

Deep learning based Metal Artifact Reduction in X-ray Computed Tomography

Chunjie Su*

College of Electronics and Information, Southwest Minzu University, Sichuan, 610200, China

* Corresponding author: Chunjie Su (Email: 1728711323@qq.com)

Abstract: Due to the presence of metal implants, computed tomography(CT) images of patients undergoing CT scans produce severe metal artifacts. In recent years, metal artifact reduction (MAR) algorithms have been developed at a high speed, and deep learning-based MAR algorithms have proved to be one of the very effective methods. However, based on the fact that most of the current deep learning-based solutions utilize simulated data for supervised training, these models are difficult to be directly applied in clinical settings. In addition, current MAR schemes still face considerable challenges in reduce streak artifacts and dark band artifacts from metal artifacts. In this paper, we propose an improved unsupervised network model based on CycleGAN suitable for small-sample training of medical images. In order to enhance the extraction of image features, the generator in this paper adopts the U-Net network model, which is commonly used in medical image processing, and the residual connection and attention modules are added between each layer of the U-Net network. In addition, in order to better repair the dark band artifacts, the input of the network is processed by mixing the original image and the prior image, and in order to make the output image retain as much as possible the effective information of the original image, this paper adds the correlation constraints of the prior image and the original image to the output of the network. Experimentally, it is proved that the method in this paper has very good ability to reduce metal artifacts.

Keywords: Image domain, Unsupervised networks, Small samples.

1. Introduction

Computed Tomography (CT) is one of the most common aids to modern medical diagnosis. However, due to the inherent physical properties of the machine and other objective reasons, the actual scanned CT images are often accompanied by artifacts, such as metal artifacts. Metal artifacts are caused by severe beam hardening, scattering and ray starvation when X-rays are scanned into high-density metals, resulting in severe black dark bands and streak artifacts in the reconstructed CT images. With the rapid development of modern medical technology, the proportion of patients with metal implants (dentures, cardiac stents, bone marrow fixation nails, hip prostheses, etc.) is gradually increasing, so it is very important to reduce CT metal artifacts to improve the quality of CT images.

The common studies of CT Metal Artifact Reduction (MAR) are divided into three main categories. The first category is the iterative correction algorithm based on iterative reconstruction [1-3]. Since the iterative reconstruction-based iterative correction algorithm relies on extremely high computing power, it is difficult to be implemented into practical applications at the current stage.

The second type of traditional interpolation algorithm is based on filter backprojection(FBP). In 1987, Kalender [4] et al. proposed a linear interpolation based metal artifact reduction algorithm (LI-MAR). The LI-MAR algorithm can reduce most of the structural stripes artifacts, but due to the unsatisfactory interpolation effect, it introduces more serious secondary artifacts. In 2009, E. Meyer [5] et al. proposed the Normalized Metal Artifact Reduction (NMAR) algorithm based on a priori image normalization, which mitigates the problem of unsatisfactory interpolation to a certain extent by adding the process of normalized smoothing, thus effectively mitigating the introduction of secondary artifacts.

The third category is deep learning-based metal artifact reduction algorithms. In 2017, Park [6] et al. proposed a U-Net-based metal artifact reduction algorithm. In 2017, Zhang and Yu [7] proposed a metal artifact reduction algorithm based on convolutional neural network (CNN-MAR). In 2019, Lin [8] et al. proposed the DuDoNet dual-domain MAR network model, et al [9-11]. All of the above methods are supervised networks, and all of the above networks use simulated paired data for network training, which leads to the inability to be directly applied to actual clinics.

This paper is an improved unsupervised network model based on CycleGAN [12] network modeling framework, which can directly use unpaired real CT image data to train the network for better clinical applications. The generator of this model adopts the U-Net network with residual connectivity and attention mechanism, which is better adapted to small-sample training, while the discriminator follows the discriminator of CycleGAN, which has been proved to be excellent in the article [12]. The model uses real CT image data as the dataset, and the test results show that the model can effectively reduce metal artifacts.

2. Proposed Method

This model uses CycleGAN as the base model architecture, which divides metal artifact-affected images and artifact-free images into two different styles of samples, allowing the model to perform the task of style migration from artifact-affected images to artifact-free images. The original CycleGAN network model was designed based on traditional images. The information content and information density of traditional images are much higher than that of medical CT images and the number of samples is sufficient, while medical CT images are simpler in structure and semantics, and the number of samples is small. Therefore the CT metal artifact

reduction task is prone to overfitting by directly using the generator of the CycleGAN network model, so this model is designed to be more suitable for the small sample features of medical images.

2.1. Overview of the overall network model architecture

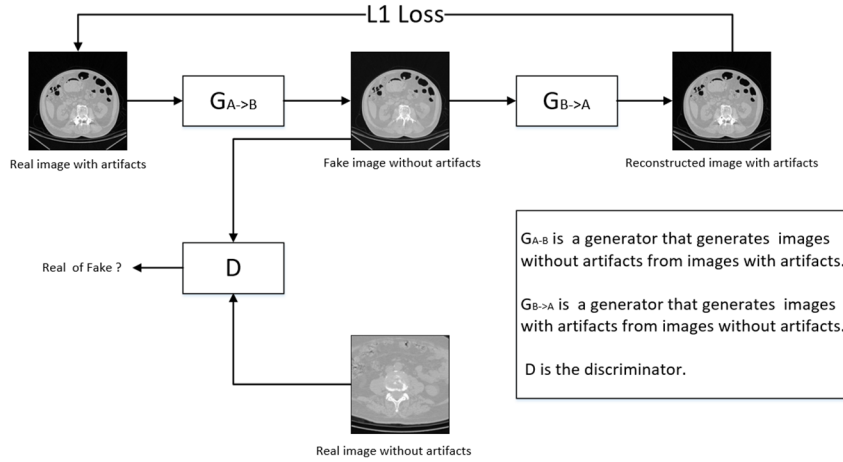


Figure 1. Overall Network Model Architecture

CycleGAN is an unpaired image style transformation model designed to transform images in one image domain to images in another image domain. Unlike traditional supervised image style migration methods, CycleGAN does not require paired images and can learn the corresponding mappings from unpaired images. The architecture of CycleGAN consists of two generators and two discriminators. Where generator $G_{A \rightarrow B}$ converts an image from domain A to domain B, generator $G_{B \rightarrow A}$ converts an image from domain B to domain A, and the discriminators are used to determine whether the generated image is true or not. CycleGAN uses adversarial training and cyclic consistency loss to learn the mapping between images while maintaining the semantic information in the original image.

2.1.1. Generator

U-Net[13] network architecture is one of the most commonly used network architectures for medical image processing, which adopts an encoder-decoder architecture with a symmetric U-shape structure. The basic principle of U-Net is to realize the fine extraction and reconstruction of features through the down-sampling and up-sampling operations, and to fuse the information of the up and down-

sampling processes through the jump connection. In the down-sampling process, U-Net uses a convolutional kernel with stride 2 to realize the compression of the feature map, which not only reduces the size of the feature map and extracts more abstract feature information, but also effectively reduces the amount of computation and memory consumption. During the up-sampling process, U-Net uses the transposed convolution operation to expand the feature map, and this process can recover the original resolution and detail information of the image. The coder-decoder connection channel and the jump connection fuse the information between the encoder and the decoder. The jump connection mechanism allows the decoder to access more low-level detail information. This low-level detail information is particularly important for the metal artifact reduction task because they often contain detailed information about edges and textures, which play a crucial role in maintaining the organization and sharpness of the CT image. In addition, in the case of small samples, the U-Net network jump-connection mechanism can fully utilize the information at all levels, which can largely solve the phenomenon of network overfitting caused by insufficient information in small samples. The specific generator structure is shown below:

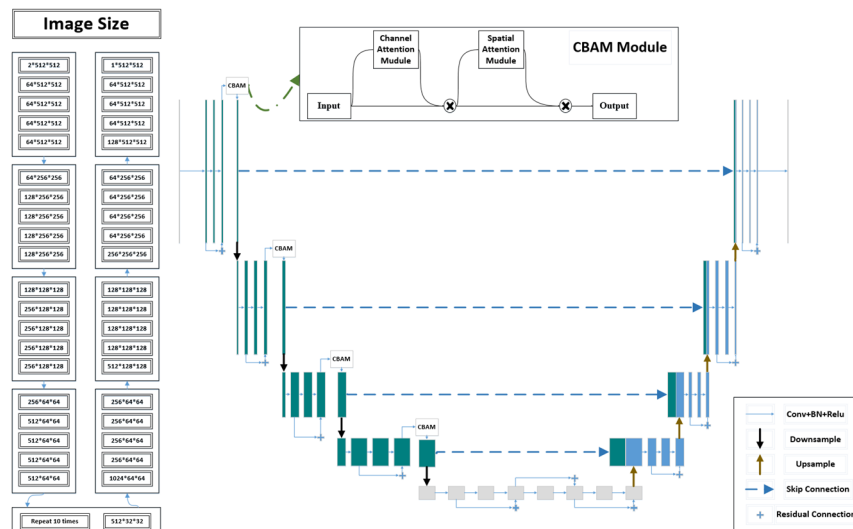


Figure 2. Generator Architecture

Both the encoder and decoder of the generator contain four layers (excluding the bottom layer), where the bottom layer is the connection channel of the codec. Each layer of the coder and decoder structure contains three convolutional modules, where the first convolutional module serves to expand the number of image channels and the next two convolutional

modules form the residual structure. Attention modules are added before each downsampling in the encoder structure. The connected channels are composed of one convolutional module and three residual blocks, the residual block structure is as follows:

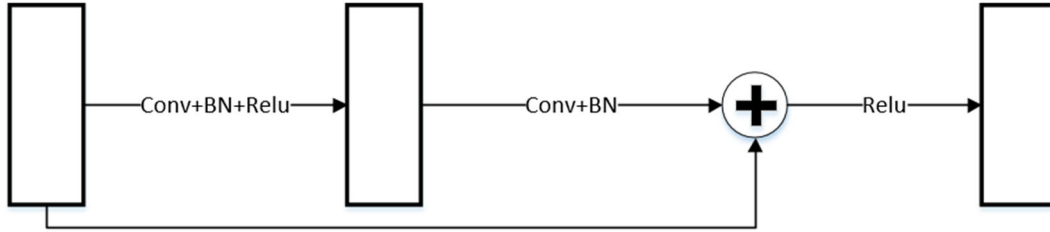


Figure3. Schematic of the residual structure

The residual structure is added to each layer of the coding and decoding structure, which further improves the utilization of low-level feature information in the encoder. The three residual blocks in the coder-decoder connection channel appropriately increase the depth of the network model, allowing the network to extract more abstract information. The inclusion of the residual connection mechanism can speed up the feature information transfer and can improve the stability and generalization ability of the model.

Attention mechanism is one of the most common ways to improve feature information extraction in deep learning models. Convolutional Block Attention Module (CBAM) [14] is a module that combines Spatial Attention and Channel Attention. This module has low computational complexity and good feature representation. The CBAM module first performs global maximum pooling and global average pooling operations on the input feature maps to aggregate the channel orientation information. The pooling result is fed into the shared fully connected layer, and the two outputs of the

shared fully connected layer perform a pixel-by-pixel summation operation to obtain the channel attention weights. Then, the CBAM module inputs the feature map processed by the channel attention mechanism into the spatial attention module. The spatial attention module first performs channel-based maximum pooling and average pooling operations on the feature map. Then, the pooled results are spliced along the channel dimension and the number of channels is compressed to 1 by a 7x7 convolutional layer. Finally, the spatial attention weights are obtained using the Sigmoid activation function. The spatial attention weights are multiplied with the feature map to obtain the enhanced feature representation. CBAM allows dynamic channel and spatial importance assessment of features. Weighting based on the importance enables a better assessment of the importance of the features, thus enabling U-Net to extract feature information in a more refined way. The final result is to improve the model's ability to reduce metal artifacts. Below is a schematic of the CBAM structure and its sub-modules:

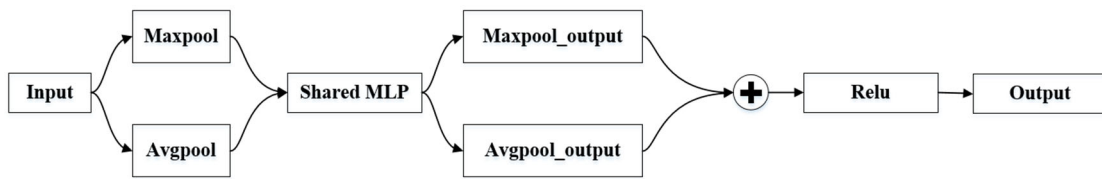


Figure 4. Channel Attention Mudule

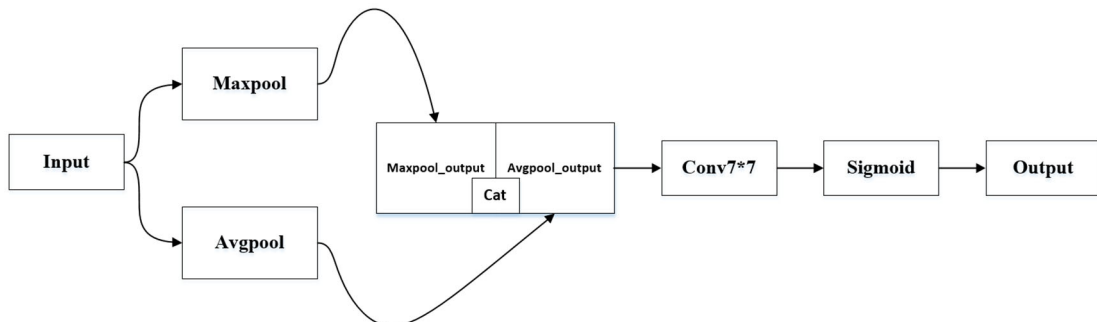


Figure 5. Spatial Attention Mudule

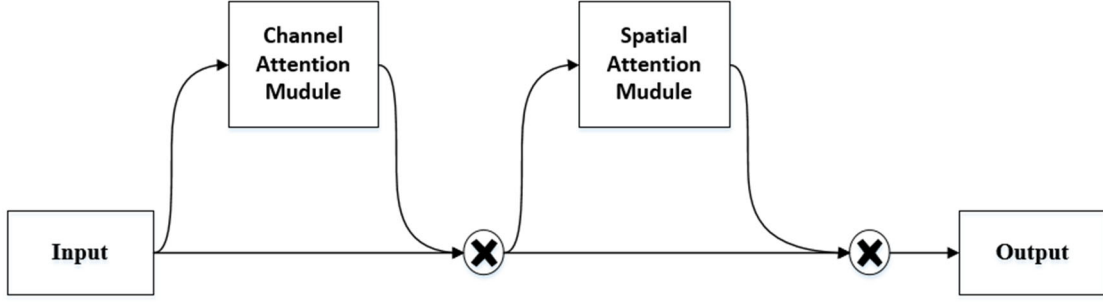


Figure 6. CBAM module

2.1.2. Discriminator

The discriminator structure uses the PatchGAN [14] structural model, which is different from traditional image discriminators, which usually take the whole image as input and ultimately give a score used to judge the quality of the whole generated image. PatchGAN, on the other hand, uses a fully convolutional network, and the final output is a matrix.

Each matrix element corresponds to a region on the original image, and the value of that matrix element represents the score of a region of the original image. The advantage of this is that PatchGAN can better capture the local information of the image and can analyze the structure of the image in more detail. Below is a schematic of the network structure of the discriminator:

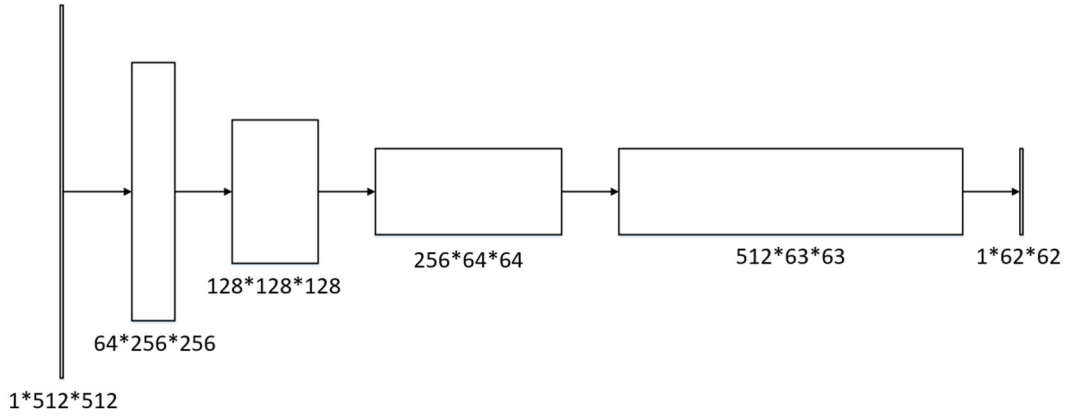


Figure 7. Discriminator structure schematic

2.2. Loss Function

The loss function of the whole model includes adversarial loss of the generator, cyclic consistency loss and L1 loss function. The adversarial loss is used to identify whether the image produced by the generator is similar in style to the real image in the target domain. This helps the generator to learn the distribution of images in the target domain. The cyclic consistency loss enforces that the image produced by the generator can be restored to the original image after a cyclic transformation between the two generators. This helps to preserve the content information in the original image and avoids distortion during the conversion process. To ensure that the image details in the non-metallic portion of the CT image are preserved as much as possible, an additional L1 loss is added between the generated image and the original image.

The adversarial loss function formula is:

$$L_{ADV} = E_{y \sim P_{data}(y)} [\log D_Y(y)] + E_{x \sim P_{data}(x)} [\log (1 - D_Y(G(x)))] \quad (1)$$

Where E represents the mathematical expectation, $P_{data}(y)$ represents the data y obeys the $P_{data}(y)$ distribution. D represents the discriminator. G represents the generator. x and y represent the images that are input to the generator and discriminator, respectively.

The circular consistency loss function formula is:

$$L_{cyc} = E_{x \sim P_{data}(x)} [\|F(G(x)) - x\|_1] + E_{y \sim P_{data}(y)} [\|G(F(y)) - y\|_1] \quad (2)$$

where G represents a generator that generates an artifact-free image from an artifact-containing image. F represents a generator that generates an artifact-affected image from an artifact-free image. x and y represent images that are input to generators G and F.

The L1 loss equation is:

$$L_1 = \frac{1}{N} \sum_{i=1}^N (\|f(x_i) \times x_{mask} - x_i \times x_{mask}\|_1 + \|f(x_i) \times (1 - x_{mask}) - NMAR \times (1 - x_{mask})\|_1) \quad (3)$$

Where f represents the generator. N represents the total number of image elements. x_i represents the input image. NMAR represents the a prior image processed by the NMAR algorithm. x_{mask} represents the non-dark band region binary mask.

3. Experimental Results

This experiment is compared with the results of BHC (beam-hardening correction), LI-MAR, NMAR and raw CycleGAN processing, respectively. The display window is [40,400] HU.

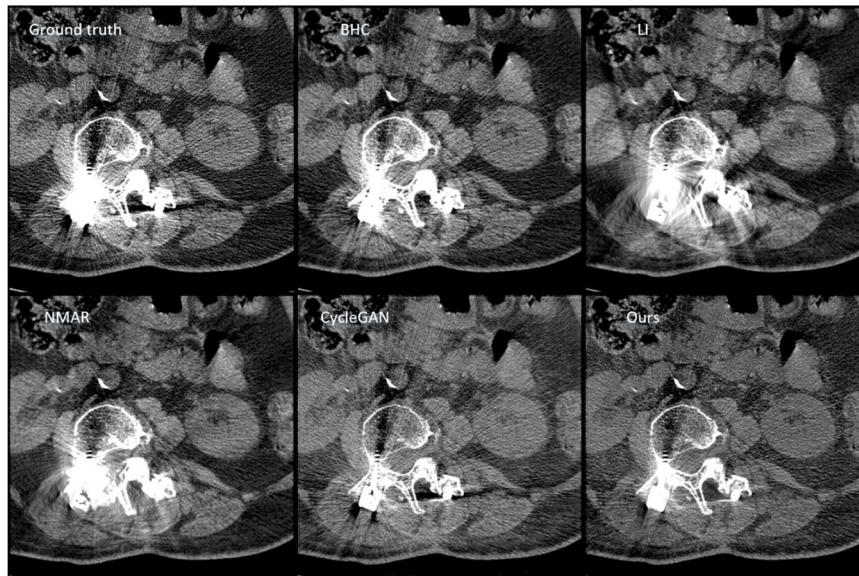


Figure 8. Results in small field of view 1

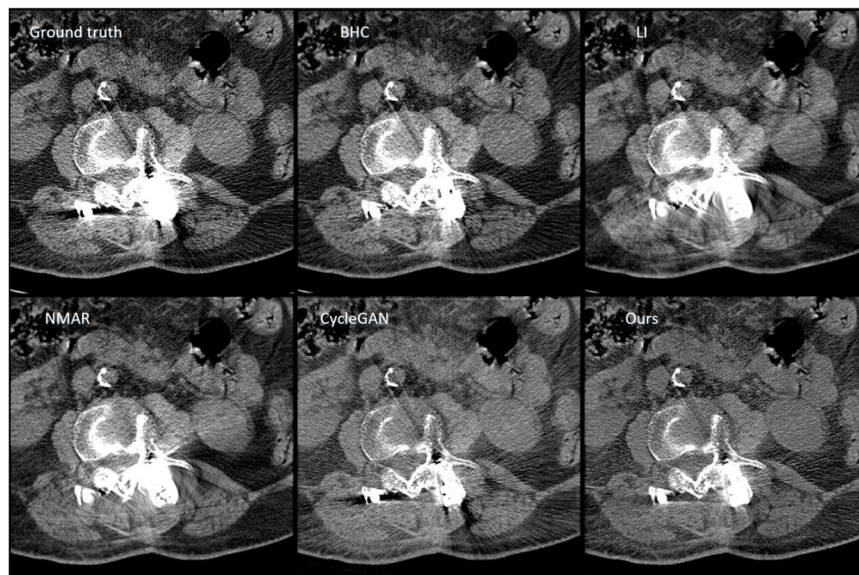


Figure 9. Results in small field of view 2

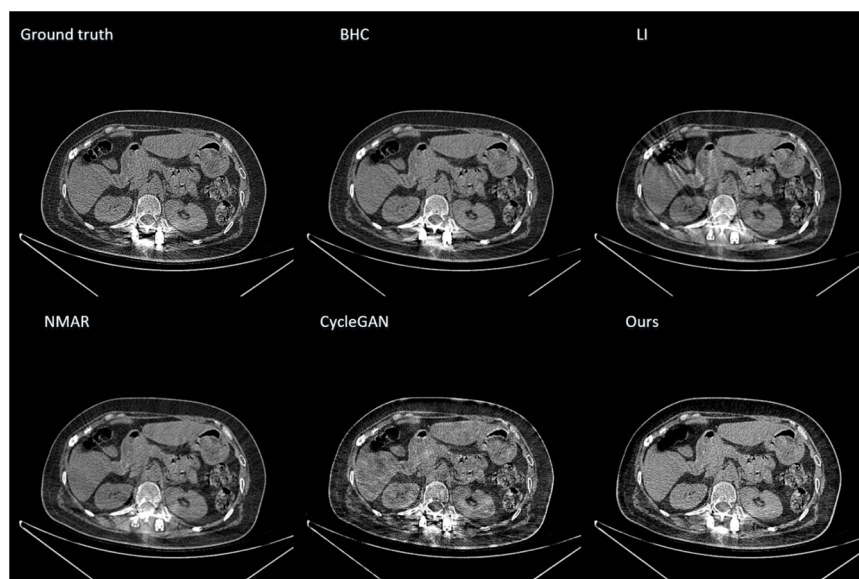


Figure 10. Results in large field of view

As can be seen in Figs.8 and 9, the BHC algorithm retains strong streak artifacts and dark bands in the small FOV(Field

of View) case, and the LI-MAR algorithm introduces strong streak secondary artifacts throughout the entire image.The

NMAR algorithm is better, but still introduces strong streak secondary artifacts near the metal, and the bone morphology around the metal is greatly changed. The cyclegan network streak artifacts removal effect is not ideal, and the processing effect is very poor for the dark band artifacts. The algorithm proposed in this paper, on the other hand, can remove the streak artifacts and repair the dark band artifacts well without introducing secondary artifacts.

As can be seen in Fig.3, the BHC algorithm introduces secondary artifacts in the dark band restoration region at large FOV. The LI-MAR algorithm processed image still introduces strong streak secondary artifacts throughout the whole image. the NMAR algorithm processed the image better, but still introduces strong streak secondary artifacts near the metal. The CycleGAN network processed image organization structure has a large deformation. The algorithm proposed in this paper maintains the organization of the original image well and has a good reduction of dark bands and streak artifacts.

4. Implementation Details

4.1. Datasets

This experimental dataset uses real CT clinical data and contains 1600 sheets in total. Among them, 1300 sheets are used for training and 300 sheets are used for validation. Among the metallic implant categories are mainly spinal fixation nails and hip prostheses.

4.2. Implementation and Training details

This experiment is based on network training on Linux system using NVIDIA GeForce RTX3090 24GB GPU, and the overall framework is built based on pytorch. In the training process, the optimizer used is the Adam optimizer with parameters $(\beta_1, \beta_2) = (0.5, 0.999)$, batch size is set to 1, the input image size is 512×512 , and the overall training epoch is set to 200. In order to make the training smoother, the learning rate strategy of this experiment adopts the warm-up mode, and the base learning rate is set to 0.0001. In order to enhance the stability of the model and the effect of artifact reduction, the cyclic consistency loss coefficient is set to 20, and the L1 loss function coefficient is set to 15. In addition, in order to make the network's artifact reduction effect better, the input of the network simultaneously takes the NMAR-processed images together as the input of the overall network, so that the dark band artifacts can be better reduced.

Regarding the calculation of L1 loss, in order to ensure that the output image is as close as possible to the information of the original artifact-affected image, the calculation of L1 loss is divided into two parts, for metal and dark band regions L1 loss is carried out between the output image and the NMAR image, and for non-metal and non-dark band regions L1 loss is carried out between the output image and the original artifact-affected image.

5. Conclusion

This paper is based on CycleGAN improved unsupervised network model, for CycleGAN's generator is dealing with natural images, this paper builds a generator based on U-Net network suitable for small sample features of medical images. In order to increase the utilization of small sample image

information, the residual structure and attention mechanism are added to the generator, which strengthens the ability of extracting image information, enhances the ability of reducing metal artifacts, and can better retain the original metal artifact-containing useful image information. Comparative experimental results show that this network can reduce metal artifacts well.

References

- [1] Gordon R , Bender R , Herman G T .Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography.[J].Journal of Theoretical Biology, 1970, 29(3):471,IN1,477-476,IN2,481.DOI:10.1016/0022-5193(70)90109-8.
- [2] Williamson J F , Whiting B R , Benac J ,et al.Prospects for quantitative CT imaging in the presence of foreign metal bodies using statistical image reconstruction[C]//Proceedings IEEE International Symposium on Biomedical Imaging.0[2023-07-24].DOI:10.1109/ISBI.2002.1029341.
- [3] Kole J S , Beekman F J .Evaluation of the ordered subset convex algorithm for cone-beam CT[J].Physics in Medicine & Biology, 2005, 50(4):613.DOI:10.1088/0031-9155/50/4/004.
- [4] W. A. Kalender, R. Hebel, and J. Ebersberger, "Reduction of CT artifacts caused by metallic implants," *Radiology*, vol. 164, no. 2, pp. 576–577, Aug. 1987.
- [5] Meyer E , Raupach R , Lell M ,et al.Normalized metal artifact reduction (NMAR) in computed tomography[J].Medical Physics, 2010, 37(10).DOI:10.1118/1.3484090.
- [6] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention*, pp. 234–41, Springer, 2015.
- [7] L. Gjestebj, Q. Yang, Y. Xi, H. Shan, B. Claus, Y. Jin, B. De Man, and G. Wang, "Deep learning methods for CT image-domain metal artifact reduction," in *Developments in X-Ray Tomography XI*, vol. 10391, p. 103910W, International Society for Optics and Photonics, 2017.
- [8] Lin W A , Liao H , Peng C ,et al.DuDoNet: Dual Domain Network for CT Metal Artifact Reduction[J].arXiv, 2019.DOI:10.48550/arXiv.1907.00273.
- [9] Lyu Y , Lin W A , Lu J ,et al.DuDoNet++: Encoding mask projection to reduce CT metal artifacts[J]. 2020.DOI:10.48550/arXiv.2001.00340.
- [10] Yu L , Zhang Z , Li X ,et al.Deep Sinogram Completion with Image Prior for Metal Artifact Reduction in CT Images[J].IEEE Transactions on Medical Imaging, 2020.DOI:10.1109/TMI.2020.3025064.
- [11] Wang T , Xia W , Lu Z ,et al.IDOL-Net: An Interactive Dual-Domain Parallel Network for CT Metal Artifact Reduction[J]. 2021.DOI:10.48550/arXiv.2104.01405.
- [12] Zhu J Y , Park T , Isola P ,et al.Unpaired Image-to-Image Translation using Cycle-Consistent Adversarial Networks[J].IEEE, 2017.DOI:10.1109/ICCV.2017.244.
- [13] Ronneberger O , Fischer P , Brox T .U-Net: Convolutional Networks for Biomedical Image Segmentation[C]//Springer, Cham.Springer, Cham, 2015.DOI:10.1007/978-3-662-54345-0_3.
- [14] Isola P , Zhu J Y , Zhou T ,et al.Image-to-Image Translation with Conditional Adversarial Networks[J].IEEE, 2016.DOI:10.1109/CVPR.2017.632.