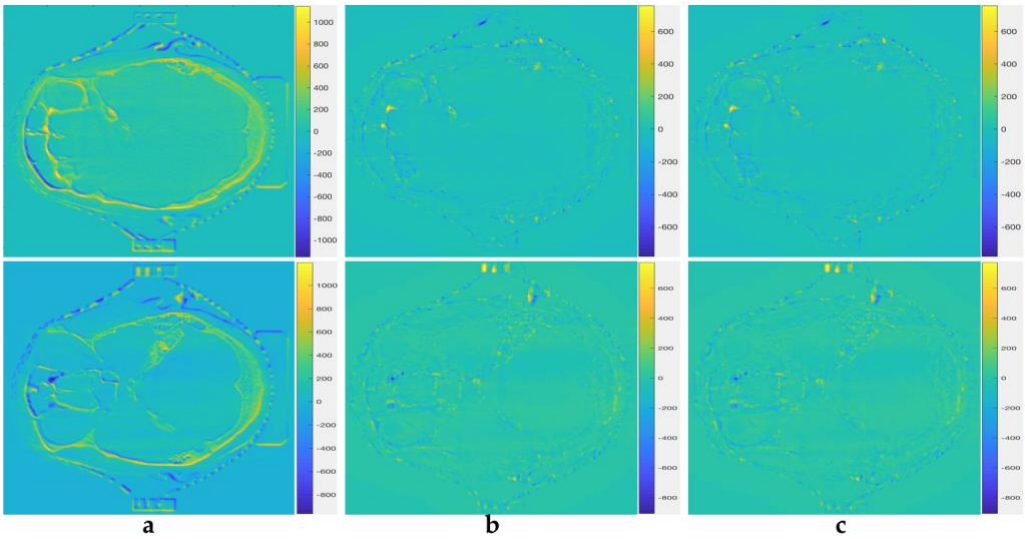



**Figure 14: Comparison of the difference images of brain patients study.**

For quantitative evaluation, Table 1 shows PSNR and SSIM of original CBCT images, augmented CBCT images by SR-CNN network, and augmented CBCT images by U-net network, measured based on the six ROIs indicated by yellow squares in Figure 12. The mean pixel value among the six ROIs are also shown in Table 1, in the HU column. Results demonstrate that the augmented CBCT images by SR-CNN or U-net always achieve higher PSNR and SSIM than the original CBCT images, especially for ROIs with artifacts and large noise. Based on the HU number, SR-CNN and U-net also considerably corrected the pixel intensity of the original CBCT images, making it close to that of the CT images. The quantitative results suggest that the noise reduction and artifacts corrections were performed effectively, and the CBCT image quality was augmented substantially close to that of CT.

Although both SR-CNN and U-net had improved original CBCT image quality, the quantitative analysis of PSNR, SSIM and HU number correction all demonstrated that U-net had a better performance.

The histograms in Figure 15 also show the comparison of PSNR and SSIM between original CBCT images and augmented CBCT images by SR-CNN and U-net model.

**Table 1: Quantitative analysis of brain patients study**

| | original CBCT | | | SR-CNN CBCT | | | U-net CBCT | | | CT |
|---|---|---|---|---|---|---|---|---|---|---|
| ROI | PSNR | SSIM | HU | PSNR | SSIM | HU | PSNR | SSIM | HU | HU |
| 1 | 31.3 | 0.927 | 109 | 40.6 | 0.973 | 27 | 42.3 | 0.975 | 26 | 24 |
| 2 | 29.5 | 0.906 | 954 | 41.2 | 0.966 | 658 | 42.6 | 0.982 | 647 | 634 |
| 3 | 27.6 | 0.901 | 126 | 38.2 | 0.959 | 32 | 40.2 | 0.970 | 29 | 27 |
| 4 | 32.3 | 0.899 | 626 | 41.6 | 0.952 | 435 | 43.1 | 0.967 | 432 | 423 |
| 5 | 33.6 | 0.933 | 136 | 42.4 | 0.979 | 33 | 43.0 | 0.984 | 33 | 32 |
| 6 | 31.5 | 0.929 | 971 | 41.0 | 0.967 | 642 | 41.5 | 0.980 | 634 | 622 |

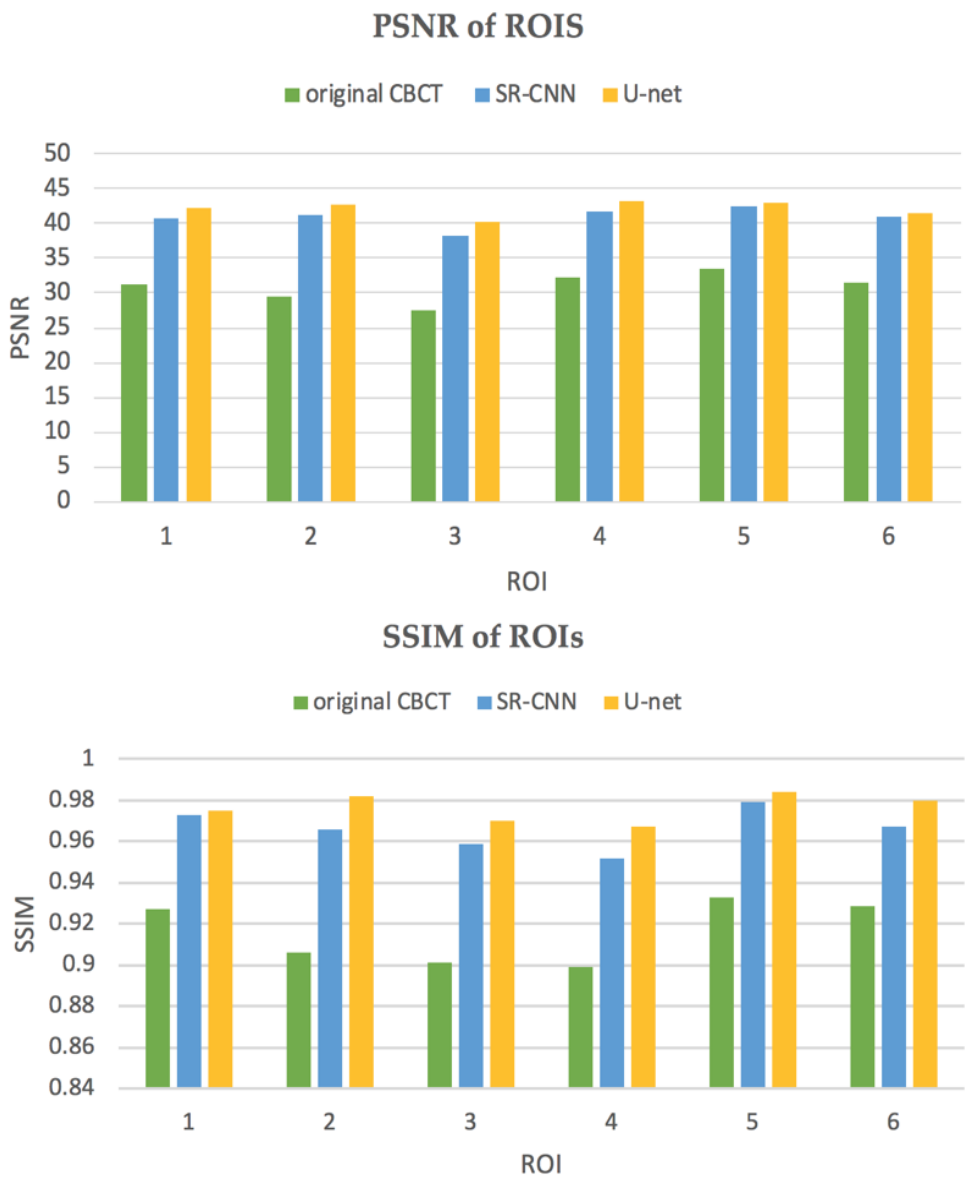



**Figure 15: Quantitative results of brain patients study. Performances of SR-CNN and U-net are shown by PSNR and SSIM in six ROIs marked in Figure 12.**

## 3.2 Lung Patients Study

The results of SR-CNN and U-net augmentation for CBCT images in the lung patient study are shown in Figure 16. Four representative slices of axial, sagittal and coronal views from the testing dataset are included to evaluate the performances of SR-CNN and U-net model. In this figure, the images in the first and second rows are two representative slices of axial view. The third and fourth row are the sagittal and coronal views. Columns (a), (b), (c) and (d) are original CBCT images, augmented CBCT images by SR-CNN network, augmented CBCT images by U-net network, and the ground truth CT images, respectively. The red arrows are included in each view to indicate anatomy details for visual inspection. The light yellow squares were selected as ROIs for PSNR and SSIM, which were shown on the axial views in column (d).

It can be seen that lung vessels, bone edges and soft tissues, highlighted by the red arrows, were blurred in the original CBCT images reconstructed with TV regularization. The streaking artifacts were evidently observed near the bone, and the overall image contrast was also relatively low due to the high scatter and noise. With the augmentation of SR-CNN or U-net, the augmented images demonstrated substantial enhancement of bone edges and lung vessels, preserving all anatomical details and suppressing most artifacts, scatter and noise. The red arrows indicate several noticeable anatomical detail enhancements in different views. The yellow circles indicate the tumor in coronal view. It's obvious that the image quality of tumor part was also improved,

and the anatomical details of tumor was preserved and enhanced. The image quality of improved CBCT images was comparable to that of the ground truth CT images.

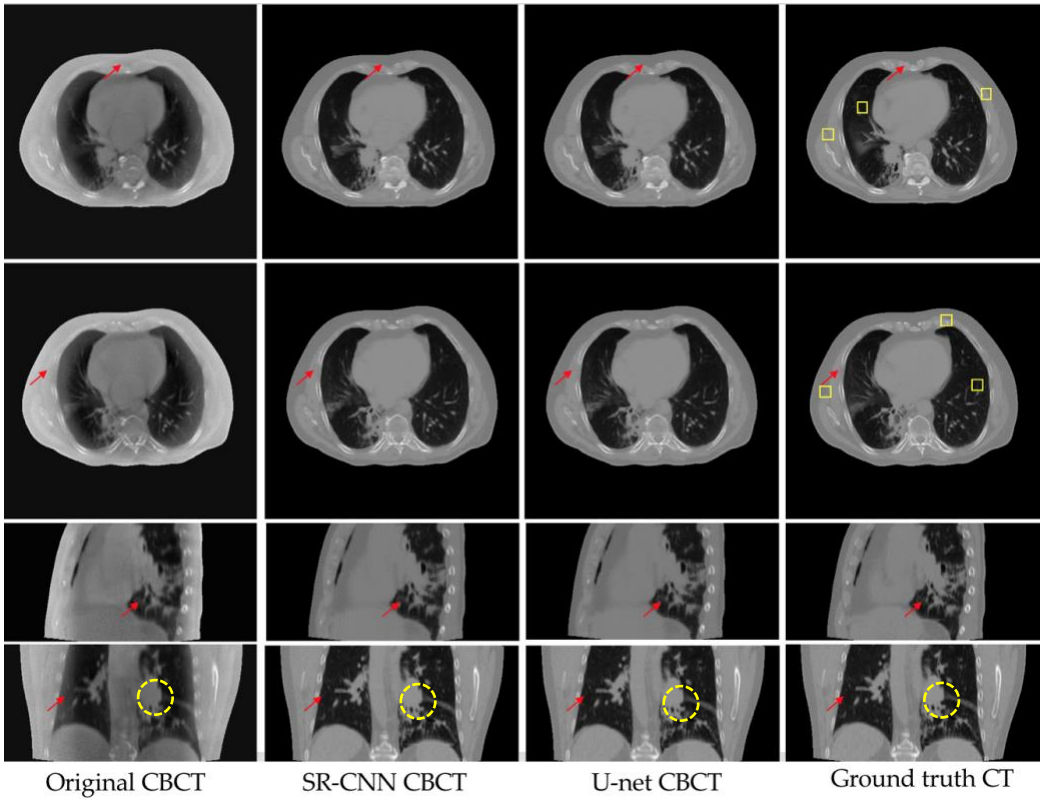



**Figure 16: Qualitative results of Monte Carlo Simulated lung patients study.**

The histograms of the axial slices are compared in Figure 17. In this figure, rows are the histogram of orginal CBCT, augmented CBCT images by SR-CNN network, augmented CBCT images by U-net network, and the ground truth CT images, respectively. The histogram of the augmented CBCT image is much closer to that of the ground truth CT image, compared to the original CBCT image. The Hounsfield units

(HU) numbers are corrected in the augmented CBCT images, which is observed from the histogram range and the position of the peaks.

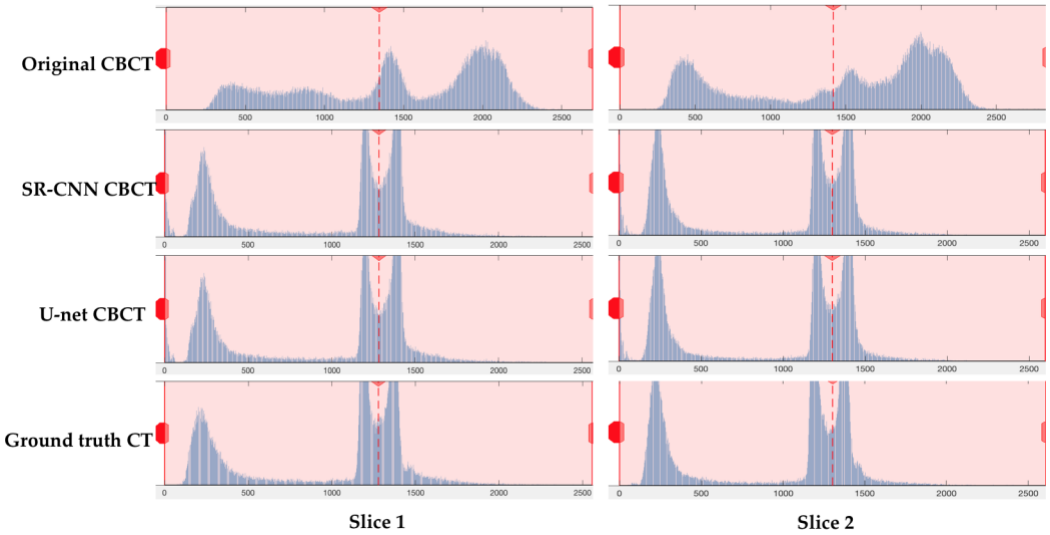



**Figure 17: Histogram comparisons of lung paitents study. The first and second columns correspond to the two axial slices shown in Firgure 16.**

The subtraction images to highlight the improvement of the CBCT image quality are shown in Figure 18. Column(a) shows original CBCT minus ground truth CT. Column(b) shows augmented CBCT images by SR-CNN minus ground truth CT. Column(c) shows augmented CBCT images by U-net minus ground truth CT.

From the difference images, the differences between original CBCT image and CT images shown in column (a) are almost eliminated in the augmented CBCT images. The image quality of the augmented CBCT images was comparable to that of the ground truth CT images.

However, from visual inspection, the augmented CBCT images by SR-CNN and U-net model are almost the same. Therefore, the comparison between the performances of SR-CNN and U-net model could not be only evaluated by the qualitative results.

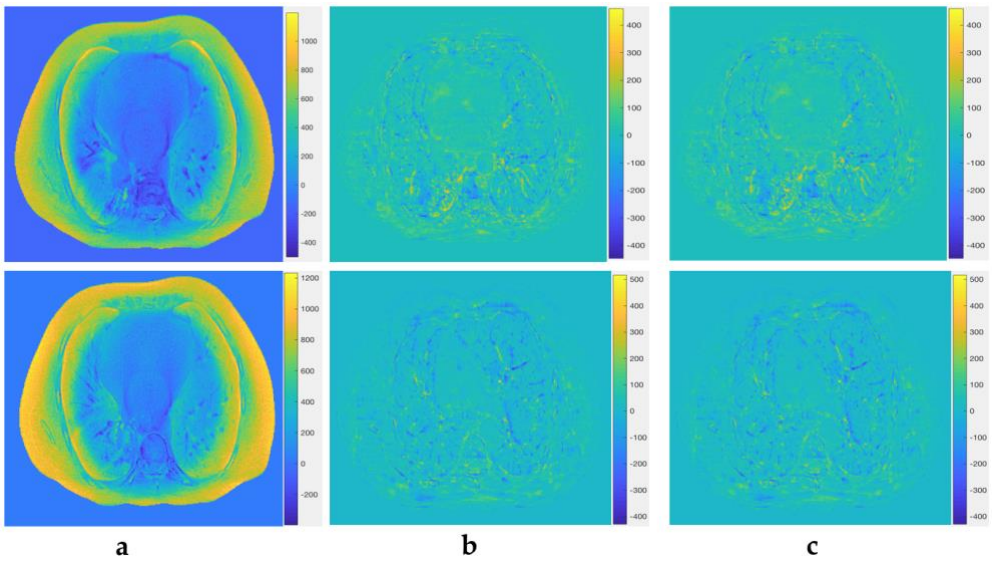



**Figure 18: Comparison of the difference images of lung patients study.**

To further show the merits of SR-CNN and U-net, the quantitative evaluation is summarized in Table 2. It includes PSNR and SSIM of original CBCT images, augmented CBCT images by SR-CNN network, and augmented CBCT images by U-net network, measured based on the six ROIs indicated by yellow squares in Figure 16. The mean pixel value among the six ROIs are also chosen for quantitative evaluation in Table 2, in the unit of Hounsfield units (HU). From the table, the augmented CBCT images were significantly better than the original CBCT images based on the PSNR and the SSIM results. The PSNR and SSIM are both higher in augmented CBCT images,

especially for scenarios with large noise. By comparing the HU number, the pixel intensities in original CBCT were also substantially corrected to that in the CT with the augmentation of SR-CNN and U-net. All these quantitative analyses shown in Table 2 demonstrated that the CBCT image quality was improved after SR-CNN or U-net enhancement for all ROIs, and high-quality CBCT images were produced.

Although both SR-CNN and U-net enhanced the original CBCT image quality, the U-net had a better performance since it achieved higher PSNR and SSIM than SR-CNN.

The histogram comparisons of the PSNR and SSIM between original CBCT images and augmented CBCT images by SR-CNN and U-net model are included in Figure 19.

**Table 2: Quantitative analysis of lung patients study**

| | original CBCT | | | SR-CNN CBCT | | | U-net CBCT | | | CT |
|---|---|---|---|---|---|---|---|---|---|---|
| ROI | PSNR | SSIM | HU | PSNR | SSIM | HU | PSNR | SSIM | HU | HU |
| 1 | 21.7 | 0.502 | -332 | 37.5 | 0.776 | -113 | 38.3 | 0.782 | -109 | -106 |
| 2 | 20.5 | 0.533 | -937 | 36.6 | 0.787 | -683 | 36.7 | 0.789 | -679 | -671 |
| 3 | 17.6 | 0.476 | 663 | 31.2 | 0.744 | 447 | 32.5 | 0.802 | 439 | 423 |
| 4 | 22.3 | 0.557 | -278 | 37.4 | 0.813 | -114 | 37.7 | 0.832 | -112 | -103 |
| 5 | 19.6 | 0.562 | -944 | 30.4 | 0.832 | -719 | 31.7 | 0.849 | -717 | -705 |
| 6 | 25.4 | 0.579 | 98 | 36.8 | 0.861 | 41 | 37.0 | 0.869 | 42 | 54 |


**Figure 19: Quantitative results of lung patients study. Performance of SR-CNN and U-net are shown by PSNR and SSIM in six ROIs marked in Figure 15.**

## 3.3 Scatter Correction in CBCT Projections

The results of scatter signals predicted by U-net in the liver patient study are shown in Figure 20. Column (a) are the CBCT projections generated by the Monte Carlo simulation, column (b) are the scatter signals of their corresponding projections predicted by U-net, and column (c) are the ground truth scatter signals, acquired in the Monte Carlo simulation process. The red dotted lines are the profiles plotted in Figure 21. From the visual inspection, the predicted scatter signals are almost the same to the ground truth scatter signals acquired in the Monte Carlo simulation process.

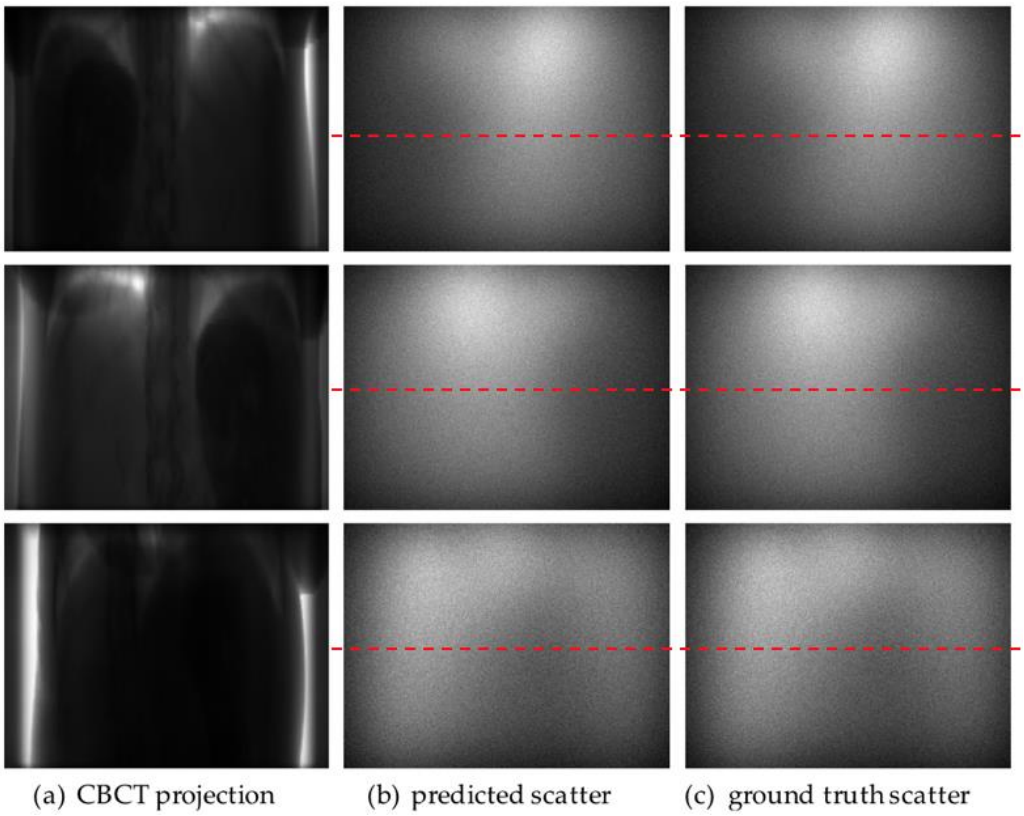



**Figure 20: Qualitative results of scatter correction of CBCT projections in liver patients study. Three representative angles chosen at 0°, 140°, 350° are included.**

Three profiles of the predicted scatter and ground truth scatter, chosen from three different projections, are compared in Figure 21. The differences between the profile of predilected scatter and that of the ground truth scatter are negligible.

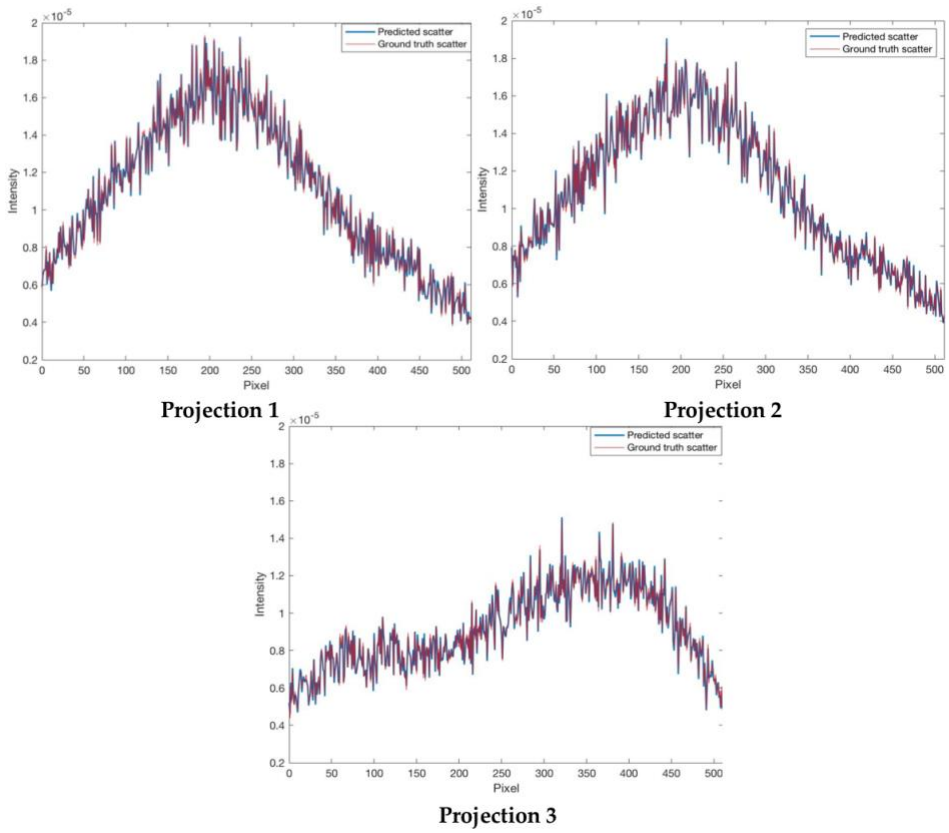
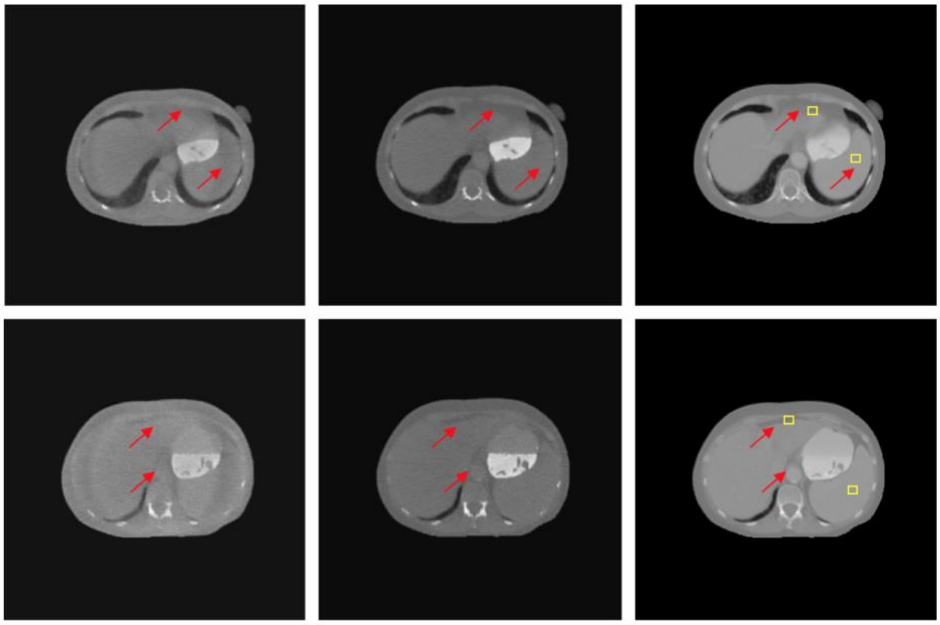



**Figure 21: Comparison of three profiles of predicted scatter and ground truth scatter, marked as red dotted lines in Figure 20.**

The predicted scatter signals were subtracted from the original CBCT projections to reduce the scatter in CBCT images. The CBCT images reconstructed by the scatter-corrected projections are compared to the CBCT images reconstructed by the original projections in the Figure 22. Column (a) are the CBCT images reconstructed from original projections, column (b) are the CBCT images reconstructed from projections

which had been corrected for scatter component, and column (c) are the CT images used for CBCT images simulation. The red arrows are included in each view to indicate anatomy details for visual inspection. The light yellow squares were selected as ROIs for SSIM calculation, which were shown on the axial views in column (c). The CT images, which were used for CBCT projections simulation, are used as reference to evaluate the scatter-corrected CBCT image quality. With the scatter correction predicted by U-net, the augmented images demonstrated substantial improvement of the image contrast and anatomical details. The red arrows indicate several noticeable anatomical structures enhancements such as artery and bone edges.



(a) original CBCT (b) scatter-corrected CBCT (c) CT

**Figure 22: Qualitative results of scatter corrected CBCT images in lung patients study. Two representative slices of axial views are shown.**

To further indicate the improvement of the scatter-corrected CBCT image quality, the quantitative evaluation is summarized in Table 3. The SSIM of CBCT images reconstructed from original projections and scatter-corrected projections were measured on the ROIs indicated by yellow squares in Figure 22 The scatter-corrected CBCT achieved higher SSIM in all selected ROIs, and the comparison are also shown in Figure 23.

**Table 3: Quantitative analysis of scatter correction**

| | original CBCT | scatter-corrected CBCT |
|---|---|---|
| ROI | SSIM | SSIM |
| 1 | 0.512 | 0.686 |
| 2 | 0.403 | 0.621 |
| 3 | 0.463 | 0.654 |
| 4 | 0.411 | 0.638 |

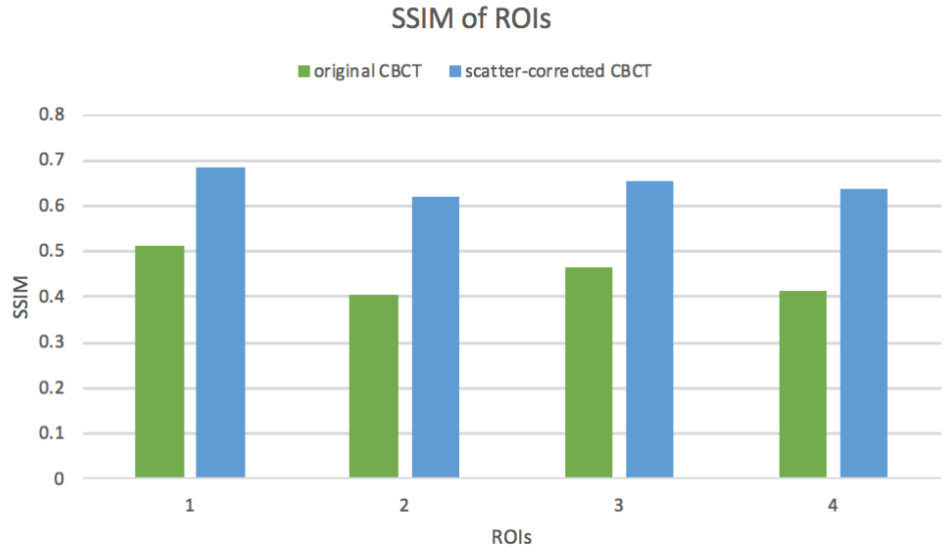



**Figure 23: Quantitative results of liver patients. Comparison of SSIM between original CBCT and scatter-corrected CBCT.**

# 4. Discussion

## *4.1 CBCT Image Quality Augmentation*

In this study, SR-CNN and U-net, two deep learning-based models, were proposed to improve the CBCT image quality, and the performances of the two models were also compared qualitatively and quantitatively. The analysis shown in the Results part demonstrates that the SR-CNN and U-net models have the capacity of learning a sophisticated restoring pattern from the training data to generate a fast and effective approach to produce high-quality CBCT images. As shown in Figure 12 and 16, both SR-CNN and U-net achieved excellent performance in correcting artifacts, suppressing noise, and enhancing anatomical details such as bone edges and lung vessels in the CBCT. Even for the tumors which are patient-specific details, the proposed networks were capable to augment the CBCT image quality in these areas. From the histogram comparisons in Figure 13 and 17, the HU numbers in the augmented CBCT images were also substantially corrected, which became closer to those in CT images. The differences images shown in Figure 14 and 18 obviously manifested that the image quality of the augmented CBCT images was comparable to that of the ground truth CT images, and was substantially improved compared to that of the original CBCT images. The quantitative evaluations of the augmentation are shown in Table 1 and 2, where the PSNR and SSIM on the improved CBCT images are higher than those on the original

CBCT images. It further demonstrated the robustness of the SR-CNN and U-net for CBCT image quality augmentation.

Although, from the visual inspection, augmented CBCT images by SR-CNN and U-net are almost the same in details preservation and image contrast improvement, U-net still has a better performance analyzed from the quantitative results in Table 1 and 2. It shows that U-net achieve higher SSIM and PSNR than SR-CNN in all ROIs, and U-net more substantially corrected the pixel intensities in the original CBCT to that in the CT. Furthermore, U-net also has faster computation time than SR-CNN in both training and testing process. In this work, all the network training and evaluation were implemented on a graphic processing unit (GPU) using Keras package, which is a deep learning library for Tensorflow and Theano. The proposed models were implemented in Python 3.7.2 on a PC (Intel Xeon CPU, 32GB RAM) using a single NVIDIA Tesla K40 GPU with 12 GB memory. For SR-CNN, the training took about 5.2 hours for each epoch, and the epoch number was empirically set to 8 based on the performance of loss function during training. Since SR-CNN was trained based on patches, the predicted CBCT images were reconstructed from predicted patches. The prediction itself took about 1.5 seconds for each slice with dimension of 450 by 450, and the prediction for each patient needs about 225 seconds. For U-net, the network training only required 22.5 hours. Once the network was trained, the augmented CBCT images were generated approximately in 20 second

for each patient which was much faster than SR-CNN, making this model more practical than SR-CNN model for clinical implementation.

In addition, it is worthy to mention that the proposed models were trained using both clinical patient data and simulated data generated by Monte Carlo simulation. It demonstrated the capability of the U-net and SR-CNN model for CBCT image quality augmentation under different scenarios, which is meaningful for the clinical application.

From the two different datasets, several limits were noticed. For the clinical data of brain patients, since CT and CBCT images were typically acquired several days or weeks apart, there usually existed some differences in the anatomical structures. Although rigid registration was applied to eliminate the misalignments between CBCT and CT images, some anatomical details' differences were still observed between registered CBCT and CT. The registration errors of the training data may result in inaccuracy of the trained model because it may be trained to make false image prediction. In Figure 12, no obvious false structures were observed on augmented CBCT images. It demonstrated that the registration errors of the training data did not cause noticeable negative effects on the accuracy of the models in brain patients study. However, for other clinical data such as liver patients, where evident anatomical differences between CBCT and CT images could be observed, this problem could become significant. One potential approach of this problem may be to improve the CT/CBCT alignment of the training data by rigid and deformable image registration

(DIR) or hybrid DIR[43]. Our future study will explore the correlation between the alignment of the training data (original CBCT and ground truth CT) and the performance of the trained models to preserve the anatomical details on augmented CBCT images.

Furthermore, both SR-CNN and U-net are post-processing models to augment CBCT image quality, which means they are incapable of creating anatomical structures out of nothing. As shown in Figure 15, some anatomical details such as lung vessels were lost in the original CBCT images due to its low image quality. Although the image quality of the augmented images was improved, and the anatomical details were enhanced, the lost anatomical structures were not recovered in the augmented CBCT images. It suggests that if the image quality of the original CBCT is extremely low, the augmentation achieved by our proposed models is limited. That is the reason for our exploration of scatter correction in CBCT projections, which are further discussed in section 4.2.

In the proposed SR-CNN and U-net models, paired patches or slices of 2D CBCT and CT were used for training and testing. In future study, multiple pairs of adjacent slices or volume patches could be applied as input and output images in the model training process, which means the imported images could be considered more comprehensively in three dimensions. It may result in more accurate augmentation for the CBCT images since more features could be extracted and preserved. Alternatively,

SR-CNN and U-net models could be trained in three different dimensions. It means the axial, sagittal and coronal slices of the 3D CBCT and CT images could be trained in the SR-CNN or U-net separately. Then the three results predicted in different dimensions may be averaged to produce the augmented CBCT.

Moreover, larger datasets including more different anatomical regions need to be included for further evaluation of the proposed tow models.

## 4.2 Scatter Correction in CBCT projections

To our best knowledge, it is the first study that investigated the feasibility of applying deep learning models for scatter correction in CBCT projections. As mentioned in section 4.1, if the image quality of input original CBCT images is extremely low, the CBCT augmentation by SR-CNN or U-net will be limited. Thus, preliminary optimization of the original CBCT by deep learning method was explored in this study. In Figure 20, the predicted scatter component in each projection was almost the same to the ground truth scatter signals which was acquired in the Monte Carlo simulation process. The differences between the profile of predilected scatter and that of the ground truth scatter, shown in Figure 21, are negligible. It further demonstrates the accuracy of the scatter signal prediction by the deep learning method.

Then, by correcting the scatter signals in the original projections, the scatter-corrected CBCT images were reconstructed from the optimized projections, as shown in Figure 22. The results showed substantial enhancement of anatomical details and image

contrast compared to the CBCT reconstructed from original projections. Although the image quality of the scatter-corrected CBCT could not be comparable to that of corresponding CT images, it was augmented substantially and could be further augmented by the SR-CNN or U-net model as discussed before.

Only U-net model was proposed for the scatter signal prediction due to its high training and predicting efficiency. The implementation was same as that in section 4.1. The network training only required 19.5 hours. Once the network was trained, the predicted scatter signals were generated approximately in 18 second for each patient. This scatter correction method is practical for clinical application due to its fast computation time. The trained U-net model could be applied to correct the scatter signals in the original projections, and further augmentation methods such as SR-CNN/U-net proposed before could be applied for the scatter-corrected CBCT images afterwards. Since the image contrast and anatomical structures' preservation in scatter-corrected CBCT images are better than the original CBCT, the further augmentation by SR-CNN or U-net model could achieve even better performance. This will be evaluated in the future study.

# 5. Conclusion

In this study, the proposed SR-CNN and U-net both demonstrated the capability of substantially augmenting the image quality of CBCT images by correcting artifacts, reducing scatter and enhancing image contrast. The networks can be directly applied to the CBCT images acquired from a clinical scanner to produce high-quality CBCT images, making them very practical for clinical application.

Besides, the scatter correction by the proposed network in CBCT projections was proved to be an effective preliminary preprocessing method for CBCT image quality augmentation.

# References


[1] M. Mahesh, "The Essential Physics of Medical Imaging, Third Edition.," *Med. Phys.*, vol. 40, no. 7, p. 077301, Jun. 2013.

[2] J.-P. Bissonnette, D. J. Moseley, and D. A. Jaffray, "A quality assurance program for image quality of cone-beam CT guidance in radiation therapy: Image quality QA of cone-beam CT," *Med. Phys.*, vol. 35, no. 5, pp. 1807–1815, Apr. 2008.

[3] T. G. Purdie *et al.*, "Cone-Beam Computed Tomography for On-Line Image Guidance of Lung Stereotactic Radiotherapy: Localization, Verification, and Intrafraction Tumor Position," *Int. J. Radiat. Oncol.*, vol. 68, no. 1, pp. 243–252, May 2007.

[4] D. R. D. Lee, "Common Image Artifacts in Cone Beam CT," p. 7.

[5] F. E. Boas and D. Fleischmann, "CT artifacts: causes and reduction techniques," *Imaging Med.*, vol. 4, no. 2, pp. 229–240, Apr. 2012.

[6] Z. Zhao, G. J. Gang, and J. H. Siewerdsen, "Noise, sampling, and the number of projections in cone-beam CT with a flat-panel detector: Noise and sampling in cone-beam CT," *Med. Phys.*, vol. 41, no. 6Part1, p. 061909, May 2014.

[7] R. Schulze *et al.*, "Artefacts in CBCT: a review," *Dentomaxillofacial Radiol.*, vol. 40, no. 5, pp. 265–273, Jul. 2011.

[8] P. G. F. Watson, "Scatter artifact correction in cone-beam CT images," p. 85.

[9] Y. Kyriakou and W. A. Kalender, "X-ray scatter data for flat-panel detector CT," *Phys. Med.*, vol. 23, no. 1, pp. 3–15, Mar. 2007.

[10] C. Nardi *et al.*, "Motion artefacts in cone beam CT: an *in vitro* study about the effects on the images," *Br. J. Radiol.*, vol. 89, no. 1058, p. 20150687, Feb. 2016.

[11] J. H. Siewerdsen and D. A. Jaffray, "Optimization of x-ray imaging geometry (with specific application to flat-panel cone-beam computed tomography)," *Med. Phys.*, vol. 27, no. 8, pp. 1903–1914, Aug. 2000.

[12] E. Samei *et al.*, "Comparative Scatter and Dose Performance of Slot-Scan and Full-Field Digital Chest Radiography Systems," *Radiology*, vol. 235, no. 3, pp. 940–949, Jun. 2005.

[13] J. Wiegert *et al.*, "Performance of standard fluoroscopy antiscatter grids in flat-detector-based cone-beam CT," presented at the Medical Imaging 2004, San Diego, CA, 2004, p. 67.

[14] J. H. Siewerdsen *et al.*, "A simple, direct method for x-ray scatter estimation and correction in digital radiography and cone-beam CT: X-ray scatter correction," *Med. Phys.*, vol. 33, no. 1, pp. 187–197, Dec. 2005.

[15] A. Akbarzadeh, M. R. Ay, H. Ghadiri, S. Sarkar, and H. Zaidi, "Measurement of scattered radiation in a volumetric 64-slice CT scanner using three experimental techniques," *Phys. Med. Biol.*, vol. 55, no. 8, pp. 2269–2280, Apr. 2010.

[16] M. Sun and J. M. Star-Lack, "Improved scatter correction using adaptive scatter kernel superposition," *Phys. Med. Biol.*, vol. 55, no. 22, pp. 6695–6720, Nov. 2010.

[17] W. Yao and K. W. Leszczynski, "An analytical approach to estimating the first order scatter in heterogeneous medium. II. A practical application: Application of an analytical approach for scatter correction," *Med. Phys.*, vol. 36, no. 7, pp. 3157–3167, Jun. 2009.

[18] J. H. Siewerdsen and D. A. Jaffray, "Cone-beam computed tomography with a flat-panel imager: Magnitude and effects of x-ray scatter," *Med. Phys.*, vol. 28, no. 2, pp. 220–231, Feb. 2001.

[19] T. Niu, M. Sun, J. Star-Lack, H. Gao, Q. Fan, and L. Zhu, "Shading correction for on-board cone-beam CT in radiation therapy using planning MDCT images: MDCT-based shading correction for CBCT," *Med. Phys.*, vol. 37, no. 10, pp. 5395–5406, Sep. 2010.

[20] L. Shi, T. Tsui, J. Wei, and L. Zhu, "Fast shading correction for cone beam CT in radiation therapy via sparse sampling on planning CT," *Med. Phys.*, vol. 44, no. 5, pp. 1796–1808, May 2017.

[21] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," *Commun. ACM*, vol. 60, no. 6, pp. 84–90, May 2017.

[22] M. D. Zeiler and R. Fergus, "Visualizing and Understanding Convolutional Networks," in *Computer Vision – ECCV 2014*, vol. 8689, D. Fleet, T. Pajdla, B. Schiele, and T. Tuytelaars, Eds. Cham: Springer International Publishing, 2014, pp. 818–833.

[23] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," *ArXiv14091556 Cs*, Sep. 2014.

[24] X. Glorot, A. Bordes, and Y. Bengio, "Deep Sparse Rectifier Neural Networks," p. 9.

[25] A. L. Maas, A. Y. Hannun, and A. Y. Ng, "Rectifier Nonlinearities Improve Neural Network Acoustic Models," p. 6.

[26] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation," *ArXiv160604797 Cs*, Jun. 2016.

[27] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115–118, Feb. 2017.

[28] R. Zhang *et al.*, "Automatic Detection and Classification of Colorectal Polyps by Transferring Low-Level CNN Features From Nonmedical Domain," *IEEE J. Biomed. Health Inform.*, vol. 21, no. 1, pp. 41–47, Jan. 2017.

[29] J. Schlemper, J. Caballero, J. V. Hajnal, A. Price, and D. Rueckert, "A Deep Cascade of Convolutional Neural Networks for MR Image Reconstruction," *ArXiv170300555 Cs*, Mar. 2017.

[30] Y. Han and J. C. Ye, "Framing U-Net via Deep Convolutional Framelets: Application to Sparse-view CT," *ArXiv170808333 Cs Stat*, Aug. 2017.

[31] B. Chen, K. Xiang, Z. Gong, J. Wang, and S. Tan, "Statistical Iterative CBCT Reconstruction Based on Neural Network," *IEEE Trans. Med. Imaging*, vol. 37, no. 6, pp. 1511–1521, Jun. 2018.

[32] X.-J. Mao, C. Shen, and Y.-B. Yang, "Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections," *ArXiv160309056 Cs*, Mar. 2016.

[33] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, and P. Liao, "Low-Dose CT with a Residual Encoder-Decoder Convolutional Neural Network (RED-CNN)," p. 12.

[34] X. Jia, H. Yan, L. Cerviño, M. Folkerts, and S. B. Jiang, "A GPU tool for efficient, accurate, and realistic simulation of cone beam CT projections: A GPU tool for CBCT projection simulation," *Med. Phys.*, vol. 39, no. 12, pp. 7368–7378, Nov. 2012.

[35] E. Y. Sidky and X. Pan, "Image reconstruction in circular cone-beam computed tomography by constrained, total-variation minimization," *Phys. Med. Biol.*, vol. 53, no. 17, pp. 4777–4807, Sep. 2008.

[36] K. He, X. Zhang, S. Ren, and J. Sun, "Delving Deep into Rectifiers: Surpassing Human-Level Performance on ImageNet Classification," *ArXiv150201852 Cs*, Feb. 2015.

[37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *ArXiv151203385 Cs*, Dec. 2015.

[38] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," *ArXiv150504597 Cs*, May 2015.

[39] Zhou Wang and A. C. Bovik, "Mean squared error: Love it or leave it? A new look at Signal Fidelity Measures," *IEEE Signal Process. Mag.*, vol. 26, no. 1, pp. 98–117, Jan. 2009.

[40] D. P. Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *ArXiv14126980 Cs*, Dec. 2014.

[41] S. L. Bangare, A. Dubal, and P. S. Bangare, "Reviewing Otsu's Method For Image Thresholding," p. 9.

[42] Q. Huynh-Thu and M. Ghanbari, "The accuracy of PSNR in predicting video quality for different video scenes and frame rates," *Telecommun. Syst.*, vol. 49, no. 1, pp. 35–48, Jan. 2012.

[43] Y. Takayama *et al.*, "Evaluation of the performance of deformable image registration between planning CT and CBCT images for the pelvic region: comparison between hybrid and intensity-based DIR," *J. Radiat. Res. (Tokyo)*, vol. 58, no. 4, pp. 567–571, Jul. 2017.