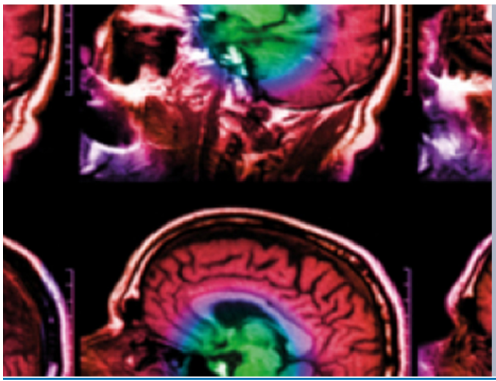
Physics in Medicine & Biology

IPEM
Institute of Physics and Engineering in Medicine

PAPER

A generalized image quality improvement strategy of cone-beam CT using multiple spectral CT labels in Pix2pix GAN

To cite this article: Yangkang Jiang *et al* 2022 *Phys. Med. Biol.* **67** 115003



IOP Publishing Phys. Med. Biol. 67 (2022) 115003 https://doi.org/10.1088/1361-6560/ac6bda

Physics in Medicine & Biology

IPEM Institute of Physics and Engineering in Medicine

CrossMark

RECEIVED
14 November 2021

REVISED
16 April 2022

ACCEPTED FOR PUBLICATION
29 April 2022

PUBLISHED
17 May 2022

PAPER

A generalized image quality improvement strategy of cone-beam CT using multiple spectral CT labels in Pix2pix GAN

Yangkang Jiang[1,2,9], **Yibao Zhang**[3,9], **Chen Luo**[2], **Pengfei Yang**[4], **Jing Wang**[5], **Xiaokun Liang**[6], **Wei Zhao**[7], **Rencang Li**[8] **and Tianye Niu**[2,5]

1 Institute of Translational Medicine, Zhejiang University, Hangzhou, Zhejiang, People's Republic of China
2 Institute of Biomedical Engineering, Shenzhen Bay Laboratory, Gaoke International Innovation Center, Guangqiao Road, Guangming District, Shenzhen, Guangdong, People's Republic of China
3 Key Laboratory of Carcinogenesis and Translational Research (Ministry of Education/Beijing), Department of Radiation Oncology, Peking University Cancer Hospital & Institute, Beijing 100142 People's Republic of China
4 School of Biomedical Engineering and Instrument Science, Zhejiang University, Hangzhou, Zhejiang, People's Republic of China
5 Sir Run Run Shaw Hospital, Zhejiang University School of Medicine, Zhejiang University, Hangzhou, Zhejiang, People's Republic of China
6 Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen, Guangdong, People's Republic of China
7 School of Physics, Beihang University, Beijing, People's Republic of China
8 Department of Mathematics, Hong Kong Baptist University, Kowloon Tong Kowloon, Hong Kong
9 These two authors contributed equally.

E-mail: niuty@szbl.ac.cn

Keywords: deep learning, multiple spectral cone-beam CT, Pix2pix GAN, scatter correction

Abstract

Objective. The quantitative and routine imaging capabilities of cone-beam CT (CBCT) are hindered from clinical applications due to the severe shading artifacts of scatter contamination. The scatter correction methods proposed in the literature only consider the anatomy of the scanned objects while disregarding the impact of incident x-ray energy spectra. The multiple-spectral model is in urgent need for CBCT scatter estimation. *Approach.* In this work, we incorporate the multiple spectral diagnostic multidetector CT labels into the pixel-to-pixel (Pix2pix) GAN to estimate accurate scatter distributions from CBCT projections acquired at various imaging volume sizes and x-ray energy spectra. The Pix2pix GAN combines the residual network as the generator and the PatchGAN as the discriminator to construct the correspondence between the scatter-contaminated projection and scatter distribution. The network architectures and loss function of Pix2pix GAN are optimized to achieve the best performance on projection-to-scatter transition. *Results.* The CBCT data of a head phantom and abdominal patients are applied to test the performance of the proposed method. The error of the corrected CBCT image using the proposed method is reduced from over 200 HU to be around 20 HU in both phantom and patient studies. The mean structural similarity index of the CT image is improved from 0.2 to around 0.9 after scatter correction using the proposed method compared with the MC-simulation method, which indicates a high similarity of the anatomy in the images before and after the proposed correction. The proposed method achieves higher accuracy of scatter estimation than using the Pix2pix GAN with the U-net generator. *Significance.* The proposed scheme is an effective solution to the multiple spectral CBCT scatter correction. The scatter-correction software using the proposed model will be available at: https://github.com/YangkangJiang/Cone-beam-CT-scatter-correction-tool.

1. Introduction

Cone-beam CT (CBCT) occupies a dominant position in the image guidance in radiation therapy and surgery. CBCT inherits the characteristics of high-definition structural imaging capability from conventional multiple detectors-based diagnostic multidetector CT (MDCT) (de la Zerda *et al* 2007). It further provides the onboard and intraoperative imaging functions to accurately supply the location and morphology of the targeted regions of interest (ROIs) during treatment (Arisan *et al* 2013). CBCT is thus ubiquitously applied for routine clinical use, for example, intraoperative patient setup (Rafferty *et al* 2006), daily or weekly patient position verification (Haworth *et al* 2009), treatment monitoring (Pouliot 2007), tumor positioning (Nickenig and Eitner 2007), etc. CBCT has been further applied in advanced clinical applications, including online target delineation (Altorjai *et al* 2012), dose accumulation and onboard treatment planning (McDermott *et al* 2008), and diagnosis decision making (Mota de Almeida *et al* 2014). CBCT can obtain the three-dimensional (3D) volumetric information efficiently with a low x-ray irradiation dose. Nevertheless, the image quality of CBCT is degraded by the shading and/or streaking artifacts due to the scatter contamination in the large cone-angle illumination procedure (Niu and Zhu 2010). When complex anatomical structures are involved in CBCT imaging, these structures significantly deteriorate the image quality due to their complicated modulation on the scatter signals (Jain *et al* 2019). These issues place restrictions on CBCT imaging for qualitative use in the clinic.

Several algorithms are developed in the literature to correct for the severe artifacts in CBCT due to scatter contamination and can be divided into two major categories: learning and non-learning-based methods according to whether learning techniques are applied in the process. Non-learning-based methods can be separated into pre- and post-processing categories according to when the scatter signal is suppressed in the measurement. The pre-processing methods prevent the scatter photons from reaching the detector which include the air gap and anti-scatter grid schemes (Sorenson and Floch 1985, Kyriakou and Kalender 2007). The pre-processing methods have the major shortcoming of increasing the x-ray dose received by the patient to maintain the appropriate signal-to-noise ratio of the projection data (Yang *et al* 2017). The post-processing techniques are more effective than pre-processing methods in that they incorporate the scatter correction at the systematic level and can be divided into measurement-based and calculation-based methods. In the measurement-based methods, the primary beam-blocker array is inserted between the x-ray source and the object to estimate the scatter samples in the blocked projection area (Niu and Zhu 2011). The primary modulation methods are developed via the low-frequency scatter extraction from the Fourier domain (Zhu *et al* 2006). These methods require hardware modification which may not be readily achievable in clinical equipment. The calculation-based methods include the Monte Carlo (MC) based methods and the analytical scatter kernel-based methods. The MC based method achieves accurate scatter estimation based on the accurate modeling of the CBCT imaging systems and simulation of the potential physical interaction between photons and matter while at the expense of heavy computational burden (He *et al* 2016). To combat the low computational efficiency, analytical scatter kernel-based methods provide a faster solution using the convolutional operations to mimic the scattering process between the primary beam and imaged object (Baer and Kachelrieß 2012, Zhao *et al* 2016). It works well for relatively uniform objects while may not function for objects with heterogeneous inner structures because of the nonlinear scattering process (Nomura *et al* 2020). The corrected CBCT images still have residual artifacts.

Learning-based methods can acquire the complicated scatter distribution by feature extraction and representation learning. The nonlinear scattering process can be approached using the deep convolutional neural network (CNN) to overcome the shortcomings of the aforementioned non-learning methods (Nomura *et al* 2019). The learning methods of scatter correction can be performed in three ways based on different working domains, i.e. direct image-to-image mapping, projection-to-projection mapping and projection-to-scatter transition. Different from scatter correction methods performed in the projection domain, the image-to-image mapping directly investigates the compensation map from the CBCT image with scatter-induced artifacts to the high-quality synthetic CT images (Harms *et al* 2019, Jiang *et al* 2019). In previous work, we proposed an image-domain scheme for CBCT shading correction using a deep residual CNN (DRCNN) combining a U-net based deep CNN and the residual of shading compensation. DRCNN trains the mapping function from the uncorrected CBCT to the high-quality CBCT image corrected using the MC simulation method. A drawback of this method is that the CBCT images in both training and testing datasets should be consistent in anatomy due to the supervised learning nature of DRCNN (Jiang *et al* 2019). To alleviate the constraint of consistent anatomy requirement in DRCNN, the cycle generative adversarial network (GAN) is integrated with residual block to investigate the mapping function from CBCT to its counterpart, e.g. the paired planning CT (Harms *et al* 2019). The GAN can obtain the high-quality samples since the parameter updating in the generator is driven by the backpropagation of the discriminator instead of the data samples (Goodfellow *et al* 2014, Zhang *et al* 2021). The cycle GAN enforces the inverse transformation and consistency to allow for higher accuracy learning (Harms *et al* 2019). Nevertheless, inconsistent anatomy training in cycle GAN may alter the anatomy of corrected CBCT images, leading to doubtful results (Harms *et al* 2019).

To minimize the anatomical structure deformation, David C Hansen *et al* proposed a U-net based network to study the transformation from uncorrected to scatter-corrected CBCT projection data (Hansen *et al* 2018). To generate the labels in the training dataset, the scatter distribution is obtained by the difference between the forward projection of the registered planning CT and the uncorrected CBCT projection. The low-pass filtration is applied to the difference to eliminate the registration error (Niu *et al* 2010). The projection domain mapping scheme alleviates the anatomy deformation issue in the image domain since the structural variation is not as significant as that in the image domain. Nevertheless, the anatomical structure deformation is difficult to be eliminated due to the small number of patients in the training dataset which cannot cover the anatomical variations found in the patient population (Hansen *et al* 2018).

To tackle the inconsistent anatomy issue, the projection-to-scatter transition is proposed to estimate the scatter signal instead of the projection data. The estimated scatter signal is subtracted from the raw projection to correct for the scatter contamination without altering the anatomical structure in the projection. For example, Yusuke Nomura *et al* applied the U-net to obtain the mapping from measured CBCT projection to the corresponding scatter distribution generated using MC simulation in nonanthropomorphic phantoms (Nomura *et al* 2019). This scheme avoids the inconsistent anatomy issue of indirect mapping methods since scatter signal is dominantly low-frequency and places little impact on the relatively high-frequency anatomical structures. In this study, we apply the projection-to-scatter transition to maintain the anatomy in the CBCT image after scatter correction.

The above methods only consider the anatomy of the scanned object to construct a scatter correction model. The x-ray energy spectrum is another major parameter affecting the estimated scatter distribution. In the research performed by Alexandr Malusek *et al* the scatter-to-primary ratio (SPR) of x-rays at 30 keV is 2.42 times more than that at 60 keV and 3.24 times more than that at 90 keV for a head-size object. For a body-size object, the SPR at 30 keV is twice more than that at 60 keV and 2.65 times more than that at 90 keV (Malusek *et al* 2003). The difference of SPRs is decreased when an anti-scatter grid is installed which is not a necessary component in all the x-ray CBCT systems with an adjustable source-to-detector distance (SID) and source-to-axis distance (SAD) including the C-arm CBCT (Orth *et al* 2008). Therefore, the multiple spectral model is in an urgent need for CBCT scatter estimation.

The energy spectra of different CBCT imaging systems vary in various clinical scenarios. Even the same device may perform different scanning protocols on the body parts of the patient. For example, the peak voltage of the x-ray source is 100 kVp in head scan and 125 kVp in abdominal scan in Varian on-board imager (OBI) system (Palm *et al* 2010). The tube voltage is 90 kVp in the dental CBCT system (Pauwels *et al* 2015) and 110 kVp in the C-arm CBCT system (Sheth *et al* 2020). It would be a very complicated and tedious task to design a model for improving CBCT image quality in each scan mode of the clinical CBCT imaging system. The multiple spectral model for CBCT image quality improvement is an innovation that can be applied as a general solution to scatter correction in different scanning modes of CBCT imaging systems.

In this study, we aim to estimate accurate scatter distributions from CBCT projections acquired at various imaging volume sizes and x-ray energy spectra to achieve the generalized image quality improvement and maintain the anatomy intact. We apply the pixel-to-pixel (Pix2pix) GAN to achieve the generalized scatter estimation since the Pix2pix GAN improves the capability of the original GAN to achieve pixel-wise image-to-image transition by the correspondence construction between the sample and label. The network architectures and loss function of Pix2pix GAN are optimized to achieve the best performance on projection-to-scatter transition. Abdominal and head patient CT data are incorporated to generate the training dataset. The trained projection to scatter mapping model is applied to the CBCT projection data of various anatomy and x-ray energy spectra to estimate the scatter distributions. Both head phantom and abdominal patient CBCT projections acquired at different energy spectra are included in the testing process.

2. Methods and materials

2.1. Concept and workflow

Any scatter estimation method that does not rely on a learning scheme can be applied to generate the training dataset such as the MC-based and the analytical scatter kernel-based methods. The MDCT data are used as one embodiment of implementing the framework in this study since scatter-induced artifacts rarely exist in the MDCT images due to its well-designed scatter suppression schemes in the past four decades (Niu *et al* 2010, Li *et al* 2019).

Figure 1 shows the workflow of the proposed scatter correction method. The part framed by the dashed line indicates the process of training dataset generation and model training. The forward projection algorithm and MC simulation are performed on the multiple spectral MDCT images to generate the scatter-free projection data and scatter signal, respectively. The scatter signal is added into the scatter-free projection to mimic the scatter-contaminated projection with a similar SPR of measured CBCT projection. The mimicked scatter-contaminated projection and scatter signal are used as the sample and label respectively in the Pix2pix GAN for the training of the scatter estimation model. The trained model is applied to the CBCT projection data to predict the scatter distribution to achieve the CBCT scatter correction.

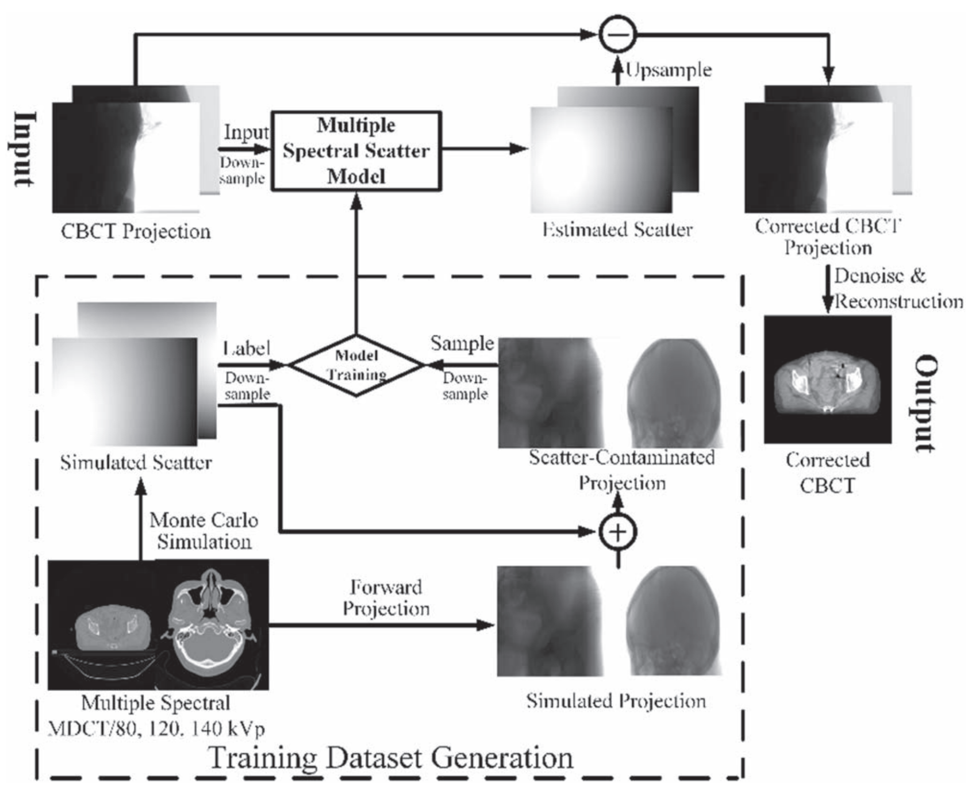


Figure 1. The workflow of the proposed scheme using the Pix2pix GAN.

2.2. Network architecture and model training

The original GAN losses the ability of user control since it cannot use the input image as a condition and learn the mapping from the input image to the output image since it generates the images from prior noise distribution (Mirza and Osindero 2014). Due to this loss of user control using the original GAN (Isola *et al* 2017), the Pix2pix GAN is applied to train the proposed model to construct the correspondence between the scatter-contaminated projection and scatter distribution. The Pix2pix GAN incorporates the data samples of the generator into the discriminator to decide the real or fake data pair (Isola *et al* 2017). The discriminator in Pix2pix GAN discriminates the real or fake estimation of a specific patch in the input image which is referred to as receptive field (RF) hereafter. As shown in figure 2(a), generator G produces the approximate scatter distribution $G(I_t)$ from the scatter-contaminated projection I_t. I_t and $G(I_t)$ are combined into the discriminator D as the data sample. The data pair of I_t and the real scatter signal label I_s are used as the data label. The discriminator acts as a dual classifier of an image to distinguish the correct data pair from the incorrect ones. The discriminated results are combined with the generator to construct the loss function.

2.2.1. Generator network

The generator G shown in figure 2(b) transforms the scatter-contaminated projection I_t into the scatter signal $G(I_t)$ to approximate the label data I_s. The symmetric encoder-decoder framework with the residual block is used to build our generator due to its well-behaved end-to-end learning capability to implement image translation (Jian *et al* 2019). The encoder part is composed of three convolutional layers and six residual blocks to balance the computation efficiency and training accuracy. The role of the convolutional layer is to extract features that contain all the textural and structural information of the input raw projection data (Krizhevsky *et al* 2017). The legend of the parameters is included in figure 2(e). The dimensions of the output image of convolutional (conv) or transposed convolutional (tconv) layer are calculated as Karpathy (2016):

$$N_{conv} = (W - K + 2P)/s + 1, \tag{1}$$

$$N_{tconv} = (W - 1) \times s - 2P + K + P_{out}, \tag{2}$$

where W is the width of the input image, K is the width of the convolution kernel, P is the number of zeros padded into one end of the input, and s is the stride of the convolutional operation. P_{out} is the number of additional zeros padded into one end of the output of the tconv operation to ensure the dimensions of the tconv output are the same as that of the conv input and is set as one in this study. n_{in} and n_{out} are the number of channels of input and output data.

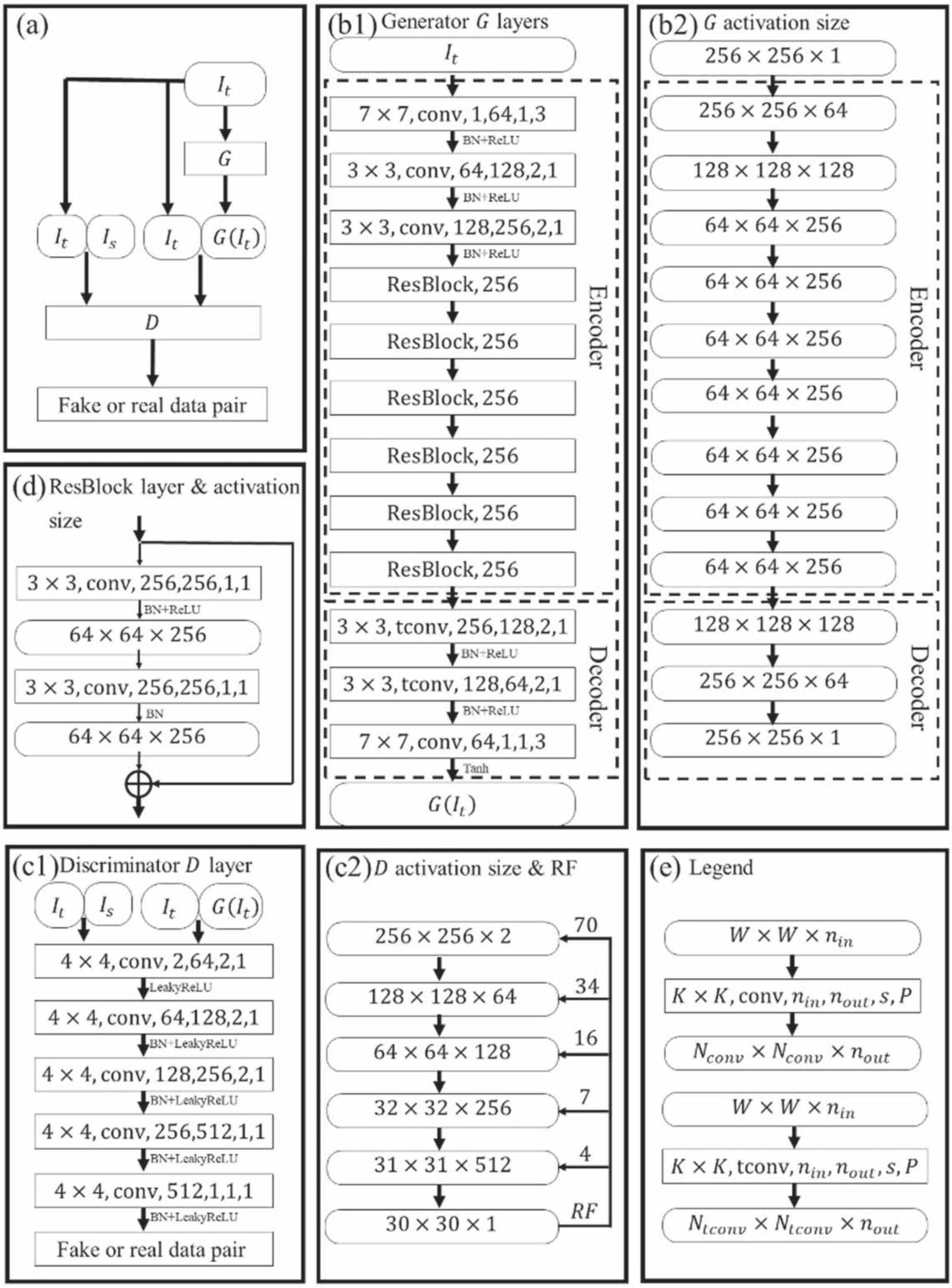


Figure 2. The architecture of the proposed Pix2pix GAN is composed of: (a) the training process of the proposed network for scatter estimation, (b1) and (c1) the generator and discriminator network architecture, and (d) the residual block used in the ResNet generator. The 'conv' and 'tconv' indicate the convolutional and transposed convolutional layer, respectively. (b2) and (c2) are the activation dimension and RF of the output of each convolutional layer for the generator and discriminator, respectively. (e) is the legend of the 'conv' and 'tconv' layers. The BN indicates the Batch Normalization and RF indicates the receptive field.

The first layer applies a 7×7 convolution kernel, the stride of one and three padded zeros, to extract a feature map with the same dimension as the input data. The second and third layers apply a 3×3 convolution kernel, the stride of two and one zero padded to one end of the input, to perform down-sampling operations and output the half-size feature maps. To improve the training accuracy without gradient vanishing and network degradation problems, six residual blocks whose structure is shown in figure 2(d) are added to the encoder part to increase the depth of the network. The gradient calculated from the output of the upper-level residual block is propagated to the input of the lower level with a shortcut connection to alleviate the gradient vanishing issue (He *et al* 2016). The two 3 × 3 convolutional layers with the stride of one are used to ensure the consistent dimension of the input and output features and achieve the highest accuracy due to the reduced complexity (Zagoruyko and Komodakis 2016). To obtain the $G(I_t)$ with the same dimension as the input I_t, the decoder part applies a symmetric transposed convolution after the six residual blocks in the encoder part. The transposed convolutional layer has the same kernel dimension and stride as the symmetrical convolutional layer to reconstruct the generated scatter signal from extracted intermediate features using an up-sampling operation. Each convolutional or transposed convolutional layer is followed by spatial batch normalization (BN) and a rectified linear activation function (ReLU) to alleviate the gradient vanishing issue and increase training stability (Zagoruyko and Komodakis 2016). The final convolutional layer applies the tanh function as the activation function since the output range of the tanh function is $(-1\ 1)$ which is readily convertible into pixel values after affine transformation (Karlik and Olgac 2011).

2.2.2. Discriminator network

Figure 2(c) shows the network architecture of the PatchGAN which is used as the discriminator in Pix2pix GAN (Isola *et al* 2017). Compared with an ordinary discriminator mapping the input into a scaler using original GAN, PatchGAN maps the input to a matrix, the entry of which represents the probability of real patch estimation in the RF (Zhao *et al* 2019). The average value of the matrix elements is defined as the final output of the discriminator, indicating the overall probability of real sample estimation. PatchGAN used in the proposed method includes five 4 × 4 convolutional layers to map the input into a 30 × 30 probability matrix. The RF dimensions of each entry of the matrix are 4 × 4, 7 × 7, 16 × 16 and 34 × 34 on the inputs from the penultimate layer to the first layer, respectively. The discriminator finally classifies whether the 70 × 70 patch in the input image is real or not.

The dimension of the RF within each layer is calculated as:

$$RF_i = (RF_{i+1} - 1) \times s_i + K_i, \tag{3}$$

where RF_i is the dimension of the RF, s_i is the stride and K_i is the kernel width of the ith convolutional layer. The dimension of the RF within the final layer is the same as the kernel of the final convolutional layer (Araujo *et al* 2019). LeakyReLU is used as the activation function in PatchGAN since the LeakyReLU is less likely to lose information than ReLU. Due to the small number of layers in the discriminator, the possible increase of network complexity using LeakyReLU may not lead to a significant computational burden (Xu *et al* 2015).

2.2.3. Label image generation

We apply the paired mimicked scatter-contaminated projections and scatter signals generated using MDCT as the sample and label in the training dataset to achieve high training performance since MDCT is of better image quality and fewer artifacts compared with CBCT.

A limited number of photons are used in the MC simulation due to the limitations of computer memory and the calculation time. The noise level is too high in the projection data directly obtained from MC simulation using the limited number of photons. Therefore, we apply Siddon's fast ray-tracing forward projection algorithm to accelerate the calculation of scatter-free line integral data (Siddon 1985). The noise-free projection is generated using the exponential transformation on the line integral data. The Poisson noise is added into the projection data to match the signal-to-noise ratio (SNR) of the reconstructed image close to that of the CBCT image. The process of generating projection data can be defined as:

$$I_p = I_0 e^{-I_l} + \lambda_n I_n, \tag{4}$$

where I_0 is the number of photons in the air scan, I_l is the line integral data generated using forward projection and I_n is the Poisson noise. λ_n is the weighting factor to tune the intensity of additive Poisson noise so that the SNR within the simulated and real CBCT images is consistent. λ_n is set as 0.9 in the head data simulation and is set as 1.1 in the abdomen data simulation.

In this study, an open-source GPU-based MC software MC-GPU (code.google.com/archive/p/mcgpu/) is applied to perform the MC simulation due to its high computational efficiency (Badal and Badano 2009). The MC-GPU code applies the PENELOPE 2006 which is a code system for MC simulation of electron and photon transport to achieve the interaction between the photons and the object and the Woodcock tracking algorithm to compute the trajectories of the photons. The object in the MC-GPU is described as a combination of voxels of different materials and mass densities. The photons treat each voxel as a uniform medium composed of the same material. In the simulation, to directly apply the MDCT images as the object into the MC-GPU program, we choose a straightforward hard-threshold segmentation to achieve the voxelization of the object and the template image composes of air, adipose, muscle and bone. The side effect of segmentation error is minimized due to the dominantly low-frequency distribution of the scatter signal in the projection domain. The spectra in the MC simulation for each MDCT dataset are the same as those in the actual scanning process. The spectra are estimated using the measurement-based spectrum estimation method proposed by Zhao *et al* (2014).

The noise appears in the simulated scatter signal and degrades the accuracy of the scatter estimation. We thus apply the two-dimensional (2D) Gaussian filtration method to smooth the simulated scatter distribution (Zbijewski and Beekman 2006). To accelerate the calculation, the logarithm of the Gaussian function is applied to convert the nonlinear to quadratic polynomial process. The 2D Gaussian function is defined as:

$$s(x, y) = (2\pi\sigma_1\sigma_2\sqrt{1-\rho^2})^{-1}\exp\left(-\frac{1}{2(1-\rho^2)}\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right)\right), \tag{5}$$

where σ_1 and σ_2 are the standard deviations of the distributions in x and y directions, μ_1 and μ_2 are the expectation of the distributions and ρ is the correlation coefficient of these two directions. After the logarithmic operation, the fitting function is converted to the following quadratic form as:

$$\begin{aligned} f(x, y) = \ln(s(x, y)) = -\ln(2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}) - \frac{1}{2(1-\rho^2)} \\ \times\left(\frac{(x-\mu_1)^2}{\sigma_1^2} - \frac{2\rho(x-\mu_1)(y-\mu_2)}{\sigma_1\sigma_2} + \frac{(y-\mu_2)^2}{\sigma_2^2}\right). \end{aligned} \tag{6}$$

These fitting parameters are calculated using the least square method. The least-squares method finds the optimal parameters in equation (6) by minimizing the sum of squared residuals between equation (6) and the scatter distribution S. The sum of squared residuals is defined as:

$$R(S, f(x, y)) = \sum_{x=1}^{X}\sum_{y=1}^{Y}(S - f(x, y))^2, \tag{7}$$

where (X, Y) is the dimension of the scatter distribution. $R(S, f(x, y))$ is minimal when the partial derivative of the $R(S, f(x, y))$ is zero. After the noise suppression, the estimated scatter signal has a high SNR while maintaining the corrected profile.

2.2.4. Model training

When the network is determined, the generator can be estimated by the optimization of a loss function. The fidelity term is generated using the output of the discriminator. The least absolute deviation (L1) term is used as the regularization item to solve the overfitting problem. The entire loss function of the Pix2pix GAN is defined as:

$$G^* = arg\ \min_G \max_D \mathcal{L}_{LSGAN}(G, D) + \lambda\mathcal{L}_{L1}(G), \tag{8}$$

where λ is the coefficient of the regularization term and set as 100.0 in this work to eliminate the overfitting while ensuring the convergence of the loss function. $\mathcal{L}_{LSGAN}(G, D)$ represents the loss function of the least-square GAN. Compared with the conventional sigmoid cross-entropy loss function used in the original GAN, the proposed method with the least-square (LS) form achieves higher-quality image generation and more stable training since it penalizes the fake samples and overcomes the gradient vanishment problem (Mao *et al* 2017). $\mathcal{L}_{LSGAN}(G, D)$ is defined as Mao *et al* (2017):

$$\mathcal{L}_{LSGAN}(G, D) = \frac{1}{M}\sum_{j=1}^{M}[(D(I_{t\,j}, I_{s\,j}))^2] + \frac{1}{M}\sum_{j=1}^{M}[(D(I_{t\,j}, G(I_{t\,j})) - 1)^2], \tag{9}$$

where M is the total number of training datasets, $D(\blacksquare)$ is the output of the discriminator. $I_{t\,j}$ and $I_{s\,j}$ are the jth scatter-contaminated projection and scatter signal, respectively.

The L1 regularization term is incorporated to constrain the difference between the generated and the real images while maintaining the boundary of the object in the image (Geng *et al* 2020):

$$\mathcal{L}_{L1}(G) = \frac{1}{M}\sum_{i=1}^{M}[I_{s\,i} - G(I_{t\,i})_1]. \tag{10}$$

The adaptive moment estimation (Adam) algorithm (Kingma and Ba 2014) is applied as the solver to optimize the loss function to determine the parameters of each convolutional layer in the G and D.

2.3. Dataset

The MDCT images of abdominal and head patients are applied to train the proposed model. The MDCT data of the abdominal patient are obtained from Philips Brilliance Big Bore CT scanner with the tube voltage of 120 kVp. Dual-energy CT data of the head patient are acquired from Siemens SOMATOM Definition Flash scanner

Table 1. Parameters of MDCT and CBCT data acquisition.

MDCT parameters	Philips brilliance big bore CT	Siemens SOMATOM definition flash
Anatomical site	Abdomen	Head
Scan mode	Helical	Helical
Scan voltage (kVp)	120	80, 140
# of patients	17	12, 12
# of images in the training dataset	18 600	
CBCT parameters	**Trilogy**	**Tabletop CBCT system**
Anatomical site	Abdomen	Head phantom
Scan mode	Half-fan	Full-fan
Scan voltage (kVp)	125	75,100,125
Bowtie filter	Yes	No
Detector shifted (mm)	148	0
Source to detector distance (mm)	1500	1300
Source to axis distance (mm)	1000	1000
Detector dimension (pixels)	1024 × 768	2816 × 2816
Pixel size (mm)	0.388	0.154
# of patients for test	20	1

using 80 and 140 kVp tube voltages. The training dataset is composed of 15 sets of 120 kVp abdominal images, ten sets of 80 kVp and ten sets of 140 kVp head images. The validation dataset has two sets of 120 kVp abdominal images, two sets of 80 kVp and two sets of 140 kVp head images.

CBCT projection data of abdominal patients and phantom are acquired to test the performance of the proposed method. Abdominal projection data are acquired using the OBI installed on the Varian Trilogy treatment system with the tube voltage of 125 kVp (Varian Medical Systems, Palo Alto, CA, USA). We collect 20 patients' data to evaluate the performance of the proposed method. The CBCT projection data of a head phantom is acquired from our customized tabletop CBCT imaging system which consists of the x-ray tube from Varian Medical Systems (Rad94, www.varian.com) and the detector from Careray Corp. (1800RF, cn.careray.com). The scanning tube voltages of the head phantom are 75 kVp, 100 kVp and 125 kVp, respectively. We also obtain the MDCT data of the head phantom from the Siemens SOMATOM Definition Flash scanner at the same kVps as the CBCT scans for a side-by-side image quality comparison instead of for model training. The purpose of this operation is to obtain the ground truth to evaluate the quality of the corrected images. The scanning parameters of both MDCT and CBCT data acquisition are listed in table 1.

For the abdominal site, the data is first downsampled from 1024 × 768 to 256 × 192 and then resize to 256 × 256 using zero padding operation. The head data is downsampled from 2816 × 2816 to 256 × 256. Image gray values are normalized to [0 1] using the Min–Max scaling method (Kahng *et al* 2002) during the training and testing process.

3. Evaluation

3.1. Implementation details

The proposed model training is implemented using PyTorch (www.pytorch.org), an open-source python machine learning library based on Torch (Paszke *et al* 2019) and an official Pix2pix project that Isola *et al* proposed in Isola *et al* (2017). The training and testing processes are implemented on a GPU workstation with 128 GB mainboard memory, Intel i7- 9700 K CPU and a single GeForce RTX 2080Ti graphics card with 11 GB graphic memory. The momentum parameters are set as $\beta_1 = 0.5$ and $\beta_2 = 0.999$. The epoch number is set as 400 to guarantee the convergence. In the first 150 epochs, the learning rate is set as a fixed value of 0.0002. In the following 250 epochs, the learning rate gradually decreases to zero to ensure the network convergence to a local minimum and avoid oscillation (You *et al* 2019). Due to the memory limitation, the batch size is set as eight in the model training.

The MC simulation is also implemented in this workstation. To balance the calculation complexity and the scatter estimation accuracy, the number of photons simulated within each projection is set as 1×10^8. In the training data generation, the number of calculated projections for each MDCT dataset is 600. They all use the same list of projection angles which is equally spaced between [0 360) degrees.

Table 2. Trained models applied in the comparison study. RF means receptive field and LS means least square loss function.

Model	Generator	Discriminator	Loss function
1	U-net	RF=70 × 70	LS
2	Resnet-3blocks	RF=70 × 70	LS
3(proposed)	Resnet-6blocks	RF=70 × 70	LS
4	Resnet-9blocks	RF=70 × 70	LS
5	Resnet-6blocks	RF=1 × 1	LS
6	Resnet-6blocks	RF=256 × 256	LS
7	Resnet-6blocks	RF=70 × 70	Sigmoid

3.2. Comparison study

In the comparison study, we evaluate the various configurations of loss functions and network structures of the generator and the discriminator. In the proposed projection-to-scatter transformation, we perform two comparison studies to choose the optimal generator. The U-net and Resnet are applied as the generators individually to compare the performance of scatter correction of the trained model. Resnet with different network depths are further compared to determine the optimal depth to achieve a trade-off between computational accuracy and efficiency. In this study, we choose the number of residual blocks as three, six and nine in the Resnet generator, respectively, while keeping the PatchGAN intact as the discriminator.

The discriminators with different dimensions of RFs are compared to determine the optimal RF and achieve a trade-off between the computational accuracy and training difficulty. We apply PatchGAN with $RF = 1 \times 1$, $RF = 70 \times 70$ and $RF = 256 \times 256$ as the discriminator, while keeping the Resnet with six residual blocks intact as the generator.

Compared with the sigmoid cross-entropy loss, the LS loss function achieves higher-quality image generation and a more stable training process since it penalizes the fake samples and overcomes the gradient vanishment problem (Mao *et al* 2017). To further verify this claim, we compare the results corrected by models trained with the LS and the sigmoid cross-entropy loss when the generator is a Resnet with six residual blocks and the discriminator is the PatchGAN with $RF = 70 \times 70$.

The projection and scatter data of head patients at 80 kVp are used as the training dataset to train the single-spectral model to compare the imaging performance with the proposed method. The network is the same as the proposed method for a fair comparison.

The trained models in the comparison study are summarized in table 2. In the model training, we maintain the same training dataset, hardware configuration, optimizer architecture, decay rate and the number of epochs to ensure a fair comparison.

3.3. Image quality metrics

The root of mean square error (RMSE), the Pearson correlation coefficient (PCC) (Biguri *et al* 2016), the mean structural similarity index (SSIM) (Nomura *et al* 2019, Lalonde *et al* 2020) and spatial nonuniformity (SNU) are selected to evaluate the correction performance of the proposed method. In all studies, the RMSE, PCC and SSIM are calculated based on the white rectangle area and the whole image and the SNU is calculated based on six ROIs selected in the uniform areas. In the evaluation of the correction results of the mimicked scatter-contaminated MDCT, the metrics are calculated between the corrected and the original MDCT images. In the evaluation of the CBCT scatter correction, the metrics are calculated between the corrected CBCT images and the corresponding high-quality MDCT images for the head phantom and the CBCT images corrected by MC simulation for abdominal patient data. The RMSE which indicates the squared differences of pixel intensities of two images is defined as:

$$RMSE = \sqrt{\frac{1}{m}\sum_{i=1}^{m}(\mu_{c\ i} - \mu_{g\ i})^2}, \tag{11}$$

where i is the index of the pixel in the selected ROI, m is the total number of the pixels in the ROI, $\mu_{c\ i}$ is the HU value of the ith pixel in the corrected CT image using the proposed method and $\mu_{g\ i}$ is the HU value of the ith pixel in the ground-truth image.

The PCC is an index to measure the linear dependence between two images and is calculated as:

$$r = \frac{\sum_{i=1}^{m}(\mu_{c\ i} - \overline{\mu_c})(\mu_{g\ i} - \overline{\mu_g})}{\sqrt{\sum_{i=1}^{m}(\mu_{c\ i} - \overline{\mu_c})^2}\sqrt{\sum_{i=1}^{m}(\mu_{g\ i} - \overline{\mu_g})^2}}, \tag{12}$$

where $\overline{\mu_c}$ and $\overline{\mu_g}$ are the mean values of the selected ROI in the corrected and ground-truth images, respectively.

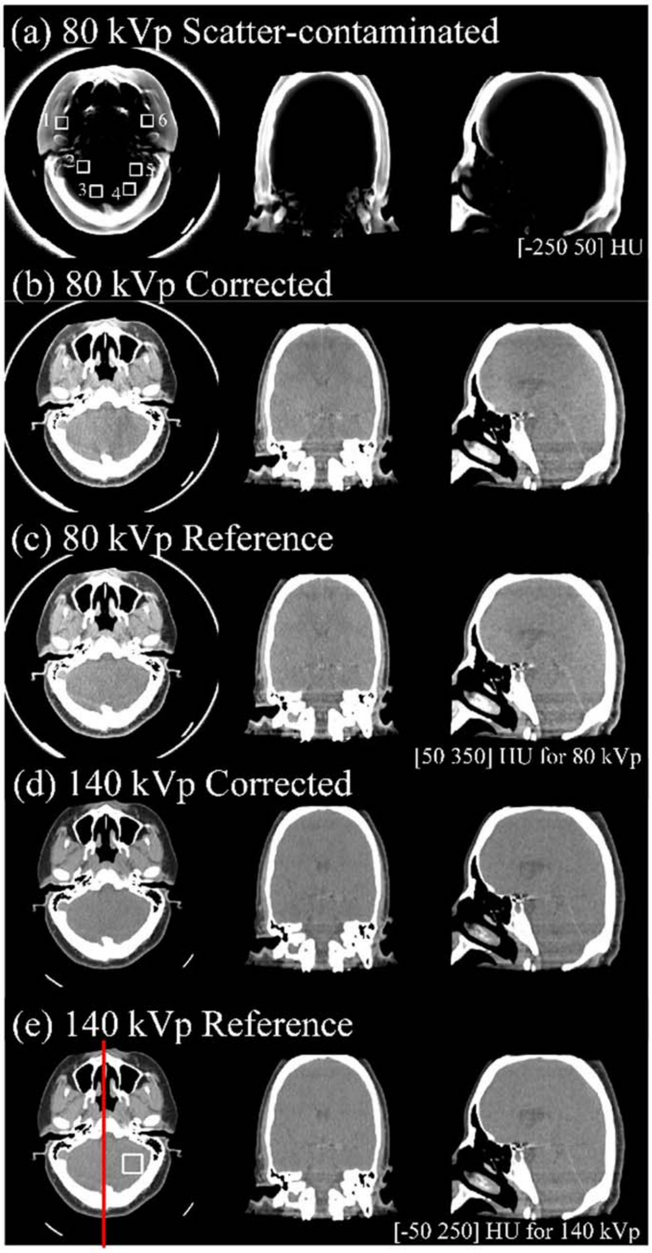


Figure 3. The axial, coronal, and sagittal views of the head patient validation data: (a) mimicked scatter-contaminated image at 80 kVp, (b) corrected CT image at 80 kVp, (c) the reference MDCT image at 80 kVp, (d) corrected CT image at 140 kVp and (e) the reference MDCT image at 140 kVp. Displace window is [−250 50] HU for (a), [50 350] HU for (b), (c) and [−50 250] HU for (d), (e).

The SSIM indicates the similarity of two images regarding intensity, contrast and structure. It is proposed to provide a good approximation of perceptual image quality and is defined as:

$$SSIM = \frac{(2\overline{\mu_c}\,\overline{\mu_g} + C_1)(2\sigma_{cg} + C_2)}{(\overline{\mu_c}^2 + \overline{\mu_g}^2 + C_1)(\sigma_c^2 + \sigma_g^2 + C_2)}, \tag{13}$$

where $C_1 = 0.01 \times d$ and $C_2 = 0.03 \times d$ are the small constants to avoid instability of SSIM calculation. Here d is the dynamic range of the pixel values in the ROI of the corrected image. σ_c and σ_g are the standard deviations of the selected ROIs in the corrected CT image and the ground-truth image, respectively. σ_{cg} is the covariance between the corrected image and the ground-truth image.

To demonstrate that our method can maintain the image uniformity, the SNU is used as a quality metric and defined as:

$$SNU = \left| \frac{\overline{HU}_{\max} - \overline{HU}_{\min}}{1000} \right| \times 100\%. \tag{14}$$

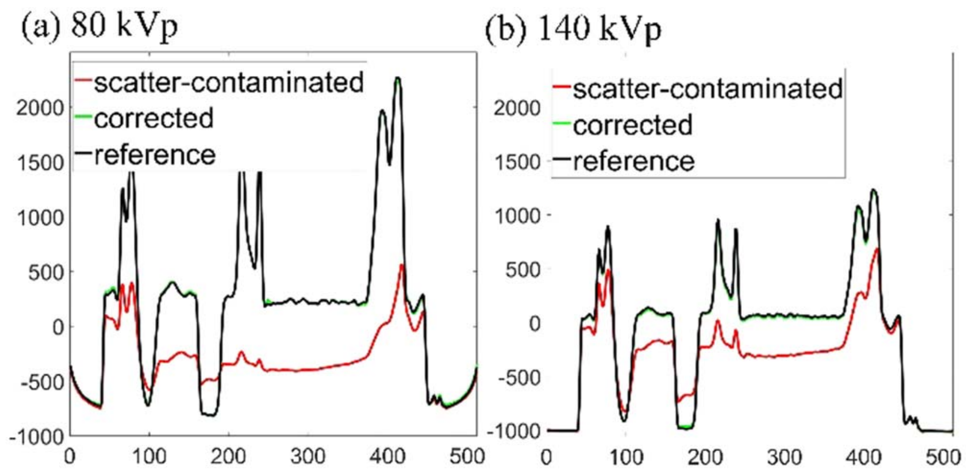


Figure 4. One-dimensional profiles along the red line in figure 3(e). (a) is for the 80 kVp study and (b) is for the 140 kVp study.

Table 3. Comparison of the RMSE, PCC, SSIM and SNU evaluated using the head validation data at 80 and 140 kVps. The values outside brackets are calculated on the white rectangle in figure 3(e) and the values in the brackets are calculated based on the whole image.

	RMSE(HU)	PCC	SSIM	SNU
Uncorrected-80 kVp	603.4(506.5)	0.1156(0.9075)	0.1007(0.6691)	9.06%
Corrected	6.2(27.3)	0.9461(0.9995)	0.9473(0.9993)	5.26%
Reference	0	1	1	4.39%
Uncorrected-140 kVp	355.4(226.9)	0.0264(0.9730)	0.0354(0.9021)	6.11%
Corrected	4.5(18.5)	0.9914(0.9996)	0.9911(0.9995)	2.13%
Reference	0	1	1	2.69%

$\overline{HU}_{\max}$ and $\overline{HU}_{\min}$ are the maximum and the minimum of the mean values of the six ROIs selected in the uniform areas.

4. Results

4.1. Head study

4.1.1. Validation dataset using MDCT

Figure 3 shows the axial, coronal, and sagittal views of the mimicked scatter-contaminated, the corrected and the reference MDCT images which are selected from the validation dataset of the head patient at the kVps of 80 and 140. We only show the mimicked scatter-contaminated image of 80 kVp for comparison. The one-dimensional (1D) profiles along the vertical red line in figure 3(e) are shown in figure 4. We select the region contoured by the white rectangle in figure 3(e) and the whole image to calculate the image quality metrics and the results are listed in table 3. The evaluation metrics on the whole image are listed in parentheses. The RMSE of the corrected image is reduced from 603.4 HU (506.5 HU) to 6.2 HU (27.3 HU) for 80 kVp data and from 355.4 HU (226.9 HU) to 4.5 HU (18.5 HU) for 140 kVp data. The PCCs between the corrected and the reference images are 0.9461 (0.9995) and 0.9914 (0.9996) while they are 0.1156 (0.9075) and 0.0264 (0.9730) before correction for the 80 kVp and 140 kVp data, respectively. The SSIMs of the corrected images of 80 and 140 kVp data are close to the ground truth, indicating that the anatomy in the image is not altered after the proposed scatter correction. The SNU is calculated based on the six ROIs shown in figure 3(a). The SNUs of the scatter-contaminated, corrected and the reference images are 9.06%, 5.26% and 4.39% in the 80 kVp study, and 6.11%, 2.13% and 2.69% in the 140 kVp study, respectively.

4.1.2. Testing data using CBCT

Figure 5 shows the axial, coronal, and sagittal views of the uncorrected, the corrected and the reference images of the CBCT data of a head phantom whose scanning kVps are 75, 100 and 125. We only show the uncorrected and reference images at 75 kVp. We apply the registered MDCT as the reference image in the head phantom study. 1D profile along the vertical red line in figure 5(e) is shown in figure 6. The ROI enclosed by the white box in

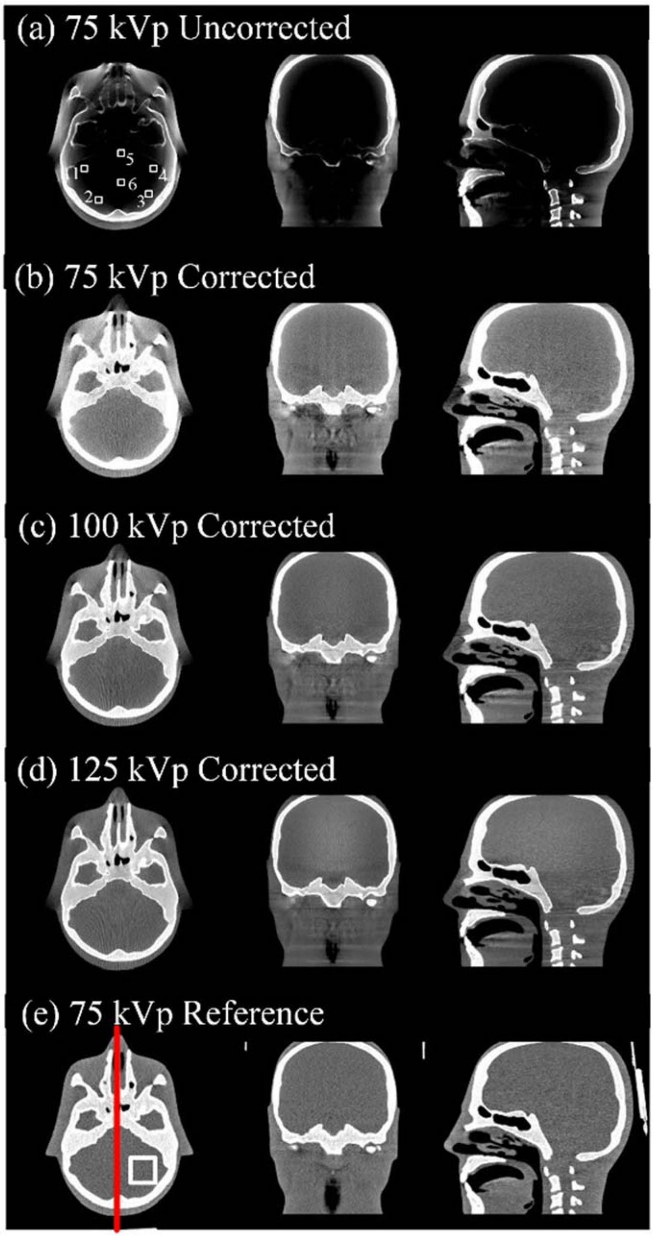


Figure 5. The axial, coronal, and sagittal views of the head phantom CBCT data: (a) uncorrected image at 75 kVp, (b)–(d) corrected image at 75 kVp, 100 kVp, 125 kVp and (e) the reference MDCT image at 75 kVp. Displace window is [−150 400] HU for all.

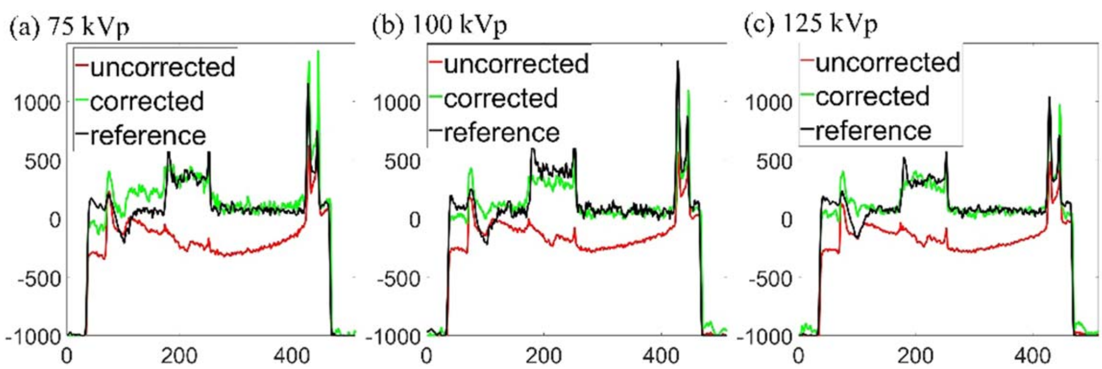


Figure 6. One-dimensional profiles along the red line in figure 5(e). (a): 75 kVp, (b): 100 kVp and (c): 125 Vp.

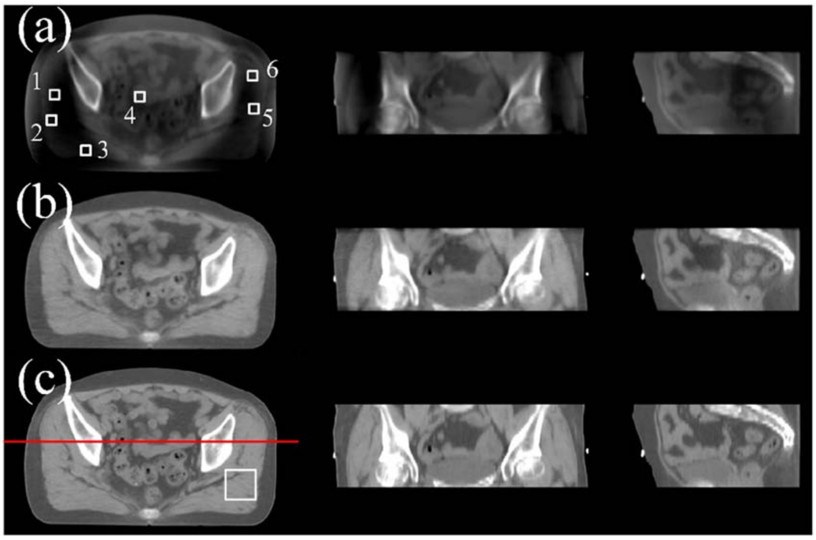


Figure 7. The axial, coronal, and sagittal views of the MDCT images of one patient in the validation dataset: (a) mimicked scatter-contaminated image, (b) corrected image using the proposed method and (c) the reference MDCT image. Displace window is [−250 300] HU for all.

Table 4. Comparison of the RMSE, PCC, SSIM and SNU evaluated using the head phantom data at 75,100 and 125 kVps. The values outside brackets are calculated on the white rectangle in figure 5(e) and the values in the brackets are calculated based on the whole image.

	RMSE(HU)	PCC	SSIM	SNU
Uncorrected-75 kVp	346.2(321.5)	0.1447(0.9520)	0.1253(0.9166)	13.77%
Corrected	37.7(51.0)	0.8612(0.9795)	0.8507(0.9778)	3.74%
Reference	0	1	1	0.64%
Uncorrected-100 kVp	287.9(230.2)	0.2007(0.9599)	0.1752(0.9107)	11.28%
Corrected	42.8(56.4)	0.8852(0.9818)	0.8424(0.9814)	1.56%
Reference	0	1	1	1.93%
Uncorrected-125 kVp	299.8(219.0)	0.2157(0.9745)	0.2274(0.9251)	10.24%
Corrected	34.8(45.7)	0.8765(0.9881)	0.8554(0.9872)	0.34%
Reference	0	1	1	0.95%

figure 5(e) and the whole image are used to calculate RMSE, PCC and SSIM. The six small ROIs in figure 5(a) are used to calculate the SNU value. The quantitative results are listed in table 4. The RMSE of the corrected image is reduced from 346.2 HU (321.5 HU) to 37.7 HU (51.0 HU), from 287.9 HU (230.2 HU) to 42.8 HU (56.4 HU) and from 299.8 HU (219.0 HU) to 34.8 HU (45.7 HU) in the 75 kVp, 100 kVp and 125 kVp studies, respectively. The RMSEs between the CBCT and MDCT images is more than 30 HU due to the energy spectra difference between the two scanning devices. The PCCs between the corrected and the reference images are 0.8612 (0.9795), 0.8852 (0.9818) and 0.8765 (0.9881) and are only 0.1447 (0.9520), 0.2007 (0.9599) and 0.2157 (0.9745) before correction for 75 kVp, 100 kVp and 125 kVp data, respectively. The SSIM of the corrected image is close to the ground truth, which indicates that the anatomy of the image is not altered after the proposed scatter correction. The SNUs of the uncorrected image, corrected image and the reference image are 13.77%, 3.74% and 0.64% for 75 kVp data, 11.28%, 1.56% and 1.93% for 100 kVp data and 10.24%, 0.34% and 0.95% for 125 kVp data, respectively.

According to the correction results of the above three groups of different scanning spectra, the proposed method can eliminate most of the shading artifacts and improve the image quality of CBCT to be closer to the high-quality MDCT.

4.2. Abdominal patient study

4.2.1. Validation data using MDCT

The corrected results using the MDCT image in the validation dataset of the abdominal patient are shown in figure 7. The 1D profiles along the horizontal red line in figure 7(c) are shown in figure 10(a). The MDCT data of

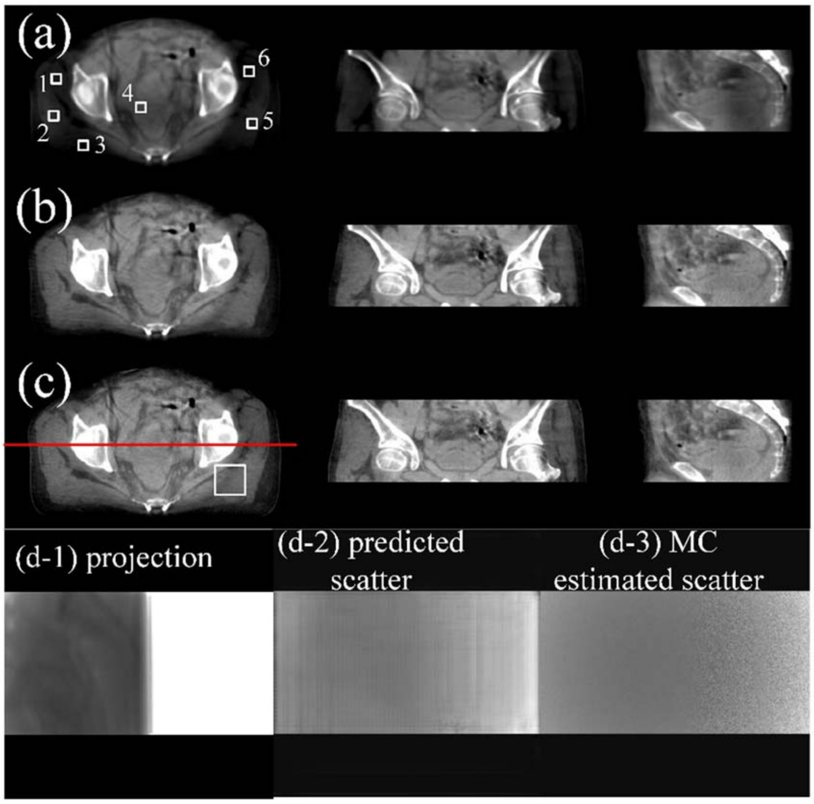


Figure 8. The axial, coronal, and sagittal views of the abdominal CBCT images of one patient whose corresponding MDCT is in the training dataset: (a) uncorrected CBCT image, (b) corrected CBCT image using the proposed method and (c) the reference CBCT image using MC simulation. (d) The original projection, the predicted scatter signal and the MC estimated scatter signal. Displace window is [−250 300] HU for (a)–(c) and [0 2400] detector intensity for (d).

Table 5. Comparison of the RMSE, PCC, SSIM and SNU evaluated using the MDCT in the validation dataset. The values outside brackets are calculated on the white rectangle in figure 7(c) and the values in the brackets are calculated based on the whole image.

	RMSE(HU)	PCC	SSIM	SNU
Uncorrected	259.4(143.9)	0.1656(0.9793)	0.1594(0.9402)	12.01%
Corrected	12.7(15.8)	0.9074(0.9901)	0.9082(0.9812)	4.84%
Reference	0	1	1	2.71%

this patient is not involved in the model training. The results of the image quality metrics are listed in table 5. The RMSE between the scatter-contaminated image and the reference image is 259.4 HU (143.9 HU) while decreasing to 12.7 HU (15.8 HU) after the proposed correction. The PCC increases from 0.1656 (0.9793) to 0.9074 (0.9901) and SNU reduces from 12.01% to 4.84% while maintaining the anatomy after the proposed correction.

4.2.2. CBCT scatter correction

Figure 8 shows the results of the correction results of the abdominal CBCT. The reference image is the CBCT image corrected by the MC simulation method. The 1D profile along the horizontal red line in figure 8(c) is shown in figure 10(b). The quality of the CBCT image is greatly improved after correction using the proposed method. The results of the image quality metrics are listed in table 6. The original projection, the predicted scatter signal and the MC estimated scatter signal are shown in figure 8(d). The RMSE of the corrected CBCT image is reduced from 186.3 HU (100.7 HU) to 11.8 HU (25.0 HU) compared with the image without correction. The PCC between the corrected image and the reference image is 0.8409 (0.9976) while it is only 0.3229 (0.9738) before correction. The SSIM of the CBCT image promotes from 0.3178 (0.9814) to 0.8386 (0.9971) and the SNU reduces from 11.17% to 6.85% after correction. We collect CBCT data of another 15 abdominal patients and performed scatter correction using the proposed method to verify that the training dataset is large enough. We compare the correction results with the MC-based method. The RMSEs of the proposed method are reduced from 200 HU to around 25 HU in this fifteen-patient study.

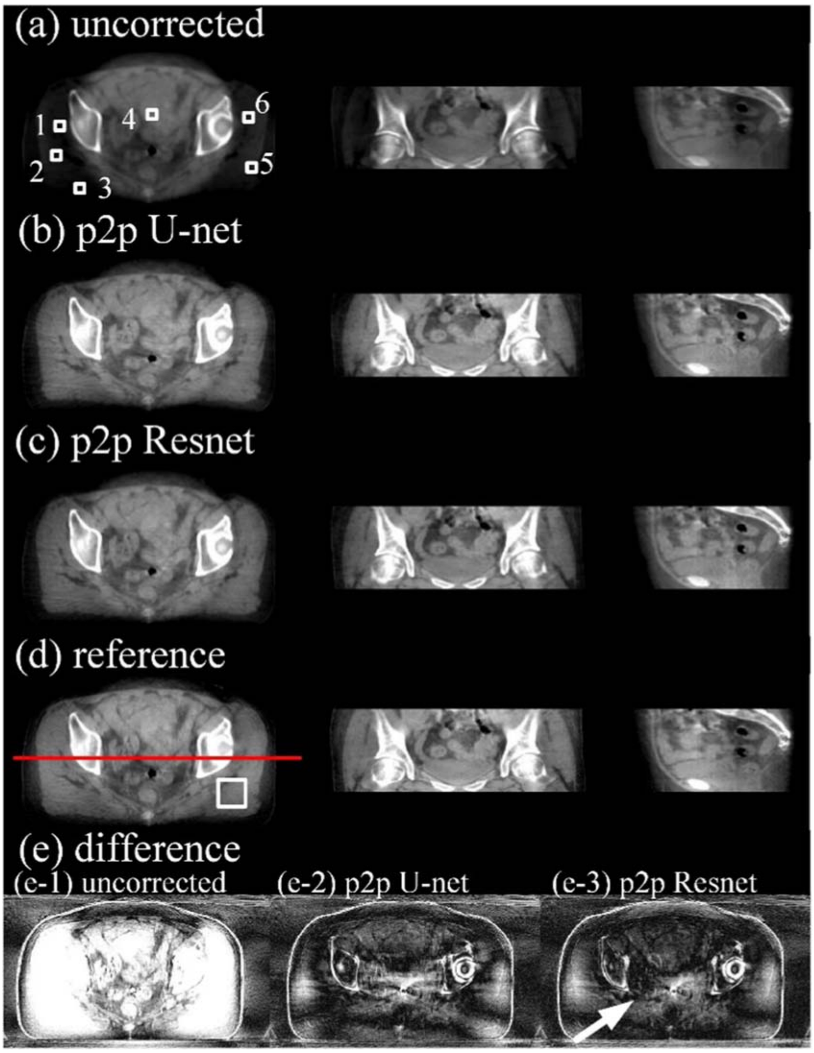


Figure 9. The axial, coronal, and sagittal views of the abdominal CBCT images of one patient whose corresponding MDCT is in the validation dataset. (a) Uncorrected CBCT image, (b) corrected CBCT image via Pix2pix GAN with a U-net generator, (c) the corrected CBCT image via the proposed method and (d) the reference CBCT image via MC simulation. (e) The absolute differences between the reference CBCT image and (e-1) uncorrected CBCT image, (e-2) the corrected CBCT image via Pix2pix GAN with a U-net generator, (e-3) the corrected CBCT image via the proposed method. Displace window is [−250 300] HU for (a)–(d) and is [0 50] HU for (e).

Table 6. Comparison of the RMSE, PCC, SSIM and SNU evaluated using the abdominal CBCT whose corresponding MDCT is in the training dataset. The values outside brackets are calculated on the white rectangle in figure 8(c) and the values in the brackets are calculated based on the whole image.

	RMSE(HU)	PCC	SSIM	SNU
Uncorrected	186.3(100.7)	0.3229(0.9738)	0.3178(0.9814)	11.17%
Corrected	11.8(25.0)	0.8409(0.9976)	0.8386(0.9971)	6.85%
Reference	0	1	1	5.60%

4.3. Comparison study

4.3.1. U-net and Resnet

To demonstrate the advantages of our method, the abdominal CBCT whose corresponding MDCT is in the validation datasets is used to quantitatively evaluate the performances of the proposed network and the Pix2pix GAN with a U-net generator. Figure 9 shows the CBCT images corrected using the two schemes and the MC simulation as the reference. The 1D profile along the horizontal red line in figure 9(d) is shown in figure 10(c). The absolute differences between the corrected image and the reference image are shown in figure 9(e). It can be seen from the residual image that the difference between the proposed method and the reference image is smaller compared with the Pix2pix GAN with the U-net generator method especially in the area pointed by the

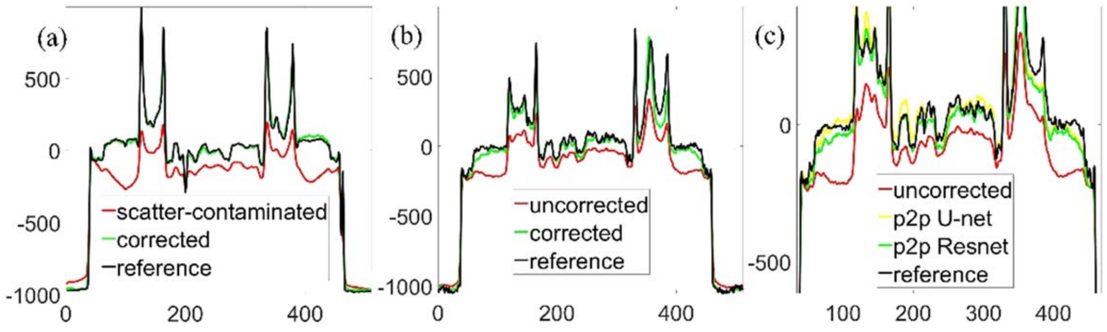


Figure 10. The one-dimensional profiles along the red line. (a) is in figure 7(c), (b) is in figure 8(c) and (c) is in figure 9(d).

white arrow in figure 9(e-3). The results of the image quality metrics are listed in table 7. The Pix2pix GAN with a U-net generator produces streaking artifacts after correction due to the incorrect scatter estimation. U-net obtains higher segmentation accuracy by classifying each pixel which is more suitable for medical image segmentation. The RMSEs of the corrected CBCT images are 21.2 HU and 16.3 HU using the Pix2pix GAN with U-net generator and the proposed method while the SNUs are 10.47% and 8.72%. The proposed Pix2pix GAN with the Resnet generator achieves higher accuracy in scatter estimation than that with the U-net generator.

4.3.2. The number of residual blocks in Resnet

Another abdominal CBCT whose corresponding MDCT is not in the training datasets is used to quantitatively evaluate the performances of the Pix2pix GAN with the Resnet generator of different numbers of residual blocks. Figures 11(b)–(d) show the CBCT images corrected using the three generators with three, six and nine residual blocks. The absolute differences between the corrected image and the reference image are shown in figures 12(a)–(c). The results of the image quality metrics are listed in table 8. As shown in the absolute difference maps, the nine-block Resnet achieves the best scatter estimation while the three-block Resnet is of the worst performance compared with the reference. The generator with six blocks has the same performance compared with the nine-block one using a much lower computational cost. We therefore apply the six-block Resnet as the generator in this study since it only needs two-thirds of the training time of the nine-block Resnet.

4.3.3. Discriminator parameters

Figures 11(c), (e), (f) show the CBCT images corrected using the discriminators with different dimensions of RF. The absolute differences between the corrected image and the reference image are shown in figures 12(b), (d), (e). Using the PatchGAN with $RF = 70 \times 70$, the difference between the corrected image and the reference image is the smallest compared with the $RF = 1 \times 1$ and the $RF = 256 \times 256$.

4.3.4. Least-square loss and sigmoid loss

The abdominal CBCT data is used in the comparison study of the least-square loss and sigmoid cross-entropy loss. The corrected image of the sigmoid loss is shown in figure 11(g). The absolute difference between the corrected image and the reference image is shown in figure 12(f). Using the least-square loss as the loss function for model training, RMSE is reduced from 32.1 to 21.5 HU compared with that using the model trained by sigmoid cross-entropy loss.

4.3.5. Single-spectral and multiple-spectral models

The mimicked scatter-contaminated MDCT of head in the validation datasets is used to quantitatively evaluate the performances of the multiple-spectral model and the single-spectral model. Figure 13 shows the axial views of the mimicked scatter-contaminated, the corrected and the reference MDCT images which are selected from the validation dataset of the head patient at the kVps of 80 and 140. It can be seen from the residual image that the single-spectral model has a similar performance compared with the proposed method in the results of the head patient at 80 kVp (0.5 HU difference in RMSE). The results of the image quality metrics are listed in table 9. In the 140 kVp head-patient study, the single-spectral model does not work very well since the overcorrection artifacts are produced while reducing the scatter artifacts compared with the proposed model. The proposed multiple-spectral model can be applied to the CBCT dataset which has different scan modes and spectra.

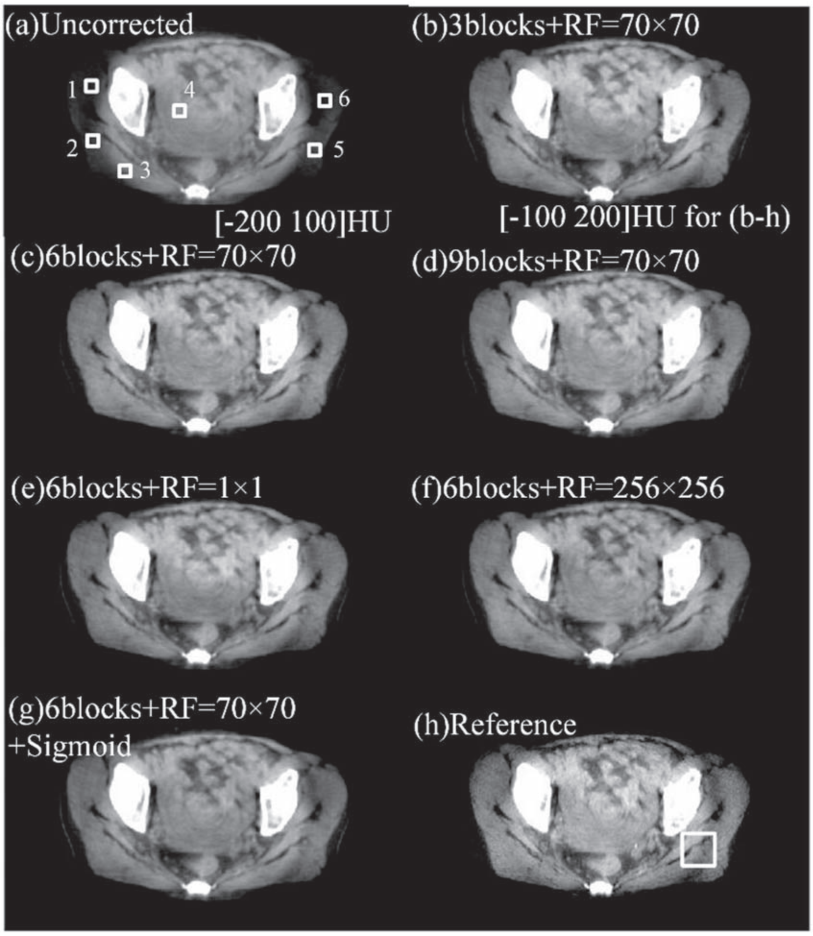


Figure 11. The axial views of the abdominal CBCT images of one patient whose corresponding MDCT is not in the training dataset. (a) Uncorrected CBCT image, (b)–(g) corrected CBCT images via Pix2pix GAN with different network parameters and (h) the reference CBCT image via MC simulation. Displace window is [−200 100] HU for (a) and [−100 200] HU for (b)–(h).

Table 7. Comparison of the RMSE, PCC, SSIM and SNU evaluated using the abdominal CBCT images in the comparison study. The values outside brackets are calculated on the white rectangle in figure 9(d) and the values in the brackets are calculated based on the whole image.

	RMSE(HU)	PCC	SSIM	SNU
Uncorrected	216.2(101.3)	0.0891(0.9941)	0.0811(0.9815)	17.88%
p2p U-net	21.2(41.5)	0.7088(0.9957)	0.7286(0.9952)	10.47%
p2p Resnet	16.3(31.9)	0.8090(0.9958)	0.8086(0.9956)	8.72%
Reference	0	1	1	8.48%

5. Discussion

In this paper, we propose a generalized image quality improvement strategy of CBCT using the multiple spectral MDCT labels. The proposed method learns the model using the Pix2pix GAN which combines the six-block Resnet as the generator and the PatchGAN as the discriminator to translate the scatter-contaminated projection to the scatter signal. In the head and abdominal patient studies, the proposed model achieves accurate scatter estimation and produces high-quality CBCT images in the projection domain.

The novelty of this study is reflected in four-folds. First of all, the x-ray energy spectrum is an important parameter to affect the estimated scatter distribution in model training. Different from other studies only working on the scatter estimation at a single spectrum, the proposed method achieves multiple spectral CBCT scatter estimation. Secondly, the proposed model is suitable for the CBCT with various imaging volumes and multiple scanning modes. For example, on the OBI system in the Trilogy, a single model we generated achieves the scatter correction for the CBCT data obtained by the half-fan and full-fan modes. Thirdly, we apply the projection-to-scatter transition to estimate the scatter distribution from the raw projection. Compared with the

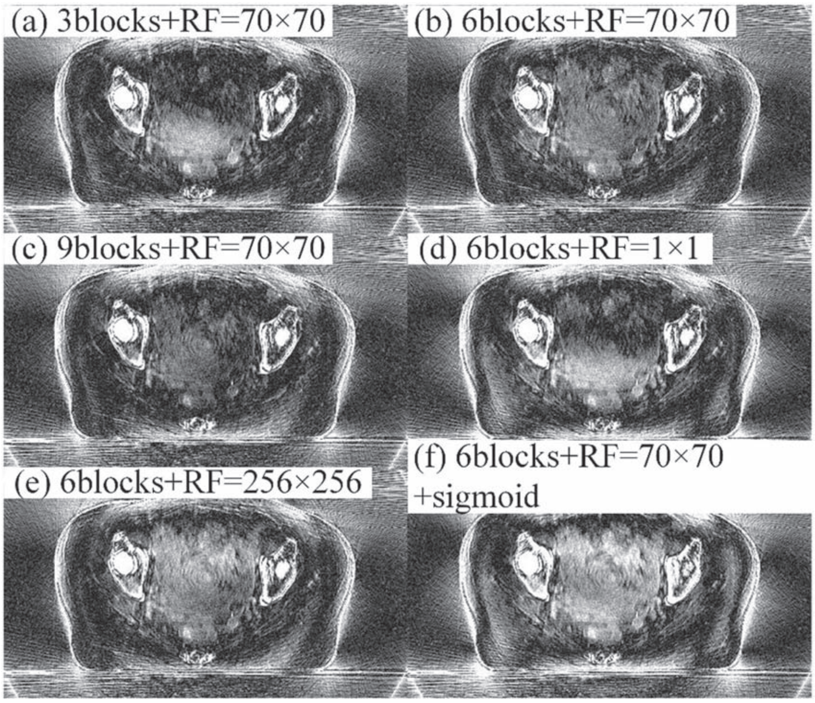


Figure 12. The absolute differences between the reference CBCT image. Displace window is [0 50] HU for all.

Table 8. Comparison of the RMSE, PCC, SSIM and SNU evaluated using the abdominal CBCT images in the comparison study. The values outside the brackets are calculated on the white rectangle in figure 11(h) and the values in the brackets are calculated based on the whole image.

	RMSE(HU)	PCC	SSIM	SNU
Uncorrected	212.5(105.0)	0.4631(0.9816)	0.4445(0.9604)	13.21%
3blocks+RF = 70 × 70	23.4(45.8)	0.8670(0.9965)	0.8466(0.9945)	4.82%
6blocks + RF = 70 × 70	21.5(44.4)	0.8794(0.9967)	0.8663(0.9951)	4.44%
9blocks + RF = 70 × 70	17.9(43.9)	0.8702(0.9968)	0.8535(0.9951)	4.36%
6blocks + RF = 1 × 1	30.8(46.6)	0.8496(0.9962)	0.8382(0.9944)	4.74%
6blocks + RF = 256 × 256	21.1(44.8)	0.8393(0.9966)	0.8258(0.9937)	4.93%
6blocks + RF = 70 × 70+sigmoid	32.1(48.7)	0.8519(0.9958)	0.8400(0.9604)	5.14%
Reference	0	1	1	4.56%

image-to-image mapping and the projection-to-projection mapping, the proposed method ensures no anatomical structure deformation of the reconstructed image during the scatter correction process due to the dominantly low-frequency behavior of scatter distribution. No extra image processing technique, e.g. image registration, is required to operate on the sample and label. Finally, the proposed method is a post-processing and real-time scatter correction scheme based on deep learning. It does not increase the radiation dose to the patient nor does it modify the existing hardware of the CBCT system. After the model training, real-time scatter estimation capability accelerates the generation of high-quality CBCT images for advanced CBCT-guided radiation therapy including patient setup time reduction, physical and psychological patient stress decreasing and accuracy and efficiency improvement.

Although the proposed method can effectively correct the scatter-induced artifacts in CBCT images, it still needs improvement in our future work. The scanning spectra of the MDCT and CBCT for the same imaging volume size are generally different. The spectral discrepancy may lead to inaccurate scatter estimation and residual artifacts in the image. For example, in the correction results of the head phantom at 125 kVp (see figure 5(d)), the scatter artifacts are over-corrected as shown in the sagittal and coronal views. Ideally, the MDCT data in the training dataset should be scanned using multiple spectra. Due to the radiation dose limitation, the patient cannot be repeatedly scanned in the hospital. In future, we plan to use phantoms for MDCT imaging under different x-ray energy spectra to construct multiple spectral training dataset. These datasets can be used to train a generalized scatter correction model by the proposed network. Without increasing the radiation dose to

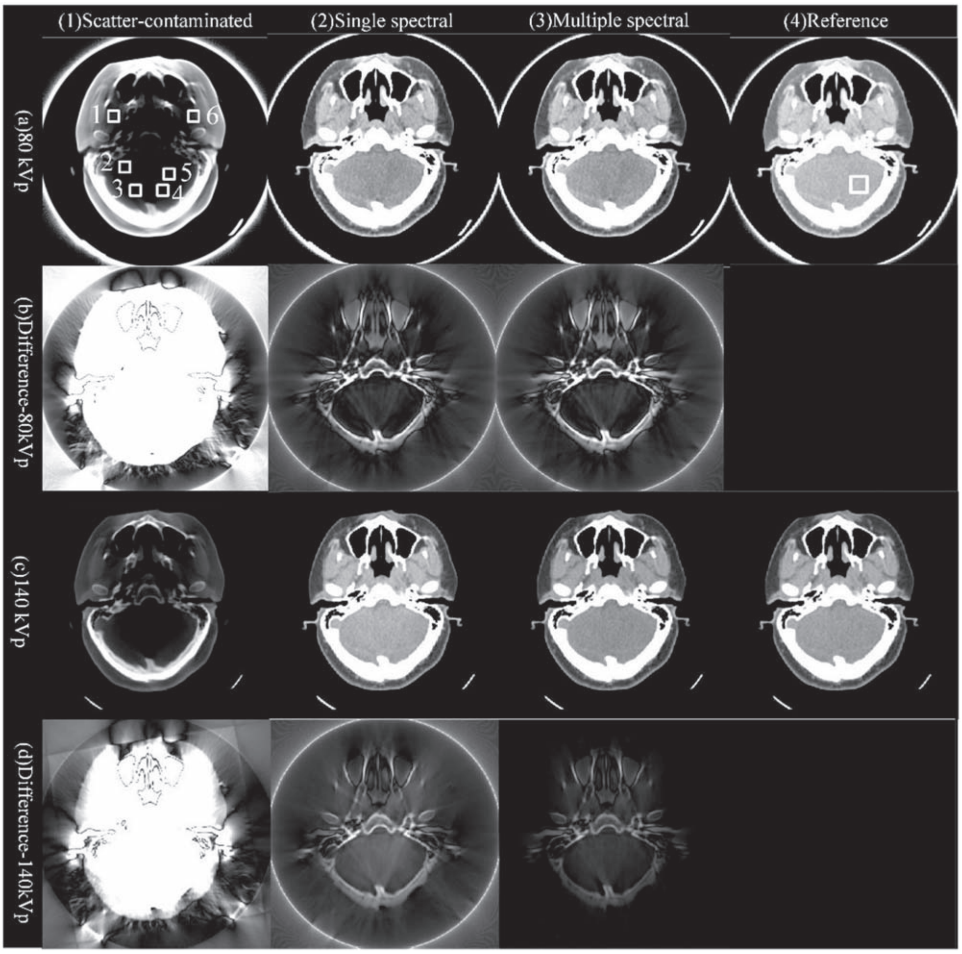


Figure 13. The axial views of the head patient validation images using: (1) mimicked scatter contamination, (2) the single-spectral model, (3) using the proposed method, (4) the reference. (a) 80 kVp images, (b) the difference image of 80 kVp, (c) 140 kVp images, (d) the difference image of 140 kVp. Displace window is [50 350] HU for (a2)–(a4), [−50 250] HU for (c2)–(c4) and [0 50] HU for (b) and (d).

Table 9. Comparison of the RMSE, PCC, SSIM and SNU evaluated using the head patient validation data in the comparison study. The values outside of the brackets are calculated within the white rectangle in figure 1(d) and the values in the brackets are calculated based on the whole image.

	RMSE(HU)	PCC	SSIM	SNU
Uncorrected-80 kVp	603.4(506.5)	0.1156(0.9075)	0.1007(0.6691)	9.06%
Single spectral	5.7(25.7)	0.9518(0.9997)	0.9517(0.9995)	5.18%
Multiple spectral	6.2(27.3)	0.9461(0.9995)	0.9473(0.9993)	5.26%
Reference	0	1	1	4.39%
Uncorrected-140 kVp	355.4(226.9)	0.0264(0.9730)	0.0354(0.9021)	6.11%
Single spectral	27.9(44.5)	0.8812(0.9921)	0.8921(0.9925)	3.76%
Multiple spectral	4.5(18.5)	0.9914(0.9996)	0.9911(0.9995)	2.13%
Reference	0	1	1	2.69%

the patient, an accurate and precise multiple spectral model can be trained for the scatter correction in clinical CBCT imaging.

In addition to scatter contamination, beam hardening effect is another issue degrading CBCT image quality with similar shading artifacts as scatter signals. In our previous work, we propose a beam hardening artifacts correction algorithm in the line-integral domain. This method estimates the polychromatic and monochromatic spectra. The scaled difference of the monochromatic reprojection data and the polychromatic reprojection is added to the raw line-integral data to achieve beam-hardening artifacts correction (Zhao *et al* 2018). This work can be transformed from the line-integral domain into the projection domain. In future, we can combine the estimation of scatter artifacts and beam-hardening artifacts to design a model that can be used for the correction of both beam hardening artifacts and scatter artifacts.

6. Conclusion

The proposed method applies the multiple spectral MDCT labels and the Pix2pix GAN to construct the multiple spectral scatter estimation model. The Pix2pix GAN combines the Resnet as the generator and the PatchGAN as the discriminator to set up the correspondence between the scatter-contaminated projection and the scatter distribution. It achieves excellent performance in scatter estimation and correction to maintain the anatomy of the patient. The proposed method does not require the increase of the radiation dose to the patient, nor does it modify the existing hardware of the CBCT system and is thus practical to be implemented in clinical applications.

Acknowledgments

This work was supported in part by the National Key R&D Program of China (2018YFE0114800, 2019YFB1311800), Natural Science Foundation of China (NSFC Grant No. 81871351).

We would thank the editors and anonymous referees for their constructive comments which significantly strengthen the paper.

ORCID iDs

Yibao Zhang https://orcid.org/0000-0003-3658-0687
Wei Zhao https://orcid.org/0000-0002-6182-4746
Tianye Niu https://orcid.org/0000-0003-4181-3641

References

Altorjai G, Fotina I, Lütgendorf-Caucig C, Stock M, Pötter R, Georg D and Dieckmann K 2012 Cone-beam CT-based delineation of stereotactic lung targets: the influence of image modality and target size on interobserver variability *Int. J. Radiat. Oncol.* Biol.* Phys.* **82** e265–272

Araujo A, Norris W and Sim J 2019 Computing receptive fields of convolutional neural networks *Distill* **4** e21

Arisan V, Karabuda Z C, Avsever H and Özdemir T 2013 Conventional multi-slice computed tomography (CT) and cone-beam CT (CBCT) for computer-assisted implant placement. I: relationship of radiographic gray density and implant stability *Clin. Implant Dentistry Related Res.* **15** 893–906

Badal A and Badano A 2009 Accelerating Monte Carlo simulations of photon transport in a voxelized geometry using a massively parallel graphics processing unit *Med. Phys.* **36** 4878–4880

Baer M and Kachelrieß M 2012 Hybrid scatter correction for CT imaging *Phys. Med. Biol.* **57** 6849

Biguri A, Dosanjh M, Hancock S and Soleimani M 2016 TIGRE: a MATLAB-GPU toolbox for CBCT image reconstruction *Biomed. Phys. Eng. Express* **2** 055010

de la Zerda A, Armbruster B and Xing L 2007 Formulating adaptive radiation therapy (ART) treatment planning into a closed-loop control framework *Phys. Med. Biol.* **52** 4137

Geng M, Tian Z, Jiang Z, You Y, Feng X, Xia Y, Yang K, Ren Q, Meng X and Maier A 2020 PMS-GAN: parallel multi-stream generative adversarial network for multi-material decomposition in spectral computed tomography *IEEE Trans. Med. Imaging* **40** 571–84

Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 Generative Adversarial Networks arXiv:1406.2661

Hansen D C, Landry G, Kamp F, Li M, Belka C, Parodi K and Kurz C 2018 ScatterNet: a convolutional neural network for cone-beam CT intensity correction *Med. Phys.* **45** 4916–4926

Harms J, Lei Y, Wang T, Zhang R, Zhou J, Tang X, Curran W J, Liu T and Yang X 2019 Paired cycle-GAN-based image correction for quantitative cone-beam computed tomography *Med. Phys.* **46** 3998–4009

Haworth A, Paneghel A, Herschtal A, Duchesne G, Williams S, Tai K, Kron T, Roxby P, Soteriou S and Laferlita M 2009 Verification of target position in the post-prostatectomy cancer patient using cone beam CT *J. Med. Imaging Radiat. Oncol.* **53** 212–220

He K, Zhang X, Ren S and Sun J 2016 *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 770–778 vol. Series pp

Isola P, Zhu J-Y, Zhou T and Efros A A 2017 *Proc. IEEE Conf. on Computer Vision and Pattern Recognition* pp 1125–1134 vol. Series pp

Jain S, Choudhary K, Nagi R, Shukla S, Kaur N and Grover D 2019 New evolution of cone-beam computed tomography in dentistry: combining digital technologies *Imaging Sci. Dentistry* **49** 179–190

Jian L, Yang X, Liu Z, Jeon G, Gao M and Chisholm D 2019 A symmetric encoder-decoder with residual block for infrared and visible image fusion arXiv:1905.11447

Jiang Y, Yang C, Yang P, Hu X, Luo C, Xue Y, Xu L, Hu X, Zhang L and Wang J 2019 Scatter correction of cone-beam CT using a deep residual convolution neural network (DRCNN) *Phys. Med. Biol.* **64** 145003

Kahng A B, Mantik S and Markov I L 2002 *Proc. 2002 Int. Symp. on Physical Design* pp 143–148 vol. Series pp

Karlik B and Olgac A V 2011 Performance analysis of various activation functions in generalized MLP architectures of neural networks *Int. J. Artificial Intelligence Expert Syst.* **1** 111–122

Karpathy A 2016 Cs231n convolutional neural networks for visual recognition *Neural Netw.* **1**
Kingma D P and Ba J 2014 Adam: a method for stochastic optimization arXiv:1412.6980
Krizhevsky A, Sutskever I and Hinton G E 2017 Imagenet classification with deep convolutional neural networks *Commun. ACM* **60** 84–90
Kyriakou Y and Kalender W 2007 Efficiency of antiscatter grids for flat-detector CT *Phys. Med. Biol.* **52** 6275
Lalonde A, Winey B, Verburg J, Paganetti H and Sharp G C 2020 Evaluation of CBCT scatter correction using deep convolutional neural networks for head and neck adaptive proton therapy *Phys. Med. Biol.* **65** 245022
Li Z-C, Zhai G, Zhang J, Wang Z, Liu G, Wu G-Y, Liang D and Zheng H 2019 Differentiation of clear cell and non-clear cell renal cell carcinomas by all-relevant radiomics features from multiphase CT: a VHL mutation perspective *Eur. Radiol.* **29** 3996–4007
Malusek A, Sandborg M P and Carlsson G A 2003 *Medical Imaging 2003: Physics of Medical Imaging* (Int. Society for Optics and Photonics) pp 740–751 vol. Series 5030
Mao X, Li Q, Xie H, Lau R Y, Wang Z and Paul Smolley S 2017 *Proc. IEEE Int. Conf. on Computer Vision* vol. Series pp 2794–2802
McDermott L N, Wendling M, Nijkamp J, Mans A, Sonke J-J, Mijnheer B J and van Herk M 2008 3D *in vivo* dose verification of entire hypo-fractionated IMRT treatments using an EPID and cone-beam CT *Radiother. Oncol.* **86** 35–42
Mirza M and Osindero S 2014 Conditional generative adversarial nets arXiv:1411.1784
Mota de Almeida F J, Knutsson K and Flygare L 2014 The effect of cone beam CT (CBCT) on therapeutic decision-making in endodontics *Dentomaxillofacial Radiol.* **43** 20130137
Nickenig H-J and Eitner S 2007 Reliability of implant placement after virtual planning of implant positions using cone beam CT data and surgical (guide) templates *J. Cranio Maxillof. Surg.* **35** 207–211
Niu T, Sun M, Star-Lack J, Gao H, Fan Q and Zhu L 2010 Shading correction for on-board cone-beam CT in radiation therapy using planning MDCT images *Med. Phys.* **37** 5395–5406
Niu T and Zhu L 2010 Overview of x-ray scatter in cone-beam computed tomography and its correction methods *Current Med. Imaging* **6** 82–89
Niu T and Zhu L 2011 Scatter correction for full-fan volumetric CT using a stationary beam blocker in a single full scan *Med. Phys.* **38** 6027–6038
Nomura Y, Xu Q, Peng H, Takao S, Shimizu S, Xing L and Shirato H 2020 Modified fast adaptive scatter kernel superposition (mfASKS) correction and its dosimetric impact on CBCT-based proton therapy dose calculation *Med. Phys.* **47** 190–200
Nomura Y, Xu Q, Shirato H, Shimizu S and Xing L 2019 Projection-domain scatter correction for cone beam computed tomography using a residual convolutional neural network *Med. Phys.* **46** 3142–3155
Orth R C, Wallace M J, Kuo M D and Radiology T A C O T S O I 2008 C-arm cone-beam CT: general principles and technical considerations for use in interventional radiology *J. Vascular Interventional Radiol.* **19** 814–820
Palm Å, Nilsson E and Herrnsdorf L 2010 Absorbed dose and dose rate using the Varian OBI 1.3 and 1.4 CBCT system *J. Appl. Clin. Med. Phys.* **11** 229–240
Paszke A, Gross S, Massa F, Lerer A, Bradbury J, Chanan G, Killeen T, Lin Z, Gimelshein N and Antiga L 2019 Pytorch: An imperative style, high-performance deep learning library *Adv. Neural Inf. Process. Syst.* **32** 8026–8037 vol. Series pp
Pauwels R, Araki K, Siewerdsen J and Thongvigitmanee S S 2015 Technical aspects of dental CBCT: state of the art *Dentomaxillofacial Radiol.* **44** 20140224
Pouliot J 2007 *IMRT, IGRT, SBRT-Advances in the Treatment Planning and Delivery of Radiotherapy* (Basel, Switzerland: Karger Publishers) pp 132–142
Rafferty M A, Siewerdsen J H, Chan Y, Daly M J, Moseley D J, Jaffray D A and Irish J C 2006 Intraoperative cone-beam CT for guidance of temporal bone surgery *Otolaryngology—Head Neck Surgery* **134** 80801–808
Sheth N, De Silva T, Uneri A, Ketcha M, Han R, Vijayan R, Osgood G and Siewerdsen J 2020 A mobile isocentric C-arm for intraoperative cone-beam CT: technical assessment of dose and 3D imaging performance *Med. Phys.* **47** 958–974
Siddon R L 1985 Fast calculation of the exact radiological path for a three-dimensional CT array *Med. Phys.* **12** 22552–255
Sorenson J A and Floch J 1985 Scatter rejection by air gaps: an empirical model *Med. Phys.* **12** 308–316
Xu B, Wang N, Chen T and Li M 2015 Empirical evaluation of rectified activations in convolutional network arXiv:1505.00853
Yang C, Wu P, Gong S, Wang J, Lyu Q, Tang X and Niu T 2017 Shading correction assisted iterative cone-beam CT reconstruction *Phys. Med. Biol.* **62** 8495
You K, Long M, Wang J and Jordan M I 2019 How Does Learning Rate Decay Help Modern Neural Networks? arXiv:1908.01878
Zagoruyko S and Komodakis N 2016 Wide residual networks arXiv:1605.07146
Zbijewski W and Beekman F J 2006 Efficient Monte Carlo based scatter artifact reduction in cone-beam micro-CT *IEEE Trans. Med. Imaging* **25** 817–827
Zhang Y, Yue N, Su M Y, Liu B, Ding Y, Zhou Y, Wang H, Kuang Y and Nie K 2021 Improving CBCT quality to CT level using deep learning with generative adversarial network *Med. Phys.* **48** 2816–2826
Zhao W, Li D, Niu K, Qin W, Peng H and Niu T 2018 Robust beam hardening artifacts reduction for computed tomography using spectrum modeling *IEEE Trans. Computat. Imaging* **5** 333–342
Zhao W, Niu K, Schafer S and Royalty K 2014 An indirect transmission measurement-based spectrum estimation method for computed tomography *Phys. Med. Biol.* **60** 339
Zhao W, Vernekohl D, Zhu J, Wang L and Xing L 2016 A model-based scatter artifacts correction for cone beam CT *Med. Phys.* **43** 1736–1753
Zhao Z-Q, Zheng P, Xu S-T and Wu X 2019 Object detection with deep learning: a review *IEEE Trans Neural Netw. Learn. Syst.* **30** 323212–3232
Zhu L, Bennett N R and Fahrig R 2006 Scatter correction method for x-ray CT using primary modulation: theory and preliminary results *IEEE Trans. Med. Imaging* **25** 1573–1587