# Limited-View Cone-Beam CT Reconstruction Based on an Adversarial Autoencoder Network With Joint Loss

**XIUBIN DAI[1], JIANAN BAI[1], TIANLIANG LIU[2], AND LIZHE XIE[3]**

[1]School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications, Nanjing 210046, China
[2]School of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China
[3]Jiangsu Province Key Laboratory of Oral Disease, Nanjing Medical University, Nanjing 210096, China

Corresponding author: Lizhe Xie (xielizhe@njmu.edu.cn)

**ABSTRACT** Limiting scan views is an efficient way to reduce radiation doses in the cone-beam computed tomography (CBCT) examinations, which unfortunately degrades the reconstructed images. Some methods on the framework of the generative adversarial network (GAN) were developed to improve low-dose CT images after CT reconstruction from the limited-view projections. However, no GAN-based methods were devoted to restoring missing CBCT projections in the sinogram domain before CT reconstruction. To avoid the trade-off between radiation dose and image quality, we propose a limited-view CBCT reconstruction method in the sinogram domain, instead of the image domain. First, this method slices the 3D CBCT projections into multiple 2D pieces. Then, an adversarial autoencoder network is trained to estimate the missing parts of these 2D pieces. To improve the prediction, we apply a joint loss function, including reconstruction loss and adversarial loss to the network. When the new limited-view 3D CBCT projections are acquired, the proposed method uses the trained adversarial autoencoder network to generate the missing parts of the 2D pieces sliced from the current 3D CBCT projections. Then, stacking the completed 2D pieces in order yields full-view 3D CBCT projections. Finally, we reconstruct the CT images from the full-view 3D CBCT projections by using the Feldkamp, Davis, and Kress algorithm. The experiments validate that our method performs well in the prediction of unknown projections and CT reconstruction and are less vulnerable to the number of unknown projections than other methods.

**INDEX TERMS** Limited-view CBCT reconstruction, adversarial autoencoder network, reduction of radiation doses, prediction of missing CBCT projections.

## I. INTRODUCTION

Cone-beam computed tomography (CBCT) is a medical imaging technique that can obtain a three-dimensional CT image quickly and directly. In recent decades, applications of CBCT have expanded to many fields, such as treatment planning [1], dentistry [2], and interventional radiology [3].

Since radiation from routine CT scans is a threat to the lives of patients, low-dose CT (LDCT) has become one of the most critical trends in CBCT [4]. Usually, limiting scan views via X-ray collimation during CBCT scans can reduce the radiation dose effectively. However, in this situation, the acquired CBCT projections, i.e., the limited-view CBCT projections, are incomplete. These limited-view CBCT projections inevitably cause the degradation of reconstructed CT images, which prohibits accurate analysis in clinical applications.

A great deal of research dedicates to low-dose CBCT reconstruction to reduce the radiation dose without compromising image quality. Among existing methods, image-based approaches are the most extensively studied ones. Image-based approaches suppress artifacts in LDCT images after CT reconstruction. There are two main categories of image-based approaches: iterative reconstruction methods [5]–[12] and machine learning methods [13]–[18]. Iterative reconstruction methods enhance reconstructed LDCT images by designing and iteratively maximizing or minimizing various prior-regularized cost functions, such as total variance-based priors [5]–[7], nonlocal mean priors [8], [9] and Huber priors [10].
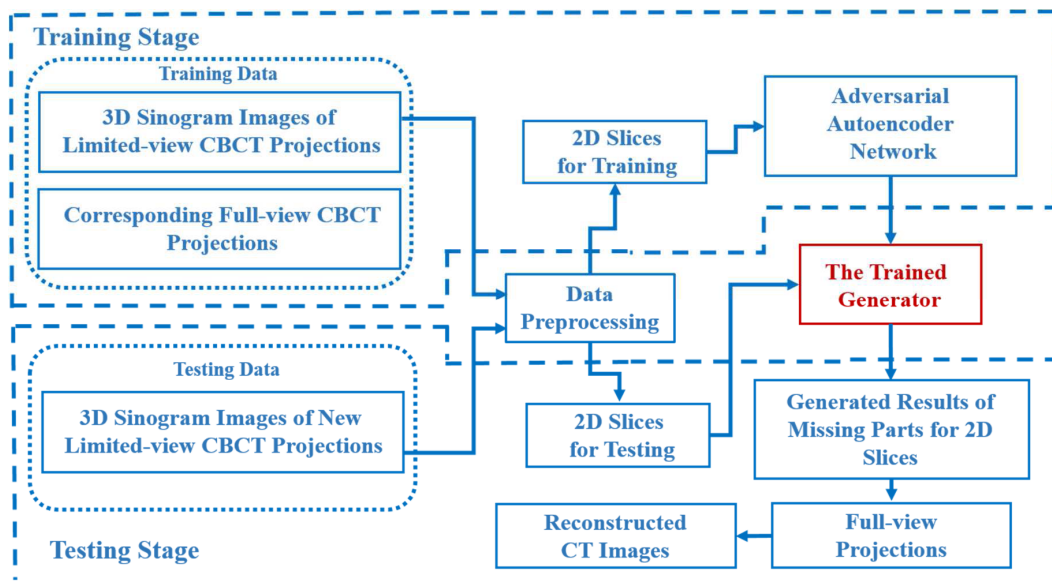
**FIGURE 1.** The pipeline of the proposed limited-view CBCT reconstruction method.

Machine learning methods emerge in the field of LDCT reconstruction in recent years. For instance, Xu *et al*. [13] constructed a global dictionary and an iteratively adaptive dictionary to address low-dose CT reconstruction. To suppress artifacts, Chen *et al*. [14] trained three discriminative dictionaries for high-frequency bands with different orientations. In addition to dictionary learning, Huynh *et al*. [16] used structured random forests and the auto-context model to estimate standard-dose CT (SDCT) images from MRI images and low-dose CT images. In Boublil's work, an artificial neural network was used to boost existing signal and image recovery methods, which was efficient for low-dose CT reconstruction [17]. Zhang *et al*. [18] adapted the Markov random field model to incorporate textural information from previous full-dose CT images for the Bayesian reconstruction of current low-dose CT images.

In addition to traditional machine learning methods, some well-known deep learning methods have also been applied to LDCT and have achieved numerous successes in medical imaging [19]–[27]. For example, a deep convolutional neural network (CNN) was constructed in the wavelet domain and trained with wavelet coefficients from CT images to detect and remove noise in LDCT images [23]. Chen *et al*. [25] built a deep CNN to learn feature mapping from LDCT images to SDCT images. Moreover, some researchers established an encoder-decoder architecture using CNN models to lessen the visibility of artifacts in LDCT images [27].

Very recently, after generative adversarial network (GAN) made significant progress in computer vision and pattern recognition [28], there were attempts to use GAN for noise reduction in LDCT images [29]–[32]. For instance, some trained a CNN as a generator to estimate LDCT images from SDCT images [29] or corresponding magnetic resonance
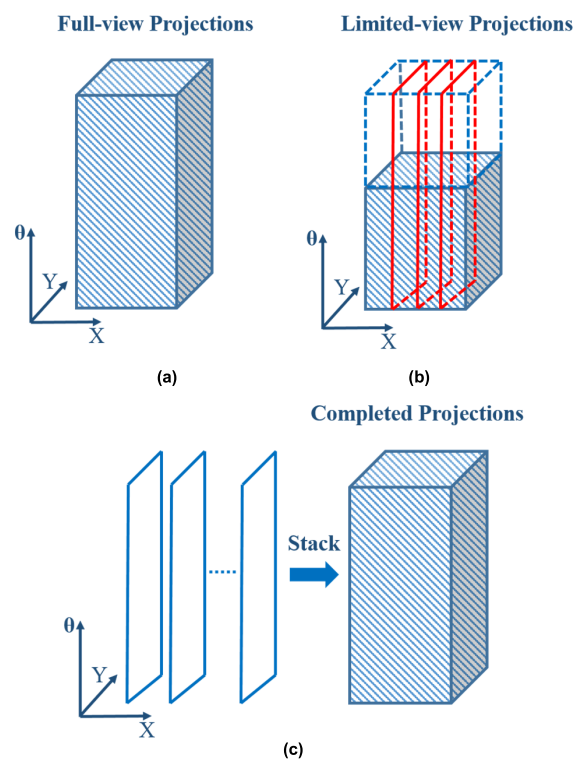


**FIGURE 2.** Data preprocessing for the proposed method. (a) Data matrix of full-view CBCT projections. (b) Data matrix of limited-view CBCT projections, in which the dotted portion indicates their missing part. The planes marked in red represent multiple 2D slices divided from the 3D data matrix along the x-axis. (c) Illustration on the derivation of 3D CBCT projections from multiple completed 2D slices.

images [32] of the same subject. Others tried to restore SDCT images by building a GAN to predict the residual between SDCT images and LDCT images [30].
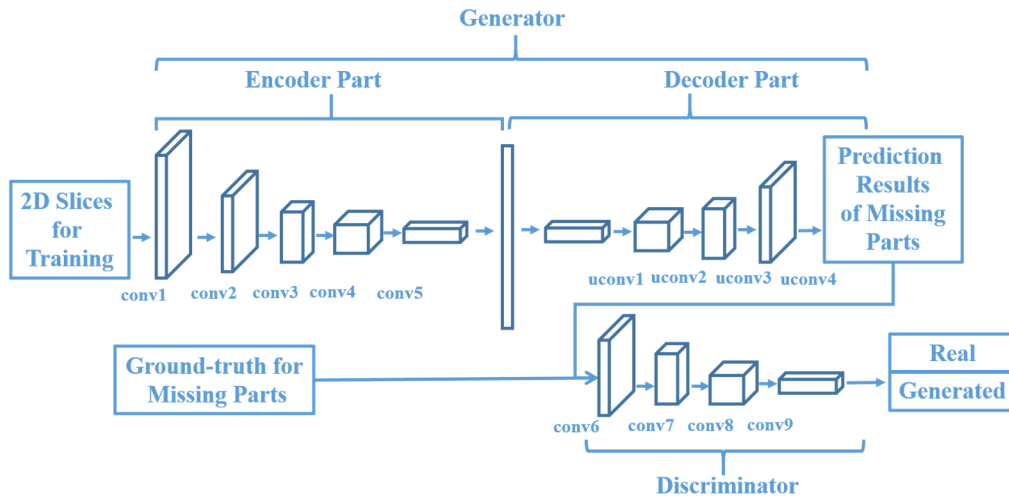
**FIGURE 3.** The architecture of the adversarial autoencoder network proposed in our method.

Apart from image-based approaches, projection-based approaches are popular for low-dose CBCT reconstruction. Projection-based approaches recover missing CBCT projections in the sinogram domain before CT reconstruction, which is entirely different from image-based approaches. Within all methods belonging to projection-based approaches, directional interpolation may be the most frequently employed one. For example, based on the structural tensor, Bertram *et al.* [33] presented an accurate non-linear directional interpolation to improve angular sampling in CBCT. Later, they developed another shape-driven directional interpolation algorithm for sparsely sampled cone-beam CT data [34]. To yield more projections for sparse angle acquisition in CBCT, Zhang *et al.* [35] extended Bertram's work by iteratively optimizing double-orientation estimation in the sinogram space. Additionally, Zhang *et al.* [36] reported a motion-weighted reconstruction method for insufficient projection in 4D CBCT reconstruction, in which a deformable registration was used to estimate local motion after directional interpolation increases the projection numbers.

Remarkably, there are no reports on the application of a GAN and its variants in the sinogram domain. Motived by this fact, we propose a limited-view CBCT reconstruction method in the sinogram domain (instead of the image domain) that can predict missing 3D CBCT projections in some scan views via a 2D adversarial autoencoder network. This adversarial network uses a CNN architecture [37] to construct a generator and a discriminator. To improve the prediction, we utilize joint loss, which can not only ensure the predicted missing parts of the sinogram images are consistent with the remainder but also capture critical high-frequency details from the ground-truth. In this way, the trained generator can derive the missing part of the sinogram of the newly acquired limited-view CBCT projection data. Finally, we reconstruct CT images from the completed

projections using the Feldkamp, Davis, and Kress (FDK) algorithm [38].

This paper is organized as follows. Section 2 describes the proposed limited-view CBCT reconstruction method. The experiments in Section 3 evaluate the performance of our approach. Finally, Section 4 gives the conclusions.

## II. METHODS

Fig. 1 illustrates the pipeline of the proposed limited-view CBCT reconstruction method, which consists of data preprocessing, training and testing. To make the 3D data feasible for the 2D network used in this paper, we perform data preprocessing, which converts 3D sinogram images of limited-view CBCT projections and their corresponding full-view ones to 2D slices before training begins. In the training stage, we train an adversarial autoencoder network [37] to produce the generator as the output. When one acquires the new limited-view CBCT projections, data preprocessing converts the new 3D sinogram images into 2D slices for testing. With the sinogram images of these 2D slices as inputs, the trained generator can predict their missing parts. After that, we stack the completed 2D slices to form full-view 3D CBCT projections. Using the FDK algorithm, our method reconstructs CT images from the completed 3D CBCT projections. The next subsections will describe the essential components of the proposed method in detail.

### A. DATA PREPROCESSING

Moving along a trajectory $\Gamma(\theta)$, the X-ray source in a regular CBCT acquisition system emits half-lines crossing the object and intersecting the 2D detector plane $T(x, y)$, which eventually yield 3D CBCT projections $C(\theta, x, y)$. Fig. 2 (a) and (b) illustrate the data matrix of full-view projections $C(\theta, x, y)$ and limited-view projections $C'(\theta', x, y)$. Note that $C(\theta, x, y)$ and $C'(\theta', x, y)$ correspond to the 3D sinogram image for full projections and limited-view projections, respectively.
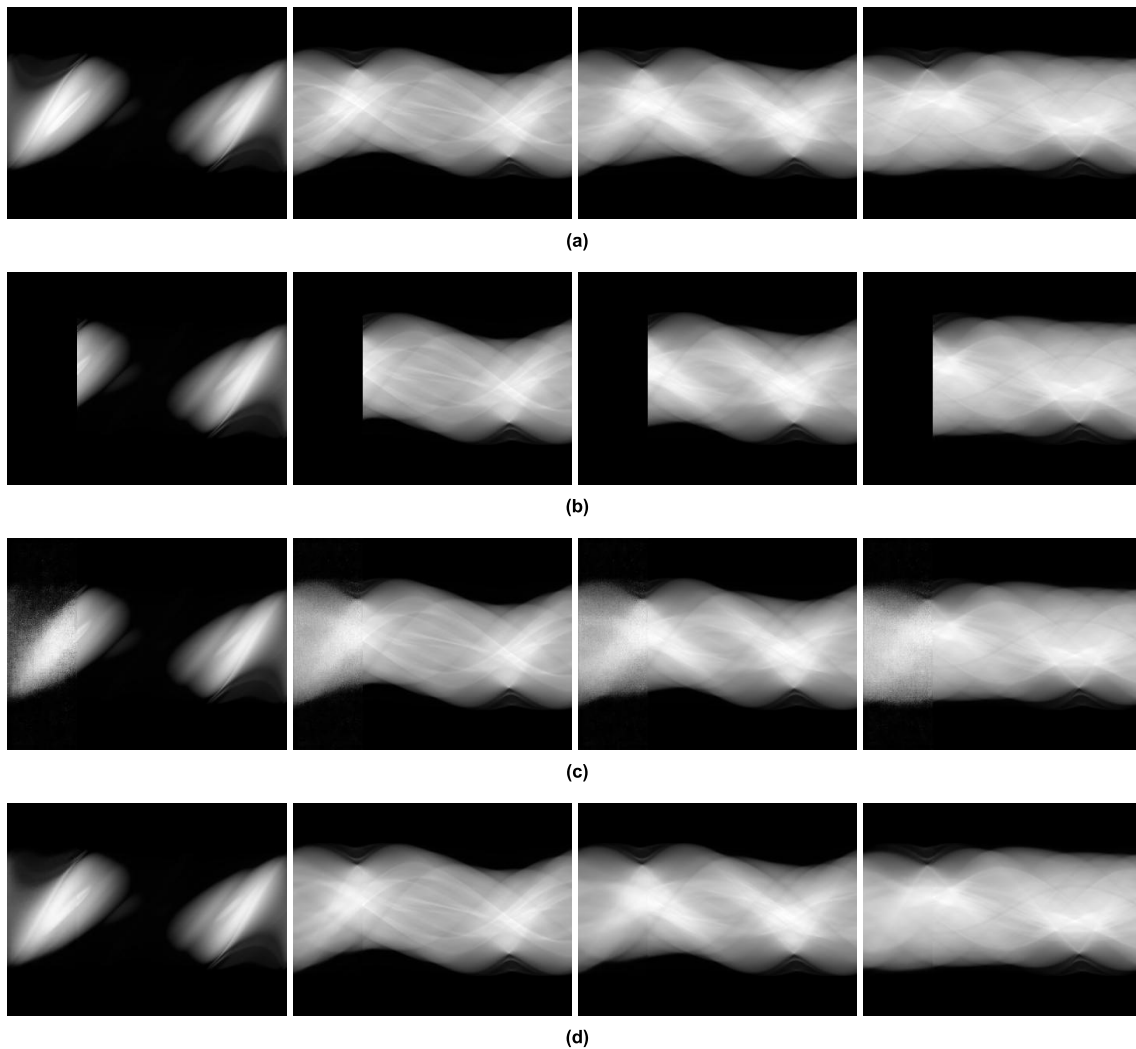
**FIGURE 4.** Resulting sinogram images of 2D slices along the x-axis after the missing parts have been predicted by GAN and the proposed method when the acquired CBCT projections at scan views ranging from 0° to 89° are unknown. (a) Sinogram images of 2D slices along the x-axis for full-view CBCT projections. (b) Sinogram images of 2D slices along the x-axis with unknown projections at scan views ranging from 0° to 89°. (c) Resulting sinogram images after missing parts have been predicted by GAN method. (d) Resulting sinogram images after missing parts have been predicted by the proposed method.

Since the adversarial autoencoder network used for prediction in our paper is two-dimensional, this network cannot handle 3D sinogram image directly. Therefore, it is necessary to slice the 3D sinogram image into 2D pieces for both the training and testing procedures. To accomplish this task, we first make the missing part of $C'(\theta', x, y)$, i.e., the dotted portion in Fig. 2 (b), equivalent to 0. Then, our method divides the 3D sinogram image $C'(\theta', x, y)$ into multiple 2D slices along the x-axis. The planes marked in red in Fig. 2 (b) represent these 2D slices from the 3D data matrix. In other words, if given a series of coordinates $x_n$ ($n = 0, 1, 2, \ldots, N$) on the x-axis, we can obtain multiple sinogram images of 2D slices $C'_{x_n}(\theta', y)$ after slicing. As these multiple 2D slices $C'_{x_n}(\theta', y)$ are compatible with the proposed 2D adversarial autoencoder network, we can generate the missing parts of $C'_{x_n}(\theta', y)$.

Before reconstructing 3D CBCT images, we must arrange the completed sinogram images of the 2D slices $C'_{x_n}(\theta', y)$ along the x-axis, as shown in Fig. 2 (c), to derive the full-view 3D CBCT projections.

## B. ADVERSARIAL AUTOENCODER NETWORK
Fig. 3 depicts the architecture of the adversarial autoencoder network used in our method. This network applies an autoencoder-like CNN model [37] that adopts the well-known AlexNet architecture [39] to train the generator. To make the prediction more accurate, we simultaneously train a discriminator by using predicted results and their corresponding ground-truth as input. The discriminator can provide loss gradients for the generator and give the possibility that the predicted results belong to real data. The training procedure does not stop until the discriminator accepts all
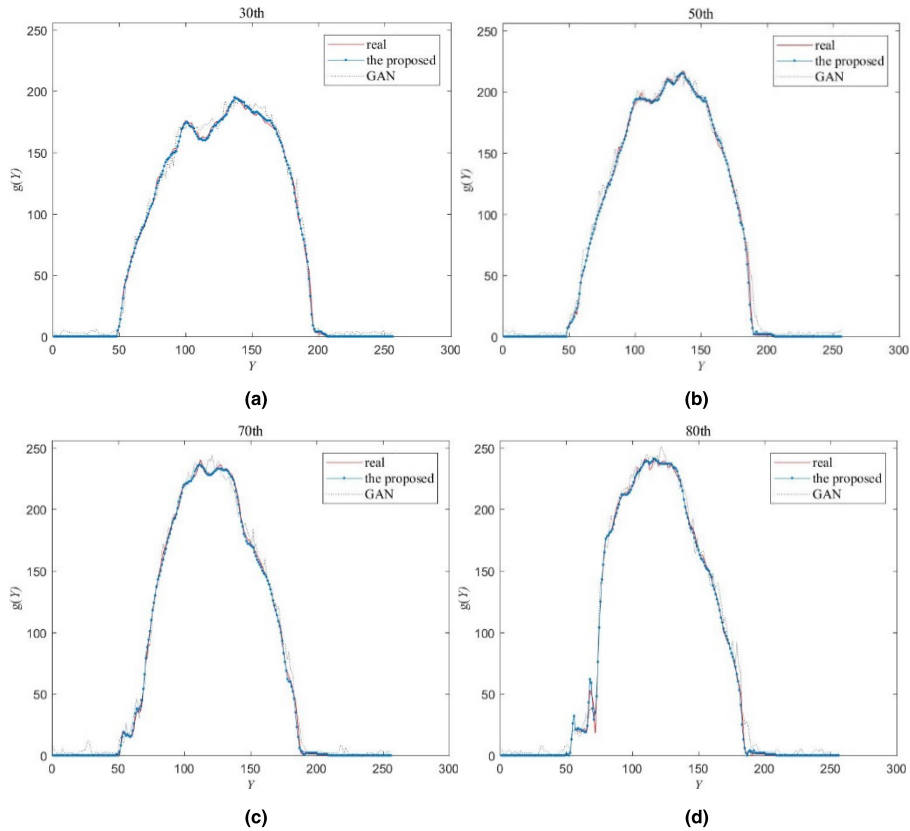
**FIGURE 5.** The vertical profiles across 30th, 50th, 70th and 80th columns of the resulting sinogram images when the acquired CBCT projections at scan views ranging from 0° to 89° are unknown. The resulting sinogram images come from the second column of Fig. 4(c)-(d). The red lines denote the real data; the purple dots and the dotted blue lines denote the vertical profiles of the resulting missing sinogram estimated by GAN and the proposed method.

the generated missing parts of the sinogram images of the limited-view CBCT projections as real ones.

The generator is composed of an encoder and a decoder. The encoder part has five convolutional layers and one fully-connected layer. All the five convolutional layers in the encoder part contain convolution, batch normalization (BN) and rectified linear unit (ReLU) operations. Specifically, the first convolutional layer filters the incomplete 2-D slices with 64 kernels of size $N_1 \times M_1$ and yields feature maps $F_1$. Next, by filtering the output of the first convolutional layer with 64 kernels of size $N_2 \times M_2$, the second convolutional layer contains feature maps $F_2$. Similarly, the third, fourth and fifth convolutional layer will acquire their feature maps $F_3$, $F_4$ and $F_5$ respectively through 2-D convolution on the output of the previous convolutional layer. Here, we let the kernel numbers for the third, fourth and fifth convolutional layers be 128, 256 and 512; and, the filter sizes for these three layers are $N_3 \times M_3$, $N_4 \times M_4$ and $N_5 \times M_5$ respectively. The fully-connected layer contains 4000 neurons which have connections to all the neurons in the fifth convolutional layer.

In the proposed networks, we do not use pooling layers because they will reduce the spatial resolution of feature maps and make the network unable to capture fine details in the sinograms precisely.

As for the decoder part of the generator network, its first layer obtains the feature maps, which have the same size as the fifth convolutional layer in encoder part, by fully connecting and reshaping the fully-connected layer of encoder part. The followings include three up-convolutional layers and one output layer. Exerting 2-D up-convolution with 256 kernels on the output of the first layer in decoder part, achieves feature maps $L_1$ in the first up-convolutional layer of the decoder part. Subsequently, by up-convolutional operation with 128 and 64 kernels on the output of the previous layer, the second and third up-convolutional layer derive the feature maps $L_2$ and $L_3$ respectively. Then, applying a 2-D up-convolution to the output of the third up-convolutional layer constructs the output layer, which corresponds to the predicted results of missing parts. The sizes of the filters used in the up-convolutional operations are $S_1 \times T_1$, $S_2 \times T_2$, $S_3 \times T_3$ and $S_4 \times T_4$. Notably, all the up-convolutional layers are followed by BN and ReLU operations, except the third one. The third up-convolutional layer adopts sigmoid as the activation function.

For the sake of making a generated result close to the real one, we train a discriminator alongside the generator. The discriminator takes both the predicted results and their ground-truth as inputs. Its output is the probability that the predicted results are drawn from the distribution of real ones.
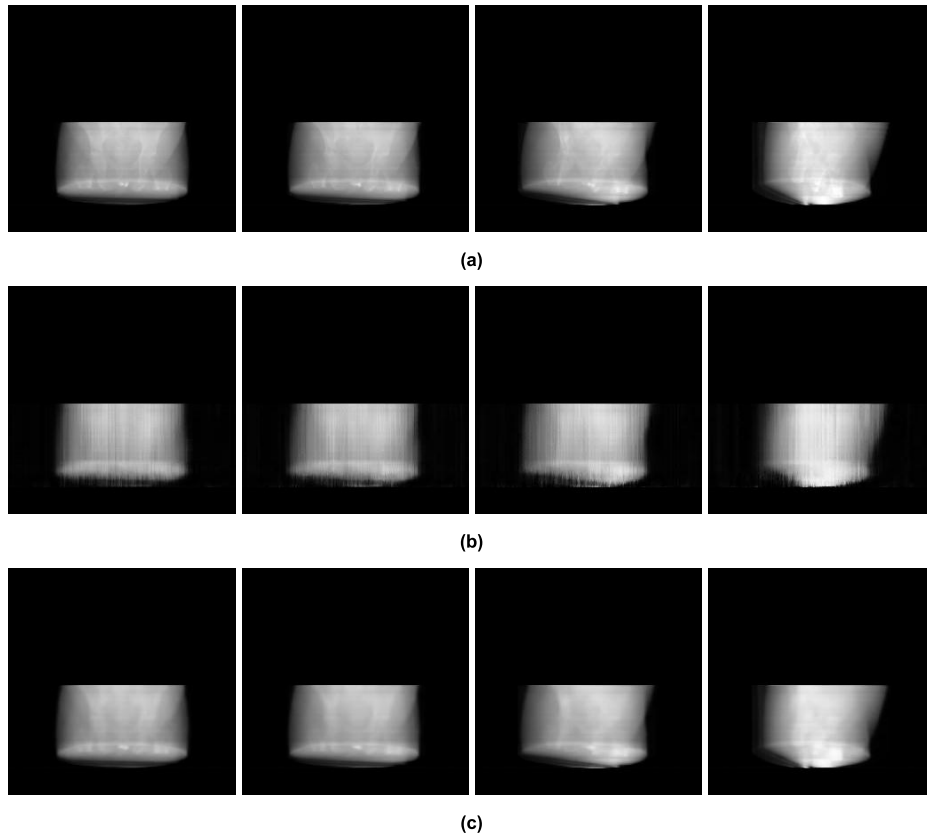
**FIGURE 6.** Predicted results for unknown CBCT projections at scan views of 7°, 13°, 30°, and 60° by GAN and the proposed method. Here, the acquired CBCT projections at scan views ranging from 0° to 89° are unknown. (a) 2D profiles of the full CBCT projections along the $\theta$-axis at scan views of 7°, 13°, 30°, and 60°. (b) Predicted results for unknown projections at scan views of 7°, 13°, 30°, and 60° by GAN method. (c) Predicted results for unknown projections at scan views of 7°, 13°, 30°, and 60° by the proposed method.

**TABLE 1.** Some parameters for each convolutional and up-convolutional operation involved in the generator and discriminator when the sizes of the sinogram image of a 2D slice and its missing part are 360 × 256 and 90 × 256 respectively. Here, Conv1~Conv5 denote the convolutional operations used in the generator; Conv6~Conv9 denote the convolutional operations used in the discriminator; uConv1~uConv4 denote the up-convolutional operations used in the discriminator.

|  | Filter size | Stride along x and y-axis (pixel) | Padding along x and y-axis (pixel) |
|---|---|---|---|
| Conv1 | 4×4 | (2, 4) | (1, 0) |
| Conv2 | 4×4 | (4, 2) | (0, 1) |
| Conv3 | 4×3 | (2, 3) | (1, 0) |
| Conv4 | 4×3 | (2, 3) | (1, 0) |
| Conv5 | 4×3 | (2, 2) | (1, 1) |
| Conv6 | 4×5 | (4, 5) | (0, 0) |
| Conv7 | 4×4 | (4, 2) | (0, 1) |
| Conv8 | 4×3 | (2, 3) | (1, 0) |
| Conv9 | 4×3 | (2, 1) | (1, 1) |
| uConv1 | 4×3 | (2, 1) | (1, 1) |
| uConv2 | 4×3 | (2, 3) | (1, 0) |
| uConv3 | 4×4 | (4, 2) | (0, 1) |
| uConv4 | 4×5 | (4, 5) | (0, 0) |

**TABLE 2.** Some parameters for each convolutional and up-convolutional operation involved in generator and discriminator when the sizes of the sinogram image of a 2D slice and its missing part are 360 × 256 and 120 × 256 respectively. Here, Conv1~Conv5 denote the convolutional operations used in the generator; Conv6~Conv9 denote the convolutional operations used in the discriminator; uConv1~uConv4 denote the up-convolutional operations used in the discriminator.

|  | Filter size | Stride along x and y-axis (pixel) | Padding along x and y-axis (pixel) |
|---|---|---|---|
| Conv1 | 4×4 | (2, 2) | (1, 1) |
| Conv2 | 4×4 | (2, 2) | (1, 1) |
| Conv3 | 4×4 | (2, 2) | (1, 1) |
| Conv4 | 4×3 | (2, 3) | (1, 0) |
| Conv5 | 4×3 | (4, 3) | (0, 0) |
| Conv6 | 4×4 | (4, 2) | (0, 1) |
| Conv7 | 4×4 | (2, 2) | (1, 1) |
| Conv8 | 4×4 | (2, 2) | (1, 1) |
| Conv9 | 4×3 | (4, 3) | (0, 0) |
| uConv1 | 4×3 | (4, 3) | (0, 0) |
| uConv2 | 4×4 | (2, 2) | (1, 1) |
| uConv3 | 4×4 | (2, 2) | (1, 1) |
| uConv4 | 4×4 | (4, 2) | (0, 1) |

The network of discriminator has similar architecture to the encoder part of the generator. It has four convolutional layers, which also include convolution, BN and ReLU operations, and one output layer. The numbers of filters used in the four convolutional operations for the discriminator are 64, 128, 256 and 512 respectively.
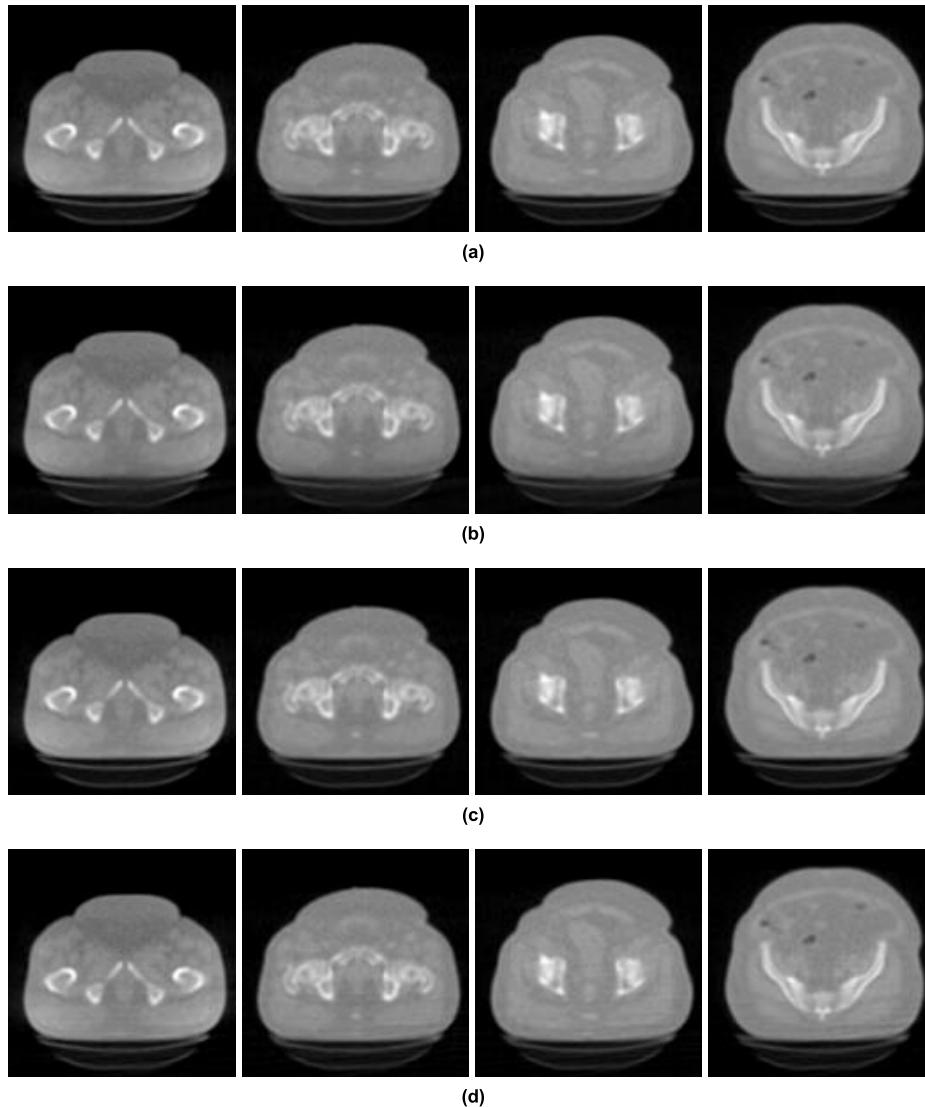
**FIGURE 7.** Comparison among transverse slices of the real scan (a) and CT images reconstructed from the resulting projections by different algorithms: (c) GAN and (d) our method. (b) are the results with missing projections set to 0. In the acquired limited-view CBCT projections, the projections at scan views ranging from 0° to 89° are unknown.

The filter sizes, stride and padding parameters for each convolutional and up-convolutional operations involved in generator and discriminator depend on the dimensions of the input sinogram images of the 2D slices, as well as their missing parts. Table 1 lists the filter sizes, stride and padding parameters when sizes of the sinogram image of a 2D slice and its missing part are $360 \times 256$ and $90 \times 256$ respectively. And, Table 2 provides the filter sizes, stride, and padding parameters when dimensions of the sinogram image of a 2D slice and its missing part are $360 \times 256$ and $120 \times 256$ respectively.

### C. JOINT LOSS FUNCTION
In the training procedure, the proposed method employs the joint loss function [37], which is comprised of reconstruction loss and adversarial loss, to ensure proper performance in prediction.

For an input sinogram image $I$ of a 2D slice, the following equation defines reconstruction loss $R(I)$ [37]:

$$R(I) = \|M \otimes (I - G((1 - M) \otimes I))\|_2 \qquad (1)$$

where $M$ is a binary mask for a missing region of $I$, $\otimes$ is the element-wise product operation, and $G(I)$ is the output of the generator for input data $I$. The reconstruction loss in (1) approximates the outline of the target well by minimizing the mean pixel-wise error. Nonetheless, reconstruction loss is deficient at catching high-frequency details, which make the predicted results look blurry. Thus, adversarial loss, whose definition is given in the next equation, participates in the training procedure as a remedy for the deficiencies of
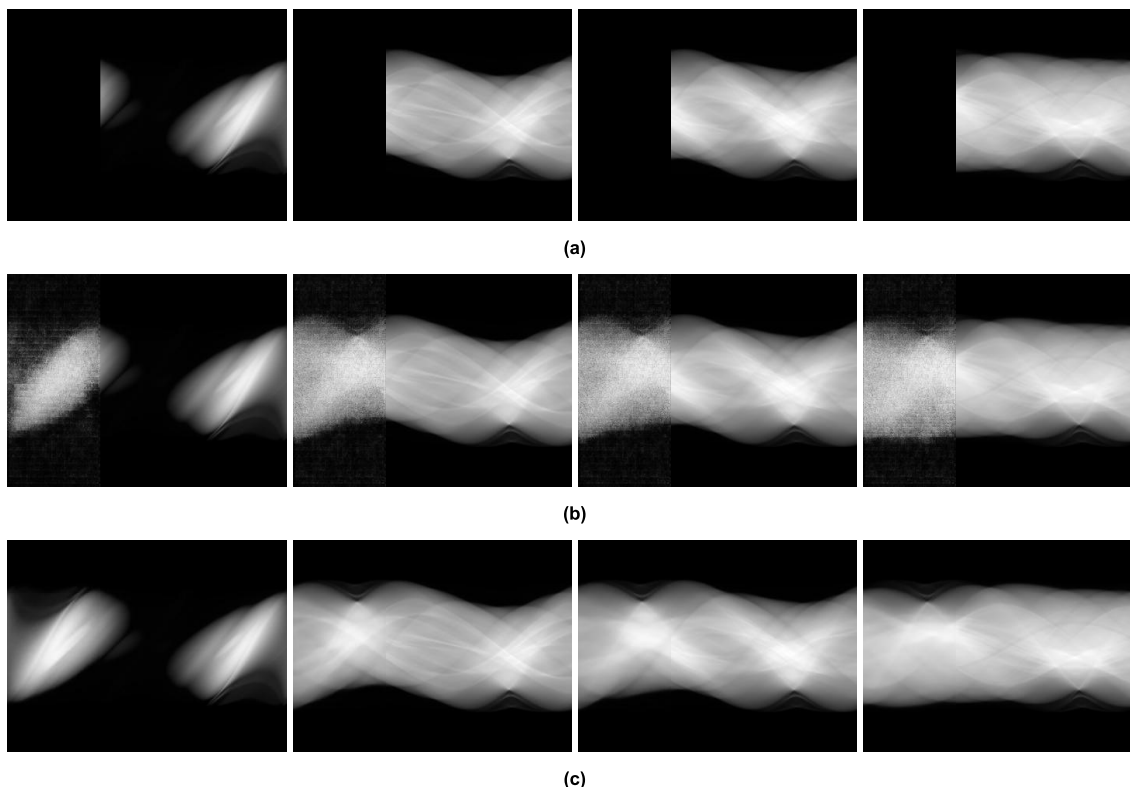
**FIGURE 8.** Resulting sinogram images of the 2D slices along the x-axis after missing parts have been predicted by the GAN and the proposed method when the acquired CBCT projections at scan views ranging from 0° to 119° are unknown. (a) Sinogram images of the 2D slices along the x-axis when projections at scan views ranging from 0° to 119° are unknown. (b) Resulting sinogram images after missing parts have been predicted by the GAN. (c) Resulting sinogram images after missing parts have been predicted by the proposed method.

reconstruction loss [37]:

$$A(I) = \max_{D} \mathrm{E}_{I \epsilon \chi} [\log(D(I)) + \log(1 - D(G((1-M) \otimes I)))]$$

(2)

where $D$ is the adversarial discriminative model. The adversarial loss in (2) can lessen the blur degradation that happens when training procedure only uses the reconstruction loss in (1). To optimize both the generative and discriminative models, we use the Adam algorithm [40] in the training procedure.

After integrating reconstruction loss with the adversarial loss, our method defines the joint loss function as follows:

$$J(I) = \lambda_r R(I) + \lambda_a A(I)$$

(3)

where $\lambda_r$ and $\lambda_a$ are the corresponding weights for reconstruction loss and adversarial loss, respectively. Since the joint loss function in (3) combines the merits of both reconstruction loss and adversarial loss, it can improve the prediction.

## III. EXPERIMENTS

This section releases some experiments to test the performance of our method. We conduct these experiments on a dataset which is collected by Jiangsu Province Key Laboratory of Oral Disease, Nanjing Medical University. The dataset

consists of 300 CT images as well as their corresponding full-view CBCT projections. The resolutions and image sizes of these CT images are $0.98 \times 0.98 \times 3$ mm$^3$ and $512 \times 512 \times 90$, respectively. And, the size of the full-view CBCT projection data is $360 \times 256 \times 256$, which implies that the dimension of the flat detector is $256 \times 256$ and the projections are sampled from 360 scan views over an arc of 360 degrees. The following experiments use 200 CT images and their corresponding projections for training. The testing step will use the remaining 100 CT images and their corresponding projections.

To compare the behavior with other methods, we also provide the experimental results of GAN [41]. The GAN uses 200 iterations, as does the proposed method.

To quantitatively evaluate the predicted projections and reconstructed CT images, this section chooses some common measurements, such as peak signal to noise ratio (PSNR) and structural similarity index (SSIM) [30]. PSNR assesses the noise and artifact suppression performance of a method, and SSIM measures the difference between the reconstructed images and their ground-truth.

The first experiment assumes an X-ray reduction of 25% to a full-view CBCT scan. More specifically, the CBCT projections at scan views ranging from 0° to 89° are unknown, and values of the missing projections are assigned to 0.
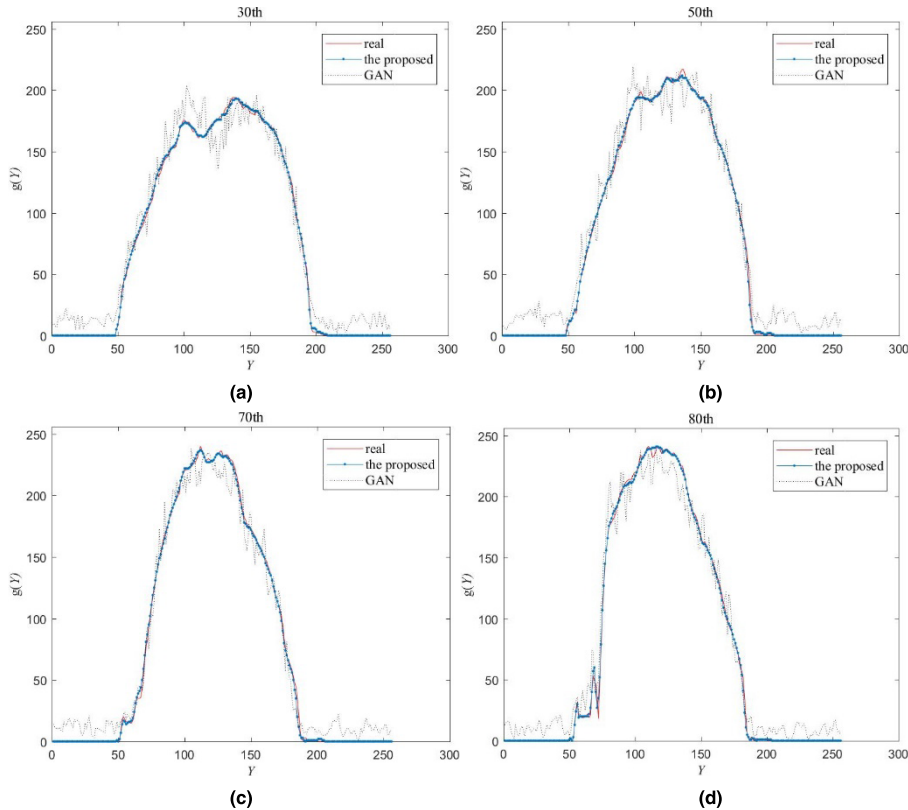
**FIGURE 9.** The vertical profiles across 30th, 50th, 70th and 80th columns of the resulting sinogram images when the acquired CBCT projections at scan views ranging from 0° to 119° are unknown. The resulting sinogram images come from the second column of Fig. 8(b)-(c). The red lines denote the real data; the purple dots and the dotted blue lines denote the vertical profiles of the resulting missing sinogram estimated by GAN and the proposed method.

Fig. 4 offers some examples of the resulting sinogram images of 2D slices along the x-axis after GAN and the proposed method predict the missing data. Fig. 4 (a) and (b) give the sinogram images of 2D slices along the x-axis for full-view and limited-view CBCT projections. Fig. 4 (c) and (d) show that compared to Fig. 4 (a), the resulting sinogram images with missing parts predicted by both GAN and the proposed method can achieve good outlines. However, the results in Fig. 4 (c) look blurry and noisy, which leads to inconsistency between the predicted results and the remainder of the sinogram images. This inconsistency might be the result of the discriminator of the GAN treating the missing parts of the data in isolation and ignoring the contextual information from the rest of the sinogram. Since the joint loss function adopted in our method tries to make the entire output look realistic, the appearance defect mentioned above does not appear in the results of Fig. 4 (d).

For better showing the performance improvement of our method, Fig. 5 gives the vertical profiles across 30th, 50th, 70th and 80th columns of the resulting sinogram images, which come from the second column of Fig. 4(c)-(d).

From Fig.5, one can observe that the curves of the estimated results by both GAN and the proposed method fit the real data well; our method has a little advantage only in the middle part of the curves where they fluctuate greatly.

To make a further comparison, we also provide the predicted results for unknown CBCT projections at scan views of 7°, 13°, 30°, and 60° using different methods, as shown in Fig. 6. While all results obtained by the GAN in Fig. 6 (b) suffer from vertical streak artifacts, no severe artifacts exist in the results of our method, as shown in Fig. 6 (c). The results in Fig. 4, Fig. 5 and Fig. 6 suggest that the proposed method can capture more high-frequency details and has better predictive performance than GAN.

For all predicted results of the missing CBCT projections, the mean PSNRs of GAN and our method are 35.98 and 44.37, respectively. The higher PSNR achieved by the proposed method confirms its superiority to GAN.

Fig. 7 provides a comparison among some transverse slices of the real scans and CBCT images reconstructed from the resulting projections by GAN and our method. For the purpose of showing the degree of improvement of the reconstructed CT images by our method, we also give some examples of the reconstructed CT images with missing CBCT projections being 0 in Fig. 7 (b). In other words, to obtain the results in Fig. 7 (b), we assign 0 to values of the missing CBCT projections and then derive the reconstructed CT images by using FDK method. In Fig. 7 (d), the reconstructed images by the proposed method are comparable to the original images; however, they do not show notable improvement
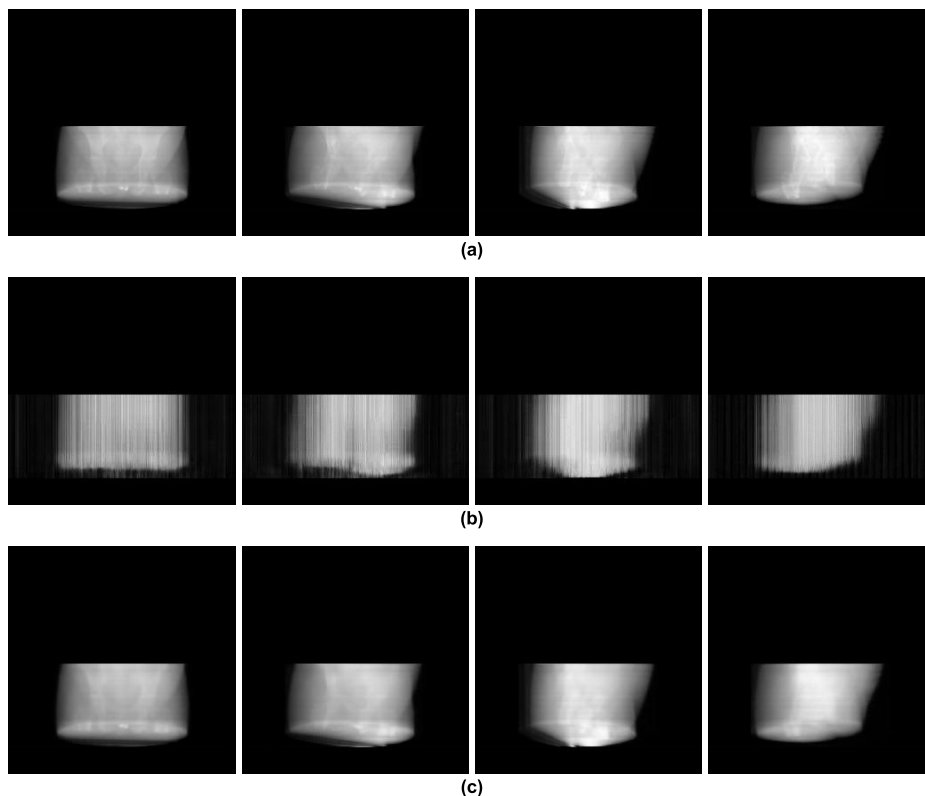
**FIGURE 10.** Predicted results for unknown CBCT projections at scan views of 7°, 13°, 30°, and 60° by the GAN and the proposed method. Here, the acquired CBCT projections at scan views ranging from 0° to 119° are unknown. (a) 2D profiles of full CBCT projections along the $\theta$-axis at scan views of 7°, 13°, 30°, and 60°. (b) Predicted results for unknown projections at scan views of 7°, 13°, 30°, and 60° by the GAN. (c) Predicted results for unknown projections at scan views of 7°, 13°, 30°, and 60° by the proposed method.

**TABLE 3.** Mean PSNR and SSIM for all reconstructed CT images by GAN and our method when the acquired CBCT projections at scan views ranging from 0° to 89° are unknown. Here, 'ZERO' denotes mean PSNR and SSIM for the reconstructed CT images when the missing CBCT projections are set to 0.

|  | ZERO | GAN | Our method |
|---|---|---|---|
| mean PSNR | 30.50 | 38.66 | 44.43 |
| mean SSIM | 0.92 | 0.97 | 0.99 |

in visual appearance in contrast to Fig. 7 (b)-(c). It is noteworthy that the main difference between the results in Fig. 7 (b) and the other two groups of results in Fig. 7 (c)-(d) is whether there is an estimation of the missing CBCT projections or not.

Table 3 lists the mean PSNR and SSIM for all reconstructed CT images by the GAN and our method, as well as the ones when missing CBCT projections are set to 0. Table 3 implies that although the mean PSNR and SSIM of the results obtained by our method are higher than the other two, the proposed method wins by a narrow margin.

To test the performance of our method with more missing projections, we assume the CBCT projections at scan views ranging from 0° to 119° unknown in the next paragraphs.

That is, the number of the acquired projections in the second experiment is decreased by 1/3 over a full-view CBCT scan.

Fig. 8 shows the resulting sinogram images of the 2D slices along the x-axis after the missing parts have been predicted by various methods when the projections at scan views ranging from 0° to 119° are unknown. Since the number of the unknown projections increases, the resulting sinogram images by the GAN, which are given in Fig. 8 (b), become noisier than the results in Fig. 4 (c), as expected. However, compared to the results in Fig. 4 (d), no noticeable degradation is observed from the resulting sinogram images by our method, as shown in Fig. 8 (c).

Fig. 9 gives the vertical profiles across the same columns of another group of resulting sinogram images, which come from the second column of Fig. 8(b)-(c). Compared to Fig.5, the advantage of our method over GAN is more obvious in Fig.9, as the curves of the estimated results by GAN in Fig.9 become volatile.

The same situation occurs in Fig. 10, which shows the predicted results for the unknown CBCT projections at scan views of 7°, 13°, 30°, and 60° by the GAN and our method. The mean PSNR values of the predicted CBCT projections by the GAN and our method when 1/3 of the projections are missing are 28.40 and 42, respectively. They decrease
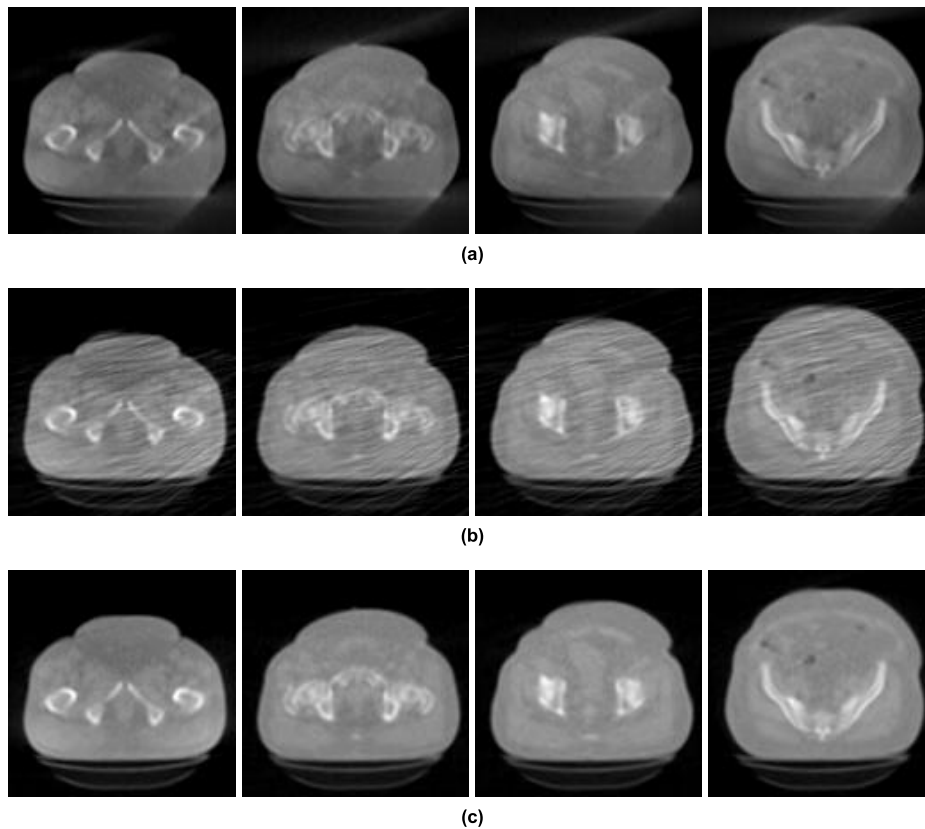
**FIGURE 11.** Comparison of transverse slices of the CBCT images reconstructed from the resulting projections by different algorithms when the acquired projections at scan views ranging from 0° to 119° are unknown. (a) are the results with missing projections set to 0. (b) are the reconstructed CT images with missing projections predicted by GAN; (c) are the reconstructed CT images with missing projections predicted by our method.

by approximately 21% and 5%, respectively, in comparison to the corresponding mean PSNR of the predicted CBCT projections when 1/4 of the projections are missing. Quantitatively, we find that the performance of our method in terms of predicting missing CBCT projections is less vulnerable to the increasing number of unknown projections than the GAN.

Fig. 11 (a) shows some transverse slices of the reconstructed CBCT images when we let the missing projections at scan views ranging from 0° to 119° equal to 0. Fig. 11 (b)-(c) give the transverse slices of the reconstructed CBCT images when GAN and our method estimate the missing projections at scan views ranging from 0° to 119°. Predictably, with more missing CBCT projections, the reconstructed CT images in Fig. 11 (a) have a much greater degradation than those in Fig. 7 (b); and, in contrast to Fig. 7 (c), the reconstructed results in Fig. 11 (b) suffer a severe deterioration. However, when our method estimates the missing projections, the corresponding reconstructed results in Fig. 11 (c) are comparable to the results in Fig. 7 (d). It suggests that the increasing number of missing projections affects the performance of our method less than those of GAN.

This conclusion is also proved by Table 4, which gives the mean PSNR and mean SSIM for all reconstructed CT

**TABLE 4.** Mean PSNR and mean SSIM for all reconstructed CT images by the GAN and our method when the acquired CBCT projections at scan views ranging from 0° to 119° are unknown. Here, 'ZERO' denotes mean PSNR and SSIM for the reconstructed CT images when the missing CBCT projections are set to 0.

|           | ZERO  | GAN   | Our method |
|-----------|-------|-------|------------|
| mean PSNR | 20.09 | 28.02 | 38.86      |
| mean SSIM | 0.74  | 0.77  | 0.97       |

images by GAN and our method when the acquired CBCT projections at scan views ranging from 0° to 119° are unknown. Table 4 also offers mean PSNR and mean SSIM for all reconstructed CT images with the missing projections being 0. Table 4 shows that, although all mean PSNRs and mean SSIMs decrease with an increasing number of missing projections, the results of our method remain the highest. In contrast to Table 3, the mean PSNRs of the GAN and the proposed method in Table 4 fall by 27.52% and 12.53%, respectively. As for mean SSIM, the values obtained by the GAN and the proposed method in Table 4 fall by 20.62% and 2%, respectively. It is clear that the decreasing rates of GAN for both mean PSNR and mean SSIM are far higher than those of the proposed method, especially for mean SSIM.

Based on the above experiments, we can conclude that our method has better and more stable performance than the other methods, particularly when the acquired projections are reduced by 1/3 over a full-view CBCT scan.

## IV. CONCLUSIONS

To improve the quality of the CBCT images reconstructed from limited-view projections, we develop a CBCT reconstruction method in the sinogram domain. In data preprocessing, the proposed method slices the 3D CBCT projections into 2D pieces along the x-axis. Then, it constructs a 2D adversarial autoencoder network that uses a CNN architecture with the joint loss to build the generator and discriminator and predict the missing 3D CBCT projections. The experiments validate the effectiveness of our method in estimating missing CBCT projections and reconstructing images.

In the future, we will extend our method to a 3D version to improve the predicted results further. The method proposed in this paper has to slice 3D CBCT projections into multiple 2D pieces along the x-axis, instead of handling 3D CBCT projections directly, before using the adversarial auto-encoder network to estimate the missing data in each sliced 2D piece of 3D CBCT projections. Thus, we do not incorporate any spatial or appearance information from the other 2D slices adjacent to the current one into the estimation. It would lead to inconsistency across slices after stacking the completed 2D pieces into full-view 3D CBCT projections. We believe that using a 3D adversarial autoencoder network might be helpful to settle down this problem.

## REFERENCES

[1] C. Walter, J. C. Schmidt, K. Dula, and A. Sculean, "Cone beam computed tomography (CBCT) for diagnosis and treatment planning in periodontology: A systematic review," *Quintessence Int.*, vol. 47, no. 1, pp. 25–37, Jan. 2016.

[2] A. Gupta, O. P. Kharbanda, V. Sardana, R. Balachandran, and H. K. Sardana, "A knowledge-based algorithm for automatic detection of cephalometric landmarks on CBCT images," *Int. J. Comput. Assist. Radiol. Surg.*, vol. 10, no. 11, pp. 1737–1752, Nov. 2015.

[3] B. Bapst, M. Lagadec, R. Breguet, V. Vilgrain, and M. Ronot, "Cone beam computed tomography (CBCT) in the field of interventional oncology of the liver," *CardioVascular Interventional Radiol.*, vol. 39, no. 1, pp. 8–20, Jan. 2016.

[4] C. Y. S. Lee, T. M. Koval, and J. B. Suzuki, "Low-dose radiation risks of computerized tomography and cone beam computerized tomography: Reducing the fear and controversy," *J. Oral Implantol.*, vol. 41, no. 5, pp. e223–e230, Oct. 2015.

[5] A. A. Sánchez, "Estimation of noise properties for TV-regularized image reconstruction in computed tomography," *Phys. Med. Biol.*, vol. 60, no. 18, pp. 7007–7033, Aug. 2015.

[6] X. Han *et al.*, "Algorithm-enabled low-dose micro-CT imaging," *IEEE Trans. Med. Imag.*, vol. 30, no. 3, pp. 606–620, Mar. 2011.

[7] J. S. Jorgensen, E. Y. Sidky, and X. Pan, "Quantifying admissible undersampling for sparsity-exploiting iterative image reconstruction in X-ray CT," *IEEE Trans. Med. Imag.*, vol. 32, no. 2, pp. 460–473, Feb. 2013.

[8] J. Ma *et al.*, "Iterative image reconstruction for cerebral perfusion CT using a pre-contrast scan induced edge-preserving prior," *Phys. Med. Biol.*, vol. 57, no. 22, p. 7519, Nov. 2012.

[9] Y. Zhang, Y. Xi, Q. Yang, W. Cong, J. Zhou, and G. Wang, "Spectral CT reconstruction with image sparsity and spectral mean," *IEEE Trans. Comput. Imag.*, vol. 2, no. 4, pp. 510–523, Dec. 2016.

[10] 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, p. 4777, Sep. 2008.

[11] G.-H. Chen *et al.*, "Time-resolved interventional cardiac C-arm cone-beam CT: An application of the PICCS algorithm," *IEEE Trans. Med. Imag.*, vol. 31, no. 4, pp. 907–923, Oct. 2011.

[12] M. G. Lubner, P. J. Pickhardt, J. Tang, and G.-H. Chen, "Reduced image noise at low-dose multidetector CT of the abdomen with prior image constrained compressed sensing algorithm," *Radiology*, vol. 260, no. 1, pp. 248–256, Mar. 2011.

[13] Q. Xu, H. Yu, X. Mou, L. Zhang, J. Hsieh, and G. Wang, "Low-dose X-ray CT reconstruction via dictionary learning," *IEEE Trans. Med. Imaging*, vol. 31, no. 9, pp. 1682–1697, Sep. 2012.

[14] Y. Chen *et al.*, "Artifact suppressed dictionary learning for low-dose CT image processing," *IEEE Trans. Med. Imag.*, vol. 33, no. 12, pp. 2271–2292, Dec. 2014.

[15] Y. Chen *et al.*, "Improving abdomen tumor low-dose CT images using a fast dictionary learning based processing," *Phys. Med. Biol.*, vol. 58, pp. 5803–5820, Aug. 2013.

[16] T. Huynh *et al.*, "Estimating CT image from MRI data using structured random forest and auto-context model," *IEEE Trans. Med. Imag.*, vol. 35, no. 1, pp. 174–183, Jan. 2016.

[17] D. Boublil, M. Elad, J. Shtok, and M. Zibulevsky, "Spatially-adaptive reconstruction in computed tomography using neural networks," *IEEE Trans. Med. Imag.*, vol. 34, no. 7, pp. 1474–1485, Jul. 2015.

[18] H. Zhang *et al.*, "Extracting information from previous full-dose CT scan for knowledge-based Bayesian reconstruction of current low-dose CT images," *IEEE Trans. Med. Imag.*, vol. 35, no. 3, pp. 860–870, Nov. 2015.

[19] G. Litjens *et al.*, "A survey on deep learning in medical image analysis," *Med. Image Anal.*, vol. 42, pp. 60–88, Dec. 2017.

[20] D. Shen, G. Wu, and H. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221–248, Jun. 2017.

[21] H. Greenspan, B. V. Ginneken, and R. M. Summers, "Guest editorial deep learning in medical imaging: Overview and future promise of an exciting new technique," *IEEE Trans. Med. Imag.*, vol. 35, no. 5, pp. 1153–1159, Mar. 2016.

[22] D. Nie, X. Cao, Y. Gao, L. Wang, and D. Shen, "Estimating CT image from MRI data using 3D fully convolutional networks," in *Medical Image Computing and Computer-Assisted Intervention—MICCAI*. Athens, Greece: Springer, 2016, pp. 170–178.

[23] E. Kang, J. Min, and J. C. Ye, "A deep convolutional neural network using directional wavelets for low-dose X-ray CT reconstruction," *Med. Phys.*, vol. 44, no. 10, pp. e360–e375, Oct. 2017.

[24] W. Yang *et al.*, "Improving low-dose CT image using residual convolutional network," *IEEE Access*, vol. 5, pp. 24698–24705, 2017.

[25] H. Chen *et al.*, "Low-dose CT via convolutional neural network," *Biomed. Opt. Express*, vol. 8, no. 2, pp. 679–694, 2017.

[26] H. Zhang *et al.* (Jul. 2016). "Image prediction for limited-angle tomography via deep learning with convolutional neural network." [Online]. Available: https://arxiv.org/abs/1607.08707

[27] H. Chen *et al.*, "Low-dose CT with a residual encoder-decoder convolutional neural network," *IEEE Trans. Image Process.*, vol. 36, no. 12, pp. 2524–2535, Dec. 2017.

[28] A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, B. Sengupta, and A. A. Bharath, "Generative adversarial networks: An overview," *IEEE Signal Process. Mag.*, vol. 35, no. 1, pp. 53–65, Jan. 2018.

[29] J. M. Wolterink, T. Leiner, M. A. Viergever, and I. Išgum, "Generative adversarial networks for noise reduction in low-dose CT," *IEEE Trans. Med. Imag.*, vol. 36, no. 12, pp. 2536–2545, Dec. 2017.

[30] Q. Yang *et al.*, "Low-dose CT image denoising using a generative adversarial network with wasserstein distance and perceptual loss," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1348–1357, Jun. 2018.

[31] H. Shan *et al.*, "3-D convolutional encoder-decoder network for low-dose CT via transfer learning from a 2-D trained network," *IEEE Trans. Med. Imag.*, vol. 37, no. 6, pp. 1522–1534, Jun. 2018.

[32] D. Nie *et al.*, "Medical image synthesis with context-aware generative adversarial networks," in *Medical Image Computing and Computer Assisted Intervention—MICCAI*. Quebec City, QC, Canada: Springer, 2017, pp. 417–425.

[33] M. Bertram, G. Rose, D. Schafer, J. Wiegert, and T. Aach, "Directional interpolation of sparsely sampled cone-beam CT sinogram data," in *Proc. 2nd IEEE Int. Symp. Biomed. Imag., Nano Macro*, Arlington, VA, USA, vol. 1, Apr. 2004, pp. 928–931.

[34] M. Bertram, J. Wiegert, D. Schafer, T. Aach, and G. Rose, "Directional view interpolation for compensation of sparse angular sampling in cone-beam CT," *IEEE Trans. Med. Imag.*, vol. 28, no. 7, pp. 1011–1022, Jul. 2009.

[35] H. Zhang, L. Ouyang, J. Ma, J. Huang, W. Chen, and J. Wang, "Noise correlation in CBCT projection data and its application for noise reduction in low-dose CBCT," *Med. Phys.*, vol. 41, no. 3, p. 031906, Mar. 2014.

[36] H. Zhang, M. Kruis, and J.-J. Sonke, "Directional sinogram interpolation for motion weighted 4D cone-beam CT reconstruction," *Phys. Med. Biol.*, vol. 62, no. 6, pp. 2254–2275, Jan. 2017.

[37] D. Pathak, P. Krähenbühl, J. Donahue, T. Darrell, and A. A. Efros, "Context encoders: Feature learning by inpainting," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Las Vegas, NV, USA, Jun. 2016, pp. 2536–2544.

[38] J. Baek and N. J. Pelc, "Local and global 3D noise power spectrum in cone-beam CT system with FDK reconstruction," *Med. Phys.*, vol. 38, no. 4, pp. 2122–2131, Apr. 2011.

[39] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," in *Proc. NIPS*, vol. 1, 2012, pp. 1097–1105.

[40] D. P. Kingma and J. Ba. (Dec. 2014). "Adam: A method for stochastic optimization." [Online]. Available: https://arxiv.org/abs/1412.6980

[41] I. J. Goodfellow *et al.*, "Generative adversarial nets," in *Proc. 27th Int. Conf. Neural Inf. Process. Syst.*, Montreal, QC, Canada, vol. 2, 2014, pp. 2672–2680.

**JIANAN BAI** was born in Qiqihaer, Heilongjiang, China, in 1993. He received the B.S. degree in electronic and information engineering from Nantong University, China, in 2015.
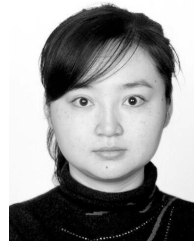
He is currently pursuing the master's degree with the Nanjing University of Posts and Telecommunications. His research interests include medical image processing and analysis.



**TIANLIANG LIU** was born in Xinfeng, Jiangxi, China, in 1980. He received the B.S. degree in telecommunication engineering from Jiangxi Normal University, China, in 2003, and the M.S. and Ph.D. degrees in biomedical engineering from Southeast University, China, in 2010.

Since 2009, he has been an Assistant Professor with the School of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications. His research interests include pattern recognition and image processing.



**XIUBIN DAI** was born in Zhenjiang, Jiangsu, China, in 1980. He received the B.S. degrees in biomedical engineering from the Huazhong University of Science and Technology, China, in 2002, and the M.S. and Ph.D. degrees in biomedical engineering from Southeast University, China, in 2009. From 2008 to 2010, he held a Postdoctoral position at Southeast University.

Since 2010, he has been an Assistant Professor with the School of Geographic and Biologic Information, Nanjing University of Posts and Telecommunications. His research interests include medical image reconstruction, medical image segmentation, and pattern recognition.



**LIZHE XIE** was born in Chengdu, Sichuan, China, in 1985. She received the B.S. and M.S. degrees in biomedical engineering from Sichuan University, China, in 2007, and the Ph.D. degree in biomedical engineering from Southeast University, China, in 2012.

Since 2012, she has been with the Jiangsu Province Key Laboratory of Oral Disease, Nanjing Medical University. Her research interests include medical image processing and analysis.

● ● ●