Generative adversarial network-based sinogram super-resolution for computed tomography imaging

Chao Tang®, Wenkun Zhang®, Linyuan Wang, Ailong Cai, Ningning Liang, Lei Li and Bin Yan

Henan Key Laboratory of Imaging and Intelligent Processing, PLA Strategic Support Force Information Engineering University, Zhengzhou, People’s Republic of China

E-mail: ybspace@hotmail.com

Keywords: CT image reconstruction, super resolution, projection domain, generative adversarial network

Abstract

Compared with the conventional 1 × 1 acquisition mode of projection in computed tomography (CT) image reconstruction, the 2 × 2 acquisition mode improves the collection efficiency of the projection and reduces the x-ray exposure time. However, the collected projection based on the 2 × 2 acquisition mode has low resolution (LR) and the reconstructed image quality is poor, thus limiting the use of this mode in CT imaging systems. In this study, a novel sinogram-super-resolution (SR) generative adversarial network model is proposed to obtain high-resolution (HR) sinograms from LR sinograms, thereby improving the reconstruction image quality under the 2 × 2 acquisition mode. The proposed generator is based on the residual network for LR sinogram feature extraction and SR sinogram generation. A relativistic discriminator is designed to render the network capable of obtaining more realistic SR sinograms. Moreover, we combine the cycle consistency loss, sinogram domain loss, and reconstruction image domain loss in the total loss function to supervise SR sinogram generation. Then, a trained model can be obtained by inputting the paired LR/HR sinograms into the network. Finally, the classic filtered-back-projection reconstruction algorithm is used for CT image reconstruction based on the generated SR sinogram. The qualitative and quantitative results of evaluations on digital and real data illustrate that the proposed model not only obtains clean SR sinograms from noisy LR sinograms but also outperforms its counterparts.

1. Introduction

Computed tomography (CT) technology has been widely applied to medical imaging (McCollough et al 2015), industrial non-destructive testing (Izumi et al 1993), and safety inspection (Mouton and Breckon 2015). In practice, the fidelity of CT image features may be enhanced by high-resolution (HR) imaging, which can better guide high-accuracy radiotherapy (Jaffray et al 2002). However, the resolution of reconstruction images in a CT imaging system is constrained by factors such as the detector element pitch, the focal spot size of the x-ray, and image reconstruction algorithms. In some important applications, such as early tumor characterization and coronary artery analysis, the inherent resolution of modern CT imaging systems is still considerably lower than the ideal resolution (Hassan et al 2010). More sophisticated hardware components can be used to improve the resolution, e.g. a detector element with a small pitch, an x-ray tube with a fine focal spot size, or better mechanical precision for CT scanning. However, these hardware-oriented methods will increase the CT system costs and radiation dose and compromise imaging speeds. Thus, developing methods to obtain higher resolution images without changing the original hardware system has gained significant research attention (Greenspan 2009).

In particular, some flat panel detectors such as Varian 2520D and 4030CB are equipped with two acquisition modes, namely 1 × 1 and 2 × 2 acquisition modes (https://www.varian.com), in most cone-beam CT imaging systems. The 2 × 2 acquisition mode enables fast scanning and improves projection collection efficiency because this mode has a higher acquisition frame rate. It can also significantly reduce the x-ray exposure time and the radiation dose under the condition of the same signal-to-noise ratio as that in the 1 × 1
acquisition mode. However, the $2 \times 2$ acquisition mode is rarely adopted by researchers because it obtains a low-resolution (LR) projection, resulting in the low quality of the reconstructed image. One possible strategy to improve the quality of the CT image under the $2 \times 2$ acquisition mode is to develop a method of super-resolution (SR) of the projection in order to obtain HR projection from LR projection.

In recent years, various SR methods have been proposed for improving the image resolution. These methods can be broadly divided into model- and learning-based approaches. The model-based approaches explicitly simulate the process of actually obtaining LR images and regularize the reconstruction on the basis of the prior information of image data (Sauer and Bouman 1993, Bouman and Sauer 1996, Farsiu et al 2004, Thibault et al 2007, Zhou et al 2011, Zhang et al 2013). These models can be broadly described as follows:

$$y^* = \arg\min_y \frac{1}{2} \| D_1 B * y - x \|_2^2 + \lambda R(y)$$

where $y$ denotes the HR image, $x$ denotes the LR image, * denotes the convolution operation, and $D_1 B$ denotes the down-sampling and blurring operations. The functional $R(\cdot)$ denotes the prior information, and multiplying it with parameter $\lambda$ sets the relative weights between two penalties. These methods promise excellent reconstruction image quality under the condition that the image priors are effectively mined. Learning-based approaches can recover the missing high-frequency image details by learning a nonlinear LR–HR mapping based on the training datasets. The approaches based on sparse representation also attract an increasing interest because it shows good performance in suppressing noise and artifacts (Yang et al 2010, Yu et al 2012, Yang and Yang 2013, Jiang et al 2018). Yang and Yang (2013) proposed a sparse representation-based SR reconstruction method, which needs to learn only two concise dictionaries and can obtain excellent results in terms of the peak signal-to-noise ratio (PSNR). Moreover, Yu et al (2012) developed a patch-based sparse representation SR method for lung CT images and obtained good performance in artifact suppression. Although these methods can enhance the quality of images, the results may have blocky artifacts and lose some subtleties.

Recently, deep learning (DL) techniques have been developed for computer vision tasks (Lecun et al 2015). In natural image SR tasks, convolutional neural networks (CNNs) have shown excellent performance (Dong et al 2014, Kim et al 2016, Chao et al 2016, Lai et al 2017). In 2014, a three-layer CNN, named SRCNN, was first proposed by Dong et al (2014). It learns a nonlinear LR–HR mapping in an end-to-end manner, exhibiting superior performance compared to traditional methods. In 2016, Wang (2016) reported that DL technology is expected to promote further development of CT imaging technology. Then, DL-based SR methods were explored by researchers for CT imaging (Yu et al 2017, Park et al 2018, Hatvani et al 2018, Umehara et al 2018, You et al 2020). Yu et al (2017) proposed two advanced CNN-based models to preserve high-frequency textures. Park et al (2018) proposed a CNN-based approach that consists of a contracting path and a symmetric expanding path to produce SR CT images. More recently, You et al (2020) proposed a semi-supervised network model based on generative adversarial networks (GANs) to generate HR CT images. These methods mainly address the SR problem of reconstructed image; however, the approaches of SR in the projection domain are not fully explored. In the $2 \times 2$ acquisition mode, the LR property of raw projection data is extended to the image domain in the reconstruction process, which intrinsically restricts the above algorithms to obtain high-quality results. To improve the quality of CT reconstruction images under the $2 \times 2$ acquisition mode, adding the effective information of the projection domain by sinogram SR is a useful approach.

The main challenges in the sinogram SR problem are as follows. First, the spatial variations, correlations, and statistical properties of the sinograms and CT images differ, limiting the existing SR methods to some extent. Second, since the blurring operations and down-sampling are ill-posed and coupled in the process of projection acquisition, it is difficult to build a numerical model to accurately depict the projection degradation process. Finally, raw projection data are sensitive, with a small change in them having a significant effect on the reconstruction image quality; therefore, there is a requirement for higher accuracy of the SR algorithm. To address these limitations, we propose a GAN-based sinogram SR method in this study. Inspired by the cycle-GAN (Zhu et al 2017, Yuan et al 2018), we propose a novel sinogram-SR GAN model (SSR-GAN). The cycle-consistent structure design of the SSR-GAN impels the network to learn the mapping relationship between the LR and HR sinograms. In addition, we design a generator based on the residual CNN, which can weaken the gradient vanishing problem. Further, we utilize the relativistic discriminator, which facilitates the generator to produce HR sinogram more efficiently. In particular, we design a filtered-back-projection (FBP) module to achieve cross-domain error backpropagation. The generated SR sinogram is more accurate under the constraint of the weighted image domain loss. The LR–HR mapping is learned by training the SSR-GAN with paired LR/HR sinograms. Then, the classic FBP algorithm (Katsevich 2001) is used to obtain a high-quality CT image from the SR sinogram. The evaluation results demonstrate
that our proposed model exhibits excellent sinogram SR performance and effectively reduces artifacts in CT images.

The rest of this paper is organized as follows. In section 2, we review the SR problems in the imaging field and introduce the network architecture of the proposed SSR-GAN model. The experimental designs and quantitative evaluations on simulated and real data are reported in sections 3 and 4, respectively. Finally, the results and related issues as well as future research direction are discussed in section 5.

2. Method

2.1. SR problem statement

In general, the conventional formulation of the ill-posed linear image SR problem (Yang et al 2010) can be formulated as

\[ x = D y \ast y + n \]  

(2)

where \( n \) denotes the noise and other factors. It is worth noting that both the down-sampling and blurring operations are nonlinear in practice. In addition, the nonlinearity exists in nonmodeled factors.

Our purpose is to convert LR sinograms obtained by the 2\( \times \)2 acquisition mode to HR sinograms. However, the sinogram SR task is a nonlinear and ill-posed problem, and there are many limitations in the sinogram domain. In recent years, DL-based methods have provided new ideas for the nonlinear and ill-posed problem. These methods use feed-forward CNNs to learn the mapping from input data to label data by optimizing parameters \( \theta \). The process can be written as

\[ \hat{y} = \Phi_{\theta}(x). \]

(3)

In the sinogram SR problem, to generate an excellent SR sinogram \( \hat{y} \), an appropriate network framework \( \Phi_{\theta} \) must be designed and the loss function should be specified to facilitate \( \Phi_{\theta} \) to obtain an SR sinogram based on the training datasets, so that

\[ \hat{\theta} = \arg \min_{\theta} \sum_{i} L(\hat{y}_{i}(\theta), y_{i}) \]

(4)

where \((x_{i}, y_{i})\) are the paired LR/HR sinograms for training.
2.2. Sinogram SR model

An overview of the proposed network model is shown in figure 1. The network model contains forward and backward cycles. In the forward cycle, the LR sinogram $x_i$ is input to the trained generator $G_H$ to obtain an SR sinogram $G_H(x_i)$. The trained generator $F_L$ translates the SR sinogram $G_H(x_i)$ back to the original LR sinogram. Similarly, in the backward cycle, the HR sinogram $y_i$ is input to the trained generator $F_L$ to obtain an LR sinogram $F_L(y_i)$. The trained generator $G_H$ translates the LR sinogram $F_L(y_i)$ back to the original HR sinogram. In addition, the discriminators $D_H$ and $D_L$ intend to identify that the sample is from real sinogram rather than generating sinogram. In the training process of the network, the generators $G_H$ and $F_L$ attempt to generate sinograms that are not easily distinguishable by the discriminators. Given an LR sinogram $x$, $G_H$ attempts to generate a new sinogram $G_H(x)$ highly similar to the corresponding HR sinogram $y$ so as to fool $D_H$. The key idea of the ‘game’ process is to perform joint/alternative training on the generator and discriminator to collaboratively improve the performance of the network. Therefore, we can describe the optimization problem as

$$\min_{G_H, F_L, D_H, D_L} \max_{L_GAN(G_H, D_H) + L_GAN(F_L, D_L).} \tag{5}$$

To promote high-precision SR sinogram generation, our proposed network model combines the following loss functions.

(a) **Relativistic adversarial GAN loss**: Adversarial losses (Goodfellow et al. 2014) are employed to prompt the generated sinograms to obey the empirical distributions in the source and target domains. Inspired by the relativistic GAN (Jolicoeur Martineau et al. 2019, Baby and Verhulst 2019), we replace the standard discriminator with a relativistic discriminator, which predicts the probability that a real HR sinogram $y$ is relatively more realistic than a fake sinogram. As shown in figure 2, the standard discriminator can be expressed as $D_H(y) = \sigma(C(y))$, where $\sigma$ denotes the sigmoid function and $C(y)$ denotes the non-transformed output of discriminator. The formulation of the relativistic discriminator is $D_H(y, G_H(x)) = \sigma(C(y) - \mathbb{E}[C(G_H(x))])$, where $\mathbb{E}[\cdot]$ denotes the expectation operator and $x$ is the LR sinogram.

Thus, the adversarial loss can be defined as

$$L^R_{GAN}(G_H, D_H) = \mathbb{E}_{x,y \sim \text{data}(x,y)} [\log D_H(y, G_H(x))] + \mathbb{E}_{x,y \sim \text{data}(x,y)} [\log(1 - D_H(G_H(x), y))]. \tag{6}$$

$$L^R_{GAN}(F_L, D_L) = \mathbb{E}_{x,y \sim \text{data}(x,y)} [\log D_L(x, F_L(y))] + \mathbb{E}_{x,y \sim \text{data}(x,y)} [\log(1 - D_L(F_L(y), x))]. \tag{7}$$

where $P_{data}$ denotes the sinogram data distribution.

(b) **Cycle consistency loss**: To ensure that the learned mapping transforms the source input successfully to the target output, we utilize the cycle consistency loss to constrain $x$ and $F_L(G_H(x))$ in the forward cycle as well as $y$ and $G_H(F_L(y))$ in the backward cycle, as expressed below:
\[ L_{\text{cycle}}(G_H, F_L) = \mathbb{E}_{x \sim P_{\text{data}}(x)} \left[ \| F_L(G_H(x)) - x \|_1 \right] \\
+ \mathbb{E}_{y \sim P_{\text{data}}(y)} \left[ \| G_H(F_L(y)) - y \|_1 \right] \tag{8} \]

where \(\| \cdot \|_1\) is the \(L_1\) norm. The cycle consistency loss urges \(F_L(G_H(x)) \approx x\) and \(G_H(F_L(y)) \approx y\) to prevent the degeneracy in adversarial learning (Li et al 2017).

(c) **Sinogram domain loss:** For the processing of sinograms, the accuracy of the generated sinogram is critical. The cycle consistency loss can only provide an indirect supervision, which is not sufficient to guarantee the accuracy of the final result. Therefore, the direct constraints to the generators are necessary, especially to \(G_H\), which directly produces the desired SR sinograms. In our sinogram task, the outcomes and labels should be matched in each pixel. Previous methods have illustrated that the MAE loss comes and labels should be matched in each pixel. Therefore, we constructed the sinogram domain loss function based on MAE loss, as follows:

\[ L_{\text{sin}}(G_H, F_L) = \mathbb{E}_{x,y \sim P_{\text{data}}(x,y)} \left[ \| F_L(y) - x \|_1 \right] \\
+ \mathbb{E}_{x,y \sim P_{\text{data}}(x,y)} \left[ \| G_H(x) - y \|_1 \right] \tag{9} \]

(d) **Reconstruction image domain loss:** For the sinogram SR task, the sinogram data are sensitive. Small errors in the generated sinogram may cause critical artifacts in the reconstructed image. In our study, to address this problem, the image domain loss is added in the proposed network model. The additional image domain loss can result in more realistic generated sinograms, achieved by the ASTRA-Toolbox (Van Aarle et al 2015). In addition, this study uses the MAE loss to compare the reconstructed images of the generated and real sinograms, as described below:

\[ L_{\text{image}}(G_H, F_L) = \mathbb{E}_{x,y \sim P_{\text{data}}(x,y)} \left[ \| A(F_L(y)) - A(x) \|_1 \right] \\
+ \mathbb{E}_{x,y \sim P_{\text{data}}(x,y)} \left[ \| A(G_H(x)) - A(y) \|_1 \right] \tag{10} \]

where \(A\) denotes the filtered-back-projection operation.

Therefore, the proposed model is trained to minimize the following total loss function:

\[ L_{\text{SSR-GAN}} = L_{\text{GAN}}^R(G_H, D_H) + L_{\text{GAN}}^R(F_L, D_L) \]

\[ + \lambda_1 L_{\text{cycle}}(G_H, F_L) + \lambda_2 L_{\text{sin}}(G_H, F_L) \]

\[ + \lambda_3 L_{\text{image}}(G_H, F_L), \tag{11} \]

where \(\lambda_1, \lambda_2,\) and \(\lambda_3\) are the weight parameters for balancing different penalties.

2.3. Network architecture

In the proposed method, the generator and discriminator of our network model play crucial roles. The specific architecture of the model is described as follows.

(a) **Generator:** As shown in figure 3(a), the overall architecture of the generator is similar to the traditional U-Net (Ronneberger et al 2015), except that, instead of the encoding–decoding structure, our generator is composed of three parts, namely the encoder, decoder, and transformation modules. In addition, the architecture of the encoder and decoder is more lightweight. A similar design of generator can also be seen in (Zhao et al 2018). As shown in figure 3(b), the transformation module is a residual network (ResNet) (He et al 2016), containing six ResNet blocks. The size of input sinograms in our generator is \(i \times 512 \times 512 \times 1\), where \(i\) is the batch size. In our generator, the encoder extracts features from the input sinogram to obtain the feature vector, with size \(128 \times 128 \times 256\). Then, the transformation module converts the feature vector extracted from source domain \(X\) to target domain \(Y\). The decoder restores the feature vector to low-level features by transposed convolution, and finally obtains the generated sinogram. Specifically, the encoder includes one Padding-7 \(\times 7\) Convolution (Conv)-Instance Norm (IN)-ReLU layer and two \(3 \times 3\) Conv-IN-ReLU layers. In the first layer, the parameter of padding is \(3\), and there are 64 filters with a stride of one. The next two layers consist of 128 and 256 filters with a
Figure 3. Architecture of the sinogram SR generators. In (a), the blue and green boxes indicate the feature vectors obtained by the network layers in the encoder and decoder. $X$ represents the input sinogram, and $Y$ represents the generated sinogram. In (b), there are six ResNet blocks, with the structure of each ResNet block being the same as the others.

Figure 4. Architecture of the discriminators. The yellow and light green boxes indicate the feature vectors obtained by the network layers in the discriminator. $Y$ represents the real sinogram, and $\hat{Y}$ represents the generated sinogram by generator.

In the transformation module, each ResNet block contains two convolution layers with 256 filters on both layers and the parameter of padding is 1. The decoder includes two $3 \times 3$ Deconvolution-IN-ReLU layers and one Padding-$7 \times 7$ Conv-Tanh layer. The two deconvolution layers consist of 128 and 64 filters with a stride of two. In the last layer, the parameter of padding is 3, and it consists of one filter with a stride of one.

(b) **Relativistic discriminator**: The structure of the relativistic discriminator is shown in figure 4, which is different from the standard discriminator $D$ in Cycle-GAN (Zhu et al. 2017). The relativistic discriminator predicts the probability that a real sinogram is relatively more realistic than a fake sinogram. The inputs of the relativistic discriminator are pairs of the real sinogram and the generated sinogram. The relativistic discriminator has five layers. The first layer is composed of a $4 \times 4$ Conv layer and leaky rectified linear units (Leaky ReLU) and consists of 64 filters with a stride of two. The second and third layers are $4 \times 4$...
Establishment of the training dataset

| Procedure for establishing the training dataset |
|-----------------------------------------------|
| **step 1.** Each CT images was subjected to value normalization. The image value was rescaled to $[0, 0.1]$. Then, the normalized images were taken as samples for generating sinogram. |
| **step 2.** Siddon’s ray-tracing algorithm (Siddon 1985) was applied to simulate the fan-beam geometry. The $1 \times 1$ acquisition mode was used to generate the HR sinograms for 360 views in $360^\circ$ with 512 linear detectors. The size of the generated HR sinograms is $360 \times 512$. Then, we added two $76 \times 512$ all-zero matrices to the generated sinograms to obtain $512 \times 512$ sinograms as the labels of the SSR-GAN. |
| **step 3.** To simulate the $2 \times 2$ acquisition mode, we retained the other parameters of the sinograms generated in step 2 and doubled the size of the detector. Then, we generated the LR sinograms for 360 views in $360^\circ$ with 256 linear detectors. The size of the generated LR sinograms is $360 \times 256$. Next, we used the bicubic interpolation algorithm to obtain the $360 \times 512$ sinograms from the generated $360 \times 256$ sinograms, and we added two $76 \times 512$ all-zero matrices to obtain $512 \times 512$ sinograms as the inputs of the SSR-GAN. In addition, we added noise to illustrate the practicality of the method. The noise is modelled as ‘Poisson + Gaussian’ model (La Riviere et al 2006):

$$g_i \sim \text{Poisson}\left\{I_0 \exp\left(-\bar{g}_i\right)\right\} + \text{Normal}\left\{0, \sigma^2\right\}$$ (12)

where $i$ denotes the pixels in the sinogram data, $g_i$ denotes the measured sinogram with added noise, $\bar{g}_i$ denotes the ideal measured sinogram. The initial intensity $I_0$ of incident photon was set to $1 \times 10^5$ and the variance $\sigma^2$ was set to 25. |

On the basis of the above procedure, 1879 pairs of input and label sinograms with a size of $512 \times 512$ were prepared.

**2.4. Image reconstruction from generated sinogram**

We obtain the generated HR sinogram by inputting the LR sinogram into the trained SSR-GAN model. Then, we obtain the reconstructed CT image from the generated HR sinogram using the classic reconstruction algorithms, such as the FBP algorithm (Katsevich 2001). By this way, a relatively high-quality reconstructed image is obtained from an LR sinogram.

**3. Evaluation**

**3.1. Experimental data and training details**

**3.1.1. Digital CT image data**

In our simulated experiment, the experimental datasets were established from a real clinical dataset that contains 2159 pleural CT images from six patients. The size of image is $512 \times 512$. The training data for training the SSR-GAN were prepared with 1879 CT images from five patients, while the test data were prepared using 280 CT images from the other one patient. The detailed procedure for preparing the training dataset is shown in table 1. For the preparation of the test dataset, the 280 CT images were used to generate LR sinograms by the process explained in step 3 in table 1.

**3.1.2. Anthropomorphic head phantom data**

To verify the potential SR capability of the proposed model for an actual CT system, we investigated a radiological anthropomorphic head phantom study for clinical applications. The phantom is shown in figure 5, and its specifications are described in the ICRU Report 48 (1992). In the real data experiment, the CT projection data were acquired by a CT scanner, which is mainly composed of the x-ray source (Hawkeye130, Thales) and a flat-panel detector (Thales 4343F). In the data acquisition, we obtained the projection datasets under different parts of the head phantom for training and testing of the SSR-GAN, respectively. The training and test datasets were completely separated; the two closest slices in the two datasets were also separated by 200 slices. The specific scanning parameters are listed in table 2.
### Table 2. Parameters set in the head phantom experiment.

| Parameters                             | Value of parameter |
|----------------------------------------|--------------------|
| Detector elements                      | 512                |
| Detector bin size (mm)                 | 0.592              |
| Distance of source to object (mm)      | 529.16             |
| Distance of source to detector (mm)    | 998.75             |
| Tube voltage (kVp)                     | 120                |
| Tube current (µA)                      | 200                |
| Scanning range (°)                     | 360                |
| Number of projections                  | 360                |
| Reconstruction size                    | 512 × 512          |

**Figure 5.** Real data experimental phantom: Chengdu dosimetric phantom.

**Figure 6.** Average RMSE of SR sinograms for different values of $\lambda_1$ when $\lambda_2 = 0$ and $\lambda_3 = 0$.

A significant difference between the simulated data and the actual collected data, we must retrain the SSR-GAN so that it can be applied in the actual CT system.

In the experiment, we collected 1002 HR sinograms of phantom slices for training the SSR-GAN. Furthermore, we collected additional 201 HR sinograms of phantom slices for preparing the test dataset. The size of the HR sinograms is 360 × 512. Limited by experimental conditions, we cannot directly obtain the LR sinograms based on the 2 × 2 acquisition mode. The 360 × 256 LR sinograms were obtained by down-sampling and blurring the real collected HR sinograms. A Gaussian kernel of size 3 × 3 with standard deviation 1 was used for blurring. Note that the value range of all sinograms was normalized to [0, 1] in the data preprocessing stage. For the training datasets, we utilized the bicubic interpolation algorithm to obtain 360 × 512 sinograms from the LR sinograms, and then added two 76 × 512 all-zero matrices to get 512 × 512 sinograms as the inputs of the SSR-GAN. Moreover, we added two 76 × 512 all-zero matrices to the real collected HR sinograms to obtain 512 × 512 sinograms as the labels of the SSR-GAN. We obtained 1002 pairs of 512 × 512 input sinograms and corresponding label sinograms. For testing the SSR-GAN, we obtained the LR sinograms by down-sampling and blurring the 201 real collected HR sinograms. Then, we
Figure 7. Average RMSE of SR sinograms for different values of $\lambda_2$ and $\lambda_3$ when $\lambda_1 = 10$. The darker the color of the square, the smaller the estimated RMSE of the sinograms.

utilized the bicubic interpolation algorithm to obtain $360 \times 512$ sinograms from the LR sinograms and added two $76 \times 512$ all-zero matrices to generate $512 \times 512$ testing sinograms.

3.1.3. Training details
All training and testing procedures were performed using the Pytorch toolbox (ver. 0.4.1) running on an AMAX workstation with two GeForce RTX 2080 Ti GPUs (NVIDIA Corporation). The parameters were updated by the adaptive moment (Adam) optimizer, the learning rate was fixed at 0.0002 in the first 100 epochs and decreased linearly from 0.0002 to 0 over the next 100 epochs, and the exponential decay rates were $\beta_1 = 0.5$ and $\beta_2 = 0.999$. In our experiment, the batch size was 2 in all the training processes. In the training process of the SSR-GAN, approximately 22.205 million parameters were to be learned. The amount of data affects the training time of the SSR-GAN. In the simulated experiment, the time cost was approximately 61 h for the overall training procedure of the SSR-GAN, whereas it was approximately 36 h in the real data experiment. The weight parameter selection of loss function $L_{SSRGAN}$ is important. A detailed discussion is presented in section 4.1.

3.2. Performance comparison
In this study, the bicubic interpolation-based (Liu et al 2013), U-Net-based (Park et al 2018), and the cycle-GAN-based (Yuan et al 2018) SR methods were adopted for comparison. We trained the U-Net- and cycle-GAN-based SR methods with different hyperparameter settings to obtain relatively good SR performance. The network structure of the U-Net-based SR method is the original U-Net framework. In addition, the heart of the cycle-GAN-based SR method is the original cycle-GAN framework (Zhu et al 2017), and we trained the cycle-GAN in the supervised manner. The reconstruction images were obtained from the SR sinograms of different methods by using the FBP algorithm (Katsevich 2001).

For the quantitative evaluation of the SR performance of the proposed model, we validated the reconstruction results in terms of three image quality metrics: PSNR, structural similarity (SSIM) (Wang et al 2004), and root mean square error (RMSE). Through extensive experiments, we compared the different SR methods on the simulated and real datasets described in section 3.1.

4. Results

4.1. Parameter selection in loss function
The parameter selection experiments are based on simulated data with gaussian noise. In loss function $L_{SSR \cdot GAN}$, the weight parameters $\lambda_1$, $\lambda_2$, and $\lambda_3$ together determine the optimal proportion of different losses in the whole training process of the SSR-GAN. In this study, $\lambda_1$ controls the weight of the cycle-consistency loss. To obtain a relatively good value of parameter $\lambda_1$, we first set $\lambda_2 = 0$ and $\lambda_3 = 0$. It means that the proposed SSR-GAN model was only optimized with the cycle consistency loss and the adversarial loss. In this scheme, we investigated the performance of the SR model on the test dataset. For fairness, each SR model
had the same parameter setting, except \( \lambda_1 \). The 20 LR sinograms from the test dataset were randomly selected for testing the SR performance of the different models. The effect of \( \lambda_1 \) was quantitatively determined by calculating the average RMSE of the SR sonograms, as shown in figure 6. From the curves of figure 6, when \( \lambda_1 = 10 \), the average RMSE of the SR sinograms reached the minimum. Therefore, we adjusted the other two parameters \( \lambda_2 \) and \( \lambda_3 \) based on \( \lambda_1 = 10 \) to test the effects of the sinogram domain loss and reconstruction image domain loss.

Weight parameters \( \lambda_2 \) and \( \lambda_3 \) are important for controlling the balance between the sinogram loss and reconstruction loss. When \( \lambda_3 \) is extremely small, the proposed model basically cannot benefit from the image domain information. Similarly, when \( \lambda_3 \) is extremely small, the constraint of the model on the output sinograms is reduced, which is not conducive to the generation of accurate HR sinograms. In our study, we also investigated the performance of the retrained SR model with different values of parameters \( \lambda_2 \) and \( \lambda_3 \) on the test dataset. The effect of \( \lambda_2 \) and \( \lambda_3 \) were quantitatively determined by calculating the average RMSE of the SR sinograms, as shown in figure 7. The color of squares indicate that when \( \lambda_2 = 1 \times 10^{-4} \) and \( \lambda_3 = 500 \), the RMSE of the SR sinograms reached the minimum. Based the optimized hyper parameter, the relative strength of each terms in equation (11) is that \( \lambda_2 L_{\text{sino}} > \lambda_3 L_{\text{image}} > \lambda_1 L_{\text{cyc}} > L_{\text{GAN}} \).

Based on the above experimental results, we set \( \lambda_1 = 10 \) and \( \lambda_2 = 1 \times 10^{-4} \) in the following simulated and real data experiments. In addition, we set \( \lambda_3 = 300 \) in simulated data experiments and \( \lambda_3 = 100 \) in real data experiments.

### 4.2. Simulated data experiments

To analyze the sinogram SR capability of the proposed SSR-GAN, four representative sinograms are shown in figure 8, and the reconstructed CT images of the first and fourth sinograms are shown in figure 9.
Figure 8 indicates that the proposed SSR-GAN captures more sinogram information. Compared with the HR sinograms obtained under the $1 \times 1$ acquisition mode, the sinograms generated by the SSR-GAN only have few errors. Compared with other SR methods, the results obtained by the proposed SSR-GAN had greater accuracy. Figure 9 shows the reconstructed images from the SR sinograms obtained by different SR methods and the error maps between the ground truth and the reconstructed images. Moreover, we selected the two regions of these two slices for further analysis of the reconstructed details, which are shown in the third and sixth rows in Figure 9. We observe that the reconstructed CT images from the LR sinograms were of relatively poor quality, and most of the image details were flooded by artifacts. Compared with other SR methods, the proposed SSR-GAN model showed excellent capabilities in image detail restoration. Based on the proposed model, we can obtain a reconstruction result that is similar to the ground truth. The error maps indicate that the reconstruction results of the proposed model have the smallest difference. Moreover, in the regions of interest (ROIs), we can clearly see that the results based on LR sinograms contained many artifacts. The bicubic interpolation-, U-Net-, and cycle-GAN-based SR methods could depress the artifacts to some extent and recover the images with relatively high quality. However, the image details indicated by arrows are still blurred. Compared with the other methods, the propose method recovered relatively clear image details in the ROIs. This indicates that the proposed method has advantages in detail restoration and blur reduction.
Figure 10. Reconstruction results of two representative test slices. (a) and (b) are slices 1 and 2 reconstructed from the HR sinograms under the 1×1 acquisition mode. The red rectangular box is the ROI of slice 1 and the yellow rectangular box is the ROI of slice 2. The display window is [0, 0.01].

Figure 11. SR results for two test sinograms of the head phantom by different methods. (a) Collected HR sinogram data under the 1×1 acquisition mode as the ground truth; (b) estimated SR sinograms by bicubic interpolation; (c) estimated SR sinograms by the classical U-Net; (d) estimated SR sinograms by the cycle-GAN; (e) estimated SR sinograms by the proposed SSR-GAN; (f) error maps of (a); (g) error maps of (a); (h) error maps of (a); (i) error maps of (a). The display window of (a)–(e) is [0, 1]. The display window of (f)–(i) is [−0.06, 0.06].

Table 4. Quantitative evaluations of reconstruction results based on LR and different SR sinograms in the anthropomorphic head study (20 testing images).

| Method      | Avg. PSNR  | Avg. RMSE  | Avg. SSIM     |
|-------------|------------|------------|---------------|
| LR          | 25.9039    | 1.1857 × 10⁻⁴ | 0.9963        |
| Bicubic     | 29.1175    | 9.6131 × 10⁻⁵ | 0.9978        |
| U-Net       | 29.5526    | 6.9118 × 10⁻⁵ | 0.9986        |
| Cycle-GAN   | 30.5004    | 4.0560 × 10⁻⁵ | 0.9993        |
| SSR-GAN     | 31.7643    | 3.3982 × 10⁻⁵ | 0.9995        |

Table 4 shows the quantitative comparison results of the reconstruction images, which can further reflect the performance of the proposed method. The evaluation items indicate the benefit of different SR methods in the image reconstruction of LR sinograms. Among them, the SSR-GAN performed the best in terms of SSIM, which indicates that the results based on SSR-GAN are the most structurally similar to the ground truth. In terms of PSNR and RMSE, the proposed method also exhibited satisfactory performance, indicating that the noise removal is relatively effective. Overall, the quantitative evaluation results of RMSE, PSNR, and SSIM indicate that the proposed method can recover high-quality CT images from LR sinograms.

4.3. Real data experiments

The performance of the proposed method was further evaluated on a radiological anthropomorphic head phantom. We selected two test slices for exhibition and the reconstruction images of these slices are shown at an enlarged field of view with 360 × 360 square pixels. Figures 10(a) and (b) show the CT images of two representative test slices reconstructed by the FBP algorithm, which were considered as the ground truth of the reconstructed images. In addition, we selected two ROIs to evaluate the ability of different methods to restore image details.

Figure 11 shows the sinograms estimated by the different SR methods and their error maps. Compared with the real HR sinograms, the results of the proposed method only have few errors. The error maps also indicate that the proposed method obtains the most similar results to the collected HR sinograms.
Figure 12 shows the reconstruction results of the anthropomorphic head phantom. The first row represents the reconstructed images from the collected HR sinograms, LR sinograms, and different SR sinograms. The second row represents the error maps between the ground truth and the resulting images. The third row represents the zoom-in ROIs of reconstruction results. The fourth to sixth rows represent the similar results to the first to third rows. The results of the first to third rows are related to slice 1, and those of the fourth to sixth rows are related to slice 2. From left to right, the columns represent the results based on the collected HR sinograms, LR sinograms, SR sinograms by bicubic interpolation, SR sinograms by U-Net, SR sinograms by cycle-GAN, and SR sinograms by the proposed SSR-GAN. The display windows of all figures in the second and fifth rows are $[-0.003, 0.003]$. The display windows of all figures in the first, third, fourth, and sixth rows are $[0, 0.01]$. The quantitative evaluation results of reconstructed images are shown in table 4. Results show that the proposed method exhibits good performance in terms of noise suppression and structure preservation; these results are consistent with the findings given in table 3. Thus, the evaluation of the head phantom also validates the efficiency of the proposed method in CT imaging under the $2\times2$ acquisition mode.
4.4. Ablation study

In order to study the effects of relativistic discriminator, sinogram domain loss, and reconstruction image domain loss in proposed SSR-GAN model, we gradually modify the baseline model and compare their differences. The 20 LR sinograms from the test dataset were randomly selected for testing the performance of different models. We validated the reconstruction results in terms of PSNR and SSIM to quantitatively evaluate the effects of each major modification in the proposed SSR-GAN. The overall quantitative comparison is shown in Table 5. Each column represents a model, and its configuration is shown in the second to fourth rows. From the 1st and 2nd configurations in Table 5, a slight improvement can be observed by using the relativistic discriminator. From the 2nd to 4th configuration, we can further observe that the weighted sinogram domain loss and reconstruction image domain loss help to promote noise suppression and structure preservation. Moreover, the weighted sinogram domain loss contributes more to noise suppression, while reconstruction image domain loss contributes more to structure preservation in our study.

5. Discussion and conclusion

In CT imaging, the $2 \times 2$ acquisition mode can improve the collection efficiency of the projection and significantly reduce the x-ray exposure time. However, this acquisition mode is rarely adopted because of the low-quality reconstruction results. The study of sinogram SR problem is beneficial to improving the reconstruction quality of CT images under the $2 \times 2$ acquisition mode. Currently, most of the existing SR methods only focus on the image domain, and the information of the projection domain has not received great attention, thus limiting the application of this acquisition mode in practice. A feasible approach to increase the effective information of LR sinograms is to use sinogram SR methods for image reconstruction. Inspired by the recent successful image-to-image translation applications of the Cycle-GAN, we proposed the SSR-GAN for supervised generation of SR sinograms. To ensure that the generated SR sinograms have high accuracy, we designed the relativistic discriminator to predict the probability that a real sinogram is relatively more realistic than a fake sinogram. Moreover, we added the weighted sinogram domain loss and reconstruction image domain loss for the proposed SSR-GAN, which can increase the fidelity of the reconstructed image. The experiment results indicate that the proposed method outperforms other SR methods.

For the training data, we performed zero-padding to avoid rounding operation when the network performs convolution. In practice, the collected sinogram may not be suitable for direct input to the network. For example, the sinograms can only be collected in a limited angular range due to the limitation of the mechanical system. Thus, the size of the collected sinogram can be $150 \times 512$, if this sinogram is direct input to the network, the problem of input and output sinogram size mismatch will present. However, the problem can be effectively solved by zero-padding operation. Therefore, the zero-padding operation makes the proposed network have strong adaptability for the collected sinogram. Although the padding function inside the network can also solve the problem of the sinogram size mismatch, such an operation will slightly increase the difficulty of network training. Compared to handling this problem inside the network, using zero-padding on the input side is a better choice. In addition, the sinogram data are sensitive, with a small change in them having a significant effect on the reconstruction image quality; therefore, we used paired data to train the proposed model in supervised manner. Although one of the major advantages of Cycle-GAN is that it can realize unsupervised training, the unsupervised generation of SR sinogram is difficult to ensure high accuracy. In order to further illustrate the necessity of supervised training, we removed the sinogram domain loss and reconstruction image domain loss in the proposed SSR-GAN, and then trained and tested the model in supervised and unsupervised modes respectively. The 20 LR sinograms from the test dataset were randomly selected for testing the performance of different training mode. Table 6 shows the quantitative comparison results of SR sinograms. The evaluation items indicate that the generated SR sinograms have the higher accuracy in supervised training mode.
Table 6. Quantitative evaluations of SR sinograms in different training mode.

| Training mode    | Avg. RMSE | Avg. PSNR |
|------------------|-----------|-----------|
| Unsupervised     | $4.4863 \times 10^{-3}$ | 43.1741   |
| Supervised       | $2.5996 \times 10^{-3}$ | 44.2013   |

For the proposed SSR-GAN, appropriate hyperparameters are the guarantee of network performance. In particular, the different weights in the total loss function significantly affect network training. By adjusting the weights, the different losses are relatively balanced, and therefore, the network can realize the continuous interaction of the projection domain and image domain information during training. Thus, the generated HR projection has high accuracy. The selection of different weights requires considerable experimental arguments, which can be seen in section 4.1. Furthermore, we can observe that the sinogram domain loss and reconstruction image domain loss are relatively more important based on the parameter selection experiments. The experiment results indicated that additional training was needed for the sinogram data collected by an actual CT system, probably due to the significant difference in the distribution of simulated data and real data. Obviously, the difference in data distribution also exists in different CT imaging systems. However, we can collect the training data from the different CT imaging systems, and then train multiple models for practical application. In addition, for a specific CT system, we can collect the projection data of different scanned objects to expand the training dataset. The robustness of the trained network model is enhanced with the increase in the collected projection data.

The proposed generator is composed of three parts, namely the encoder, decoder, and transformation modules. The encoder extracts features from the input sinogram; the transformation module converts the features extracted from source domain to target domain, and the decoder uses the features of target domain to generate the SR sinogram. For the proposed generator, we refer to the architecture from Johnson et al. (2016) who have shown impressive results for neural style transfer and natural image SR. In addition, the parameter setting is empirical in the generator, which also refers to the setting in the U-Net (Ronneberger et al. 2015) and the ResNet (He et al. 2016). In our study, the sinogram SR problem can be framed as an image transformation task. We use the transformation module to replace the convolutional layer between the encoder and decoder in the original U-Net. The transformation module uses the ResNet to achieve accurate mapping between different domain features. The experimental results illustrate that the proposed network has good performance to generate the HR sinogram.

Although the proposed model has achieved compelling SR results, there still exist some limitations. First, compared with the general DL-based methods, a considerably longer training time was required by the proposed SSR-GAN. In this aspect, we should consider designing more efficient architecture that would render the network capable of learning more complex features of sinograms with less computational cost and lower model complexity in the future. Second, although the proposed method can recover image details with high quality, some subtle features may not be always faithfully recovered. To address this issue, future work will explore an improved reconstruction algorithm that uses the SR sinogram as the initial iteration value and utilize the scanned LR sinogram to construct the fidelity term. The improved reconstruction algorithm may further enhance the image details.

In conclusion, this study proposed an efficient method to obtain HR sinograms from LR sinograms under the $2 \times 2$ acquisition mode by taking advantage of the neural network in nonlinear mapping and utilizing the information interaction between the projection domain and the image domain. Based on the HR sinograms obtained by the proposed SSR-GAN, relatively high-quality reconstruction images can be obtained by the FBP algorithm. The proposed method effectively promotes the practical application of the $2 \times 2$ acquisition mode and has considerable potential for reducing the x-ray radiation dose of CT imaging, increasing the CT scanning speed and reducing the CT production cost.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (Grant No. 61601518) and the China Postdoctoral Science Foundation (Grant No. 2019M663996).

ORCID iDs

Chao Tang  🌐 https://orcid.org/0000-0002-3589-0256
Wenkun Zhang  🌐 https://orcid.org/0000-0003-4720-2593
References

Baby D and Verhelst S 2019 Segan: speech enhancement using relativistic generative adversarial networks with gradient penalty Int. Conf. on Acoustics Speech and Signal Processing pp 106–10

Bouman C A and Sauer K 1996 A unified approach to statistical tomography using coordinate descent optimization IEEE Trans. Image Process. 5 480–92

Chao D, Chen C L and Tang X 2016 Accelerating the super-resolution convolutional neural network European Conf. on Computer Vision (https://doi.org/10.1007/978-3-319-46475-6_25)

Dong C, Loy C C, He K and Tang X 2014 Image super-resolution using deep convolutional networks IEEE Trans. Pattern Anal. Mach. Intell. 38 295–307

Farsiu S, Robinson M D, Elad M and Milburn P 2004 Fast and robust multiframe super resolution IEEE Trans. Image Process. 13 1327–44

Ganguli S, Garzon P and Glaser N 2019 GeoGAN: a conditional GAN with reconstruction and style loss to generate standard layer of maps from satellite images arXiv: Computer Vision and Pattern Recognition (arXiv:1902.05611)

Goodfellow I et al 2014 Generative adversarial nets neural information processing systems pp 2672–80

Greenspan H 2009 Super-resolution in medical imaging Comput. J. 52 43–63

Hassan A, Nazir S and Alkadhdi H 2010 Technical challenges of coronary CT angiography: today and tomorrow Eur. J. Radiol. 79 161–71

Hatvani J, Horvath A, Michelti J, Basarab A and Gyongy M 2018 Deep learning-based super-resolution applied to dental computed tomography IEEE Trans. Radiat. Plasma Med. Sci. PP 1–1

He K, Zhang X, Ren S and Sun J 2016 Deep residual learning for image recognition Computer Vision and Pattern Recognition pp 770–8

ICRU 1992 Phantoms and computational models in therapy, diagnosis and protection Report No. 48, Bethesda, MD, USA

Izumi S, Kamata S, Satoh K and Miyai H 1993 High energy x-ray computed tomography for industrial applications IEEE Trans. Nucl. Sci. 40 158–61

Jaffrey D, Siewerdsen J, Wong J and Martinez A 2002 Flat-panel cone-beam computed tomography for image-guided radiation therapy Int. J. Radiat. Oncol. Biol. Phys. 53 1337–49

Jiang C, Zhang Q, Rui F and Hu Z 2018 Super-resolution CT image reconstruction based on dictionary learning and sparse representation Sci. Rep. 8 8799

Johnson J, Agha A and Feifei L 2016 Perceptual losses for real-time style transfer and super-resolution European Conf. on Computer Vision pp 694–711

Jolicoeurmartineau A 2019 The relativistic discriminator: a key element missing from standard GAN Int. Conf. on Learning Representations

Katsevich A 2001 Theoretically exact FBP-type inversion algorithm for spiral CT SIAM J. Appl. Math. 62 2012–26

Kim J, Lee J K and Lee K M 2016 Accurate image super-resolution using very deep convolutional networks Computer Vision and Pattern Recognition pp 1646–54

La Riviere P J, Bian J and Vargas P 2006 Penalized-likelihood sinogram restoration for computed tomography IEEE Trans. Med. Imaging 25 1022–36

Lai W, Huang J, Ahuja N and Yang M 2017 Deep laplacian pyramid networks for fast and accurate super-resolution Computer Vision and Pattern Recognition pp 5835–43

Lecun Y, Bengio Y and Hinton G 2015 Deep learning Nature 521 436–44

Li C et al 2017 ALICE: towards understanding adversarial learning for joint distribution matching Neuronal Information Processing Systems pp 5495–503

Liu J, Gan Z and Zhu X 2013 Directional bicubic interpolation: a new method of image super-resolution 3rd Int. Conf. on Multimedia Technology (ICMT-13) (Atlantis Press) (https://doi.org/10.2991/icmt-13.2013.37)

Mccollough C, Leng S, Yu L and Fletcher J 2015 Dual- and multi-energy CT: principles, technical approaches, and clinical applications Radiology 276 637–53

Mouton A and Breckon T 2015 Materials-based 3D segmentation of unknown objects from dual-energy computed tomography imagery in baggage security screening Pattern Recognit. 48 1961–78

Park J, Hwang D, Kim K Y, Kang S K, Kim Y K and Lee J S 2018 Computed tomography super-resolution using deep convolutional neural network Phys. Med. Biol. 63 145011

Qiao T, Zhang W, Zhang M, Ma Z and Xu D 2019 Ancient painting to natural image: a new solution for painting processing Workshop on Applications of Computer Vision pp 521–30

Ronneberger O, Fischer P and Brox T 2015 U-Net: convolutional networks for biomedical image segmentation Int. Conf. on Medical Image Computing and Computer-Assisted Intervention (https://doi.org/10.1007/978-3-319-24574-4_28)

Sauer K and Bouman C 1993 A local update strategy for iterative reconstruction from projections IEEE Trans. Signal Process. 41 534–48

Siddon R L 1985 Fast calculation of the exact radiological path for a three-dimensional CT array Med. Phys. 12 252–5

Thibault J B, Sauer K D, Bouman C A and Hsieh J 2007 A three-dimensional statistical approach to improved image quality for multislice helical CT Med. Phys. 34 p. 4526

Umehara K, Ota J and Ishida T 2018 Application of super-resolution convolutional neural network for enhancing image resolution in chest CT J. Digit. Imaging 31 441–50

Van Aarle W, Palenstijn W J, De Beenhouwer J, Altantzis T, Bals S, Batenburg K J and Sijbers J 2015 The ASTRA toolbox: a platform for advanced algorithm development in electron tomography Ultramicroscopy 157 35–47

Wang G 2016 A perspective on deep imaging IEEE Access 4 8914–24

Wang Z, Bovik A C, Sheikh H R and Simoncelli E P 2004 Image quality assessment: from error visibility to structural similarity IEEE Trans. Image Process. 13 600–12

Yang C Y and Yang M H 2013 Fast direct super-resolution by simple functions IEEE Int. Conf. on Computer Vision (https://doi.org/10.1109/ICCV.2013.75)

Yang J, Wright J, Huang T S and Ma Y 2010 Image super-resolution via sparse representation IEEE Trans. Image Process. 19 2961–73

You C et al 2020 CT super-resolution GAN constrained by the identical, residual, and cycle learning ensemble (GAN-CIRCLE) IEEE Trans. Med. Imaging 39 168–90

Yu H et al 2017 Computed tomography super-resolution using convolutional neural networks Int. Conf. on Image Processing pp 3944–8

Yu Z, Wu G, Yap P T, Feng Q and Shen D 2012 Reconstruction of super-resolution lung 4D-CT using a patch-based sparse representation 2012 IEEE Conf. on Computer Vision and Pattern Recognition (https://doi.org/10.1109/CVPR.2012.627767)

Yuan Y, Liu S, Zhang J, Zhang Y, Dong C and Lin L 2018 Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks Computer Vision and Pattern Recognition pp 701–10
Zhang R, Thibault J-B, Bouman C, Sauer K and Hsieh J 2013 Model-based iterative reconstruction for dual-energy x-ray CT using a joint quadratic likelihood model IEEE Trans. Med. Imaging 33 117–34
Zhao B, Chang B, Jie Z and Sigal I 2018 Modular generative adversarial networks European Conf. on Computer Vision pp 157–73
Zhou Y, Thibault J B, Bouman C A, Sauer K D and Jiang H 2011 Fast model-based x-ray CT reconstruction using spatially nonhomogeneous ICD optimization IEEE Trans. Image Process. 20 161–75
Zhu J, Park T, Isola P and Efros A A 2017 Unpaired image-to-image translation using cycle-consistent adversarial networks Int. Conf. on Computer Vision pp 2242–51