



**HAL**  
open science

## **Cas-PRED: A limited-angle cone beam CT reconstruction research based on prior information**

Yuan Gao, Liu Zhang, Wenxue Yu, Shijie Wang, Pinzheng Zhang, Cheng Xue, Chunfeng Yang, Shipeng Xie, Gouenou Coatrieux, Jean-Louis Coatrieux, et al.

### ► **To cite this version:**

Yuan Gao, Liu Zhang, Wenxue Yu, Shijie Wang, Pinzheng Zhang, et al.. Cas-PRED: A limited-angle cone beam CT reconstruction research based on prior information. IEEE Transactions on Instrumentation and Measurement, 2024, 73, pp.1-16. <10.1109/TIM.2023.3335513>. <hal-04467402>

**HAL Id: hal-04467402**

**<https://univ-rennes.hal.science/hal-04467402v1>**

Submitted on 25 Apr 2024

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

# Cas-PRED: A limited-angle cone beam CT reconstruction research based on prior information

Yuan Gao, Liu Zhang, Wenxue Yu, Shijie Wang, Pinzheng Zhang, Cheng Xue, Chunfeng Yang, Shipeng Xie, Gouenou Coatrieux, *Senior Member, IEEE*, Jean-Louis Coatrieux, *Fellow, IEEE*, Yang Chen, *Senior Member, IEEE*

**Abstract**—Due to limitations in radiation dose or scanning conditions, there are instances where CT data can only be collected within a restricted scanning angle. However, the use of limited-angle projection data often results in reconstructed images that exhibit significant noise and artifacts, which can significantly impact diagnostic accuracy. Conventional iterative reconstruction algorithms, such as the total variation algorithm, consume considerable time simulating the cone beam forward and backward projection processes. Furthermore, these algorithms often struggle to address the challenges of limited-angle reconstruction due to their inherent limitations. To tackle these challenges and achieve superior reconstruction outcomes, we introduce a novel approach known as the Cascaded Prior-Based Residual Encoder-Decoder Network (Cas-PRED). This method employs a multi-level cascaded network as its core architecture, leveraging full-angle CT reconstruction results as prior information to guide the deep learning network’s training process. Our approach consistently delivers superior results compared to existing reconstruction methods, with impressive performance metrics, including a peak signal-to-noise ratio (PSNR) of  $33.69 \pm 0.90$  dB, a structural similarity index (SSIM) of  $0.9557 \pm 0.0029$ , and a root mean squared error (RMSE) of  $22.73 \pm 1.56$  for the result of  $[0, 150^\circ]$  reconstruction, highlighting its effectiveness in enhancing image quality and accuracy.

**Index Terms**—Cone beam CT reconstruction, Limited-angle, Deep learning, Prior information.

Manuscript created April, 2023; This work was supported in part by the State Key Project of Research and Development Plan under Grants 2022YFC2408500 and 2022YFC2401600, in part by the National Key Research and Development Program of China under Grant 2022YFE0116700, in part by the National Natural Science Foundation of China under Grant T2225025, in part by the Key Research and Development Programs in Jiangsu Province of China under Grants BE2021703 and BE2022768, in part by Jiangsu Province Science Foundation for Youths under Grant BK20220825. First author: Yuan Gao, Co-first author: Liu Zhang, Corresponding author: Chunfeng Yang, Co-corresponding author: Shipeng Xie.

Yuan Gao, Liu Zhang, Wenxue Yu, Shijie Wang, Pinzheng Zhang, Cheng Xue, Chunfeng Yang, and Yang Chen are with the Lab of Image Science and Technology, Key Laboratory of New Generation Artificial Intelligence Technology and Its Interdisciplinary Applications (Southeast University), Ministry of Education China, Jiangsu Provincial Joint International Research Laboratory of Medical Information Processing, School of Computer Science and Engineering, Southeast University, Nanjing 211189, China (e-mail: yuangao@seu.edu.cn, liuzhang@seu.edu.cn, ywx.list@seu.edu.cn, mike@seu.edu.cn, luckzpz@seu.edu.cn, cxue@seu.edu.cn, chunfeng.yang@seu.edu.cn, chenyang.list@seu.edu.cn). Shipeng Xie is with the College of Telecommunications and Information Engineering, Nanjing University of Posts and Telecommunications, Nanjing 210003, China (e-mail: xie@njupt.edu.cn). Gouenou Coatrieux is with the Information and Image Processing Department, Institut Mines-Telecom, Telecom Bretagne, Brest 29238, France (e-mail: gouenou.coatrieux@telecom-bretagne.eu). Jean-Louis Coatrieux is with the Centre de Recherche en Information Biomédicale Sino-Francais, Inserm, University of Rennes 1, Rennes 35042, France (e-mail: jean-louis.coatrieux@univ-rennes1.fr).

The opinions expressed here are entirely that of the author. No warranty is expressed or implied. User assumes all risk.

## I. INTRODUCTION

COMPUTED tomography (CT) imaging technology based on X-ray absorption and attenuation characteristics is widely used in medical diagnosis [1], [2], [3], safety inspection [4], [5], and industrial detection [6], [7], [8]. In the ideal scenario, CT scanning can provide high-resolution medical images, which can be used to assist lesion localization, segmentation [9], [10], and disease diagnosis. However, the imaging quality will be limited by specific imaging conditions in practical applications. For example, X-ray radiation harms human cells, so it is necessary to reduce radiation dose [11]. Standard methods include reducing the tube current/voltage [12], using sparse scanning ways, and using limited-angle scanning ways [13], [14]. Among these three methods, limited-angle scanning is the most commonly used method in practical applications. For example, in dynamic organ imaging, the heart and lungs contract and relax over time. Since too long scanning can cause motion blur, doctors use limited-angle scanning to reduce scanning time effectively. For some imaging scenarios with limited scanning angles, such as dental examination, breast tomography, and active pipeline scanning, doctors can only collect data at a limited-angle in actual scenarios. Although limited-angle imaging can reduce scanning time and radiation dose, artifact noise will appear in the reconstructed image and affect the imaging quality. The artifact noise may cover adequate information, such as lesions and anatomical structures. In recent years, researchers have proposed three algorithms for the accurate reconstruction of CT images: analytic reconstruction algorithm, iterative reconstruction algorithm, and deep learning algorithm. These algorithms can address the artifact noise problem described above to varying degrees.

FDK [15] is a typical analytical algorithm for cone beam CT reconstruction of complete data sets. It has been widely used because of its strong stability, simple calculation, fast speed, and easy implementation. However, this algorithm relies heavily on the completeness of projection data. For incomplete data sets, the reconstruction results will contain many noise artifacts. Therefore, it is generally used as the initial processing step in the scene of limited-angle CT reconstruction and then combined with an iterative algorithm or deep learning algorithm [16], [17], [18].

Iterative reconstruction algorithms can use existing projection data to bring better results than FDK [19]. In particular, it has been used to try to solve artifacts in limited reconstruction

results [20]. In order to improve reconstruction results of incomplete projection data sets, researchers consider introducing additional prior information or regular term [21], [22] into the iterative algorithm. The commonly used regularization methods include the iterative reconstruction method based on total variation (TV) or dictionary learning. The sparsity of the image gradient-domain can be controlled through specific adjustable parameters [23]. The implementation of the TV algorithm is relatively simple. The minimization process of TV iterative reconstruction can take data fidelity and gradient sparsity into account, which helps obtain a better processing effect compared with simultaneous iterative reconstruction (SART) algorithm [24], [25], [26]. In order to avoid block and gradient artifacts, Wei Yu *et al.* [27] proposed to use the L0 norm to replace the L1 norm in the TV algorithm, which improves the processing effect near the edge. Meng Cao *et al.* [28] used OMP algorithm [29] to learn two dictionaries corresponding to effective structures and artifacts, respectively. The two dictionaries were used to represent low-quality images sparsely. Xu Moran *et al.* [30] combined dictionary learning algorithm with L0 norm constraint in image gradient-domain to further protect image edges and eliminate shadow artifacts. However, this type of iterative algorithm cannot completely solve the problem of artifacts in the reconstruction results and the problem of missing effective information under imaging conditions where projection data is seriously missing.

With the improvement of computer computing power and the development of deep learning algorithm, image processing algorithm based on the convolutional neural network has achieved good performance in many medical fields, such as medical lesion segmentation and detection, medical disease diagnosis, and multimodal medical image registration [31], [32], [33], [34], [35]. Many researchers have applied deep learning algorithms to solve limited-angle CT reconstruction and achieve remarkable results. Hanming Zhang *et al.* [36] tried to train a simple full convolution neural network to process limited-angle CT images directly and proved the effectiveness of a full convolution network. However, restricted by the number of network parameters and learning methods, it can only suppress noise artifacts of limited-angle CT images to a certain extent. It was inspired by the network structure [37] proposed by Kaiming He *et al.*, YoSeob Han *et al.* [38] tried to solve the problem using residual learning in 2015. The network obtained competitive results, which showed the promoting effect of residual learning in limited CT reconstruction. Unlike natural images, CT reconstruction involves projection and image domains. Therefore, to fully use the useful information of different domains, Qiyang Zhang *et al.* [39] constructed an end-to-end network combining the projection domain and image domain. The propagation of the gradient can cross-domain and realize mutual promotion. However, this method must realize two large networks simultaneously, and it introduces propagation error in filtered back projection. For 3D cone beam CT reconstruction, the calculation pressure of the whole process cannot be ignored. The processing of limited-angle CT image needs to recover the real anatomical structure of the scanned object from the noise artifact. To make the reconstruction results closer to the real distribution,

some researchers [40] use the generative adversarial network (GAN). Compared with simple L1 and L2 loss functions, the discriminator network can provide better convergence results. The patch-GAN network was constructed by Zicheng Li *et al.* [41] used an encoder-decoder network as a generator. It attempts to recover the missing projection data in the projection domain. Shipeng Xie *et al.* [42] used condition GAN network to solve limited-angle reconstruction in the image domain instead of the projection domain. This method combined with various loss functions helps obtain more complete results. However, several GAN network methods mentioned above require training two networks at the same time. The training effects between the two networks will seriously affect each other. In addition, the training data is also required to be highly consistent. In the limited-angle dual-energy CT reconstruction field, Yikun Zhang *et al.* [43] proposed the PIE-ARNet to exploit prior information to perform dual-energy computed tomography (DECT) imaging with two complementary quarter scans, reducing radiation dose, and shortening the scan time. However, this method only addresses the limited-angle reconstruction problem of spiral CT, and the limited-angle reconstruction problem of cone-beam CT requires further research.

One of the reasons why deep learning algorithms can achieve excellent performance is that neural networks can represent a considerable solution space. However, training a complex neural network for medical image reconstruction tasks requires elaborate design. Compared with complex type acquisition scenes of natural images, medical images mainly collect the scanning data of various human body parts. The same parts of different people and the symmetrical part of the same person have similar anatomical structures. Based on this, we propose a new method named Cas-PRED. To solve this problem, we introduced Cas-PRED, which uses prior information to guide the limited-angle reconstruction task of cone beam CT to better control the training of neural networks, improve the training effect, reduce blur problems, and speed up the training process. This solves the problem that the reconstructed image may be blurry or unclear due to the low radiation dose during cone beam CT reconstruction at limited angles. This is critical in the field of medical imaging, where doctors need accurate images for diagnosis and treatment planning while reducing radiation damage to patients. Cas-PRED is an important tool with the potential to improve the limited-angle reconstruction quality of cone beam CT, helping to find a better balance between radiation dose and image quality to meet the needs of the medical field. The experimental results show that satisfactory results can be obtained for limited-angle reconstruction. The main contributions of this paper include the following aspects:

1. Cas-PRED uses a two-stage encoder decoder cascade network, which can process limited-angle CT images progressively. The cascaded encoder-decoder structure realizes layer-by-layer processing, avoids the repeated forward projection and back projection in the traditional iterative algorithm, and dramatically improves the accuracy of reconstruction results.
2. Cas-PRED extracts prior information to complement the critical feature extraction capabilities of the reconstruction

network. The features extracted by the pre-trained encoder will be used as the prior features of the secondary network to help get better reconstruction results.

3. A multi-scale loss function is introduced for Cas-PRED to improve the neural network training effect. A pre-trained network is used to extract multi-scale features. The multi-scale features are compared with the features of the cascaded encoder-decoder network to calculate the multi-scale error.

An overview of this article is as follows. Section II describes the principles of the cascade network structure, prior information selection and extraction, and multi-scale loss functions used in this method. Section III introduces the data and illustrates the analysis of the experimental results. Section IV summarizes the contributions and future developments of this study. Details are shown below.

## II. THEORETICAL DESCRIPTIONS

The cascaded residual encoder-decoder network (Cas-PRED) is mainly used to solve the problem of cone beam CT limited-angle reconstruction. The network consists of several components: an FDK reconstruction module, a cascaded encoder decoder module, and a prior information selection and extraction module. The overall network structure is shown in Fig. 1.

### A. FDK reconstruction module

For the cone beam scanning reconstruction problem, the FDK algorithm is a common analytical reconstruction algorithm for preliminary reconstruction. The FDK algorithm transforms the cone beam image reconstruction into a fan beam image reconstruction problem [15]. The FDK algorithm can obtain accurate reconstruction results of projection data simply and quickly. However, the results contain severe artifacts and miss information in the case of missing projection data. In this paper, the FDK algorithm is used as a processing module for training deep convolutional networks. The FDK algorithm reconstructs a series of [full-angle image, limited-angle image] pairs. We use the full-angle image as the training target and the limited-angle image as the training input. Case-PRED divides the FDK algorithm into three steps. 1) projection data preprocessing, 2) preprocessed data filtering, and 3) weighted inverse projection. The distance from the reconstructed point determines the weights to the focal point. The equation for this process is defined as follows.

$$f(r, \phi, z) = \frac{1}{2} ABg(l, \hat{z}, \beta)h(l' - l)dld\beta \quad (1)$$

$$S = r \sin(\phi - \beta) \quad (2)$$

$$l' = \frac{Dr \cos(\phi - \beta)}{D - r \sin(\phi - \beta)} \quad (3)$$

$$A = \int_0^{2\pi} \frac{D}{D - S} \quad (4)$$

$$B = \int_{-\infty}^{\infty} \frac{D}{\sqrt{D^2 + l + \hat{z}^2}} \quad (5)$$

where  $l$  is the coordinate of projection data on the detector,  $\hat{z}$  is the coordinate of projection data on the rotating axis,  $\beta$  is the rotation angle,  $D$  is the focal length,  $g(l, \hat{z}, \beta)$  is cone beam projection data,  $h(l' - l)$  is Hann filter,  $f(r, \phi, z)$  is the reconstruction value obtained at the corresponding reconstruction point.

### B. Cascaded residual encoder decoder module

The reconstruction of cone beam CT limited-angle often requires to be achieved from a complex neural network. The disadvantage of a complex network is that the network training is difficult. In order to solve this problem, multiple cascaded networks can be used to reconstruct the full-angle image gradually. The cascaded networks at different stages undertake different functions in the image reconstruction process to obtain better processing results. Modular design has been successfully applied in many deep learning tasks because of its simple structure, standardized definition, and good processing effect [44], [45], [46], [47]. In this paper, a two-stage cascaded residual encoder-decoder network is used. Each stage of the encoder-decoder structure has different sizes and similar structures. The encoder performs continuous down-sampling through max-pooling operation, which can reduce the feature scale, expand the receptive field, and extract features of different granularity. The decoder comprises a series of deconvolution operations, which can recover high-dimensional features from the encoder's output results. In the process of down-sampling and up-sampling, the practical information can be restored without disturbances and artifacts [47]. At the same time, to solve the structural distortion caused by the encoder during continuous down-sampling and up-sampling [48], a series of residual blocks are introduced to enable the encoder and decoder to fuse features of different scales. Each residual block contains a three-layer convolution combination operation including convolution, batch normalization, and Leaky relu. The structure of the encoder and decoder is shown in Fig. 2.

### C. Prior information selection and extraction module

Different from the complex and diverse scenes of natural images, the main scene of medical image research is the human body. Although the sections of similar positions of different people are not the same in detail, they still have many similar structural features. For example, the CT sections of the lung include the lung, nodules, blood vessels, spine, muscle and other organs. Since there are currently no conditions to obtain tomographic images of the same person at different times, we study the extraction of prior information from full-angle images of others. Suppose  $P$  is a prior image library, which contains many complete angle prior images. There is a very high probability that an image has a great similarity with the image currently processed by the algorithm. By comparing the MSE value and SSIM value, the closest image can be selected from  $P$ . If an image  $P_i$  is selected as the prior image, we register  $P_i$  with the processed image. Through the affine transformation including translation, rotation and scaling, the two images have the largest feature similarity. Therefore, the

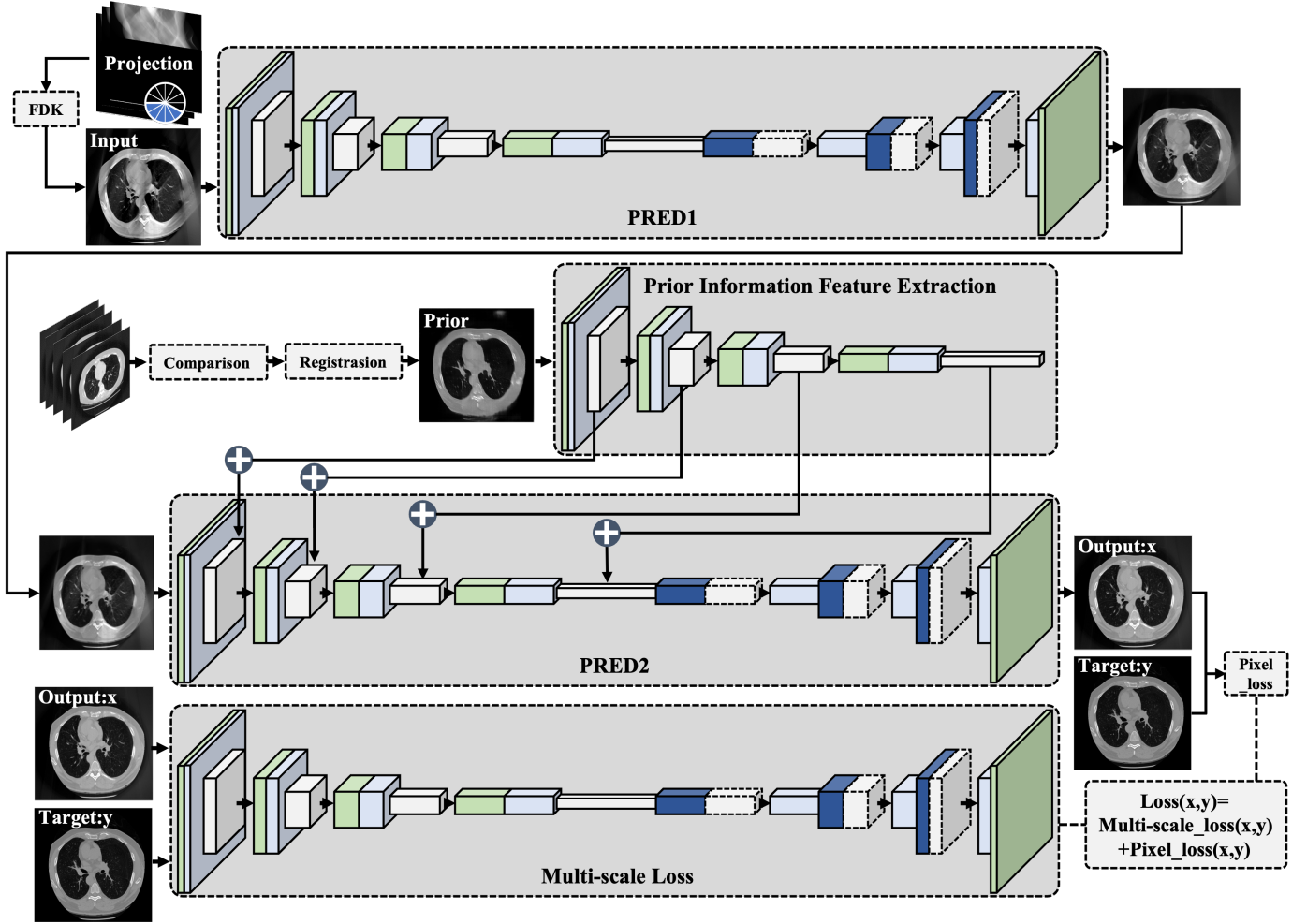


Fig. 1. Cas-PRED network. The network architecture is a two-stage cascaded encoder-decoder network. PRE1 and PRE2 perform coarse-grained and fine-grained reconstruction, respectively. The prior images are fed into the prior information feature extraction. Each layer of the feature extractor extracts prior features of different sizes. Layer-by-layer prior features are fed into PRE2 to assist PRE2 to incorporate prior information. In addition, pixel loss is fused with multi-scale loss to ensure that the network can effectively converge.

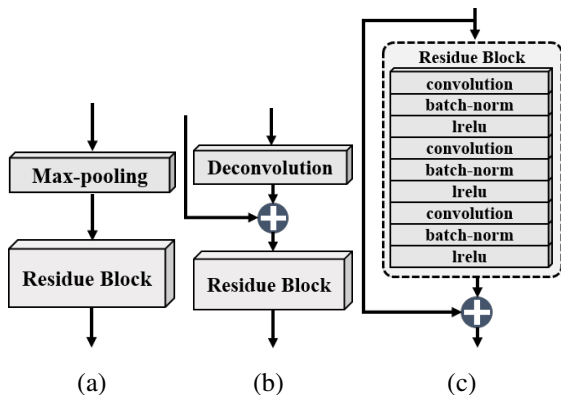


Fig. 2. The structure of encoder and decoder. (a) represents the structure of encoder. (b) represents the structure of decoder. (c) represents the structure of residue block.

full-angle cone beam reconstruction results can be used as prior information to guide the limited-angle reconstruction of other patients.

1) *Prior information choice:* We select the existing full-angle reconstruction results as the prior database for the limited-angle image reconstruction task. The 3D cone beam reconstruction results of CT images include multi-layer slices. For any limited-angle CT slice to be reconstructed, selecting the prior slice with the most significant similarity is necessary. Because the CT images of the same part have certain structure similarities, it can directly select in the image domain without using the deep learning network to extract features in higher dimensions for comparison. We use two indicators, namely MSE and SSIM, to measure the similarity between the limited-angle image and the CT image slices in the prior database:

$$MSE(x, y_i) = \frac{1}{M} \sum_{m=1}^M (x_m - y_m)^2 \quad (6)$$

$$SSIM(x, y_i) = \frac{(2\mu_x\mu_{y_i} + c_1)(2\sigma_{xy_i} + c_2)}{(\mu_x^2 + \mu_{y_i}^2 + c_1)(\sigma_x^2 + \sigma_{y_i}^2 + c_2)} \quad (7)$$

where  $x$  represents the current limited-angle CT image to be processed,  $y_i$  represents the  $i$ -th full-angle image selected from

the priori database,  $\mu_x, \mu_{y_i}$  represents the mean value of image  $x$  and image  $y$  respectively.  $\sigma_x, \sigma_{y_i}$  represents the variance value of image  $x$  and image  $y$ .  $\sigma_{xy_i}$  represents the covariance of two images.  $c_1$  and  $c_2$  are to maintain the calculation's stability and avoid the zero denominator.

The prior image with the highest MSE metric or SSIM metric can be selected and sent to the prior information extraction network to extract compelling prior features. After image comparison, there are specific differences in details due to scanning position, human structure, and other reasons. Therefore, the selected prior images need to be registered with the limited-angle images. The prior images can be adjusted to the proper size and orientation by image registration including amplification, minification, translation, and rotation. The process of comparison and registering is shown in Fig. 3.

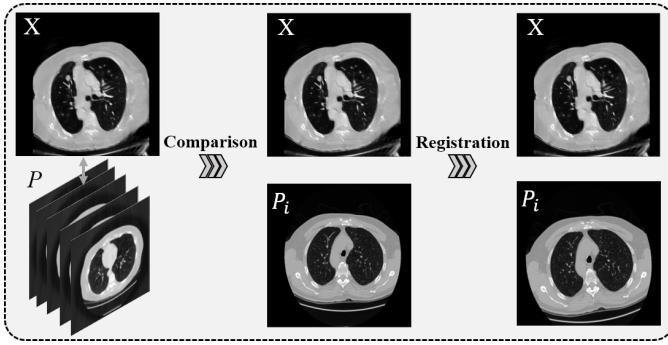


Fig. 3. Comparison and registration of prior images.

2) *Prior information extraction*: The prior information extraction network uses the encoder-decoder network structure. In order to reduce the training difficulty of the prior information extraction network and improve the stability of prior feature extraction, the prior information extraction network is pre-trained with the following steps: 1. The head and tail residual connections in the network are removed. 2. Each image in the prior database is input into the encoder network. 3. Features extracted from each encoder layer are sent to the decoder to train the whole network. After training, the encoder of the prior information extraction network is connected in parallel with the encoder in Cas-PRED to supplement prior features.

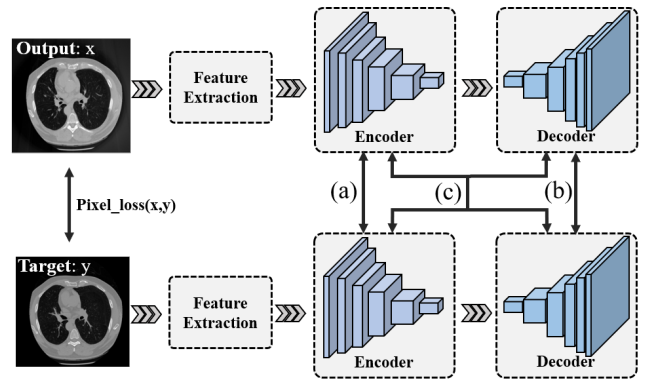
$$\theta^* = \underset{\theta}{\operatorname{argmin}}(L_2(f_{E \rightarrow D}^\theta(I_p), I_p)) \quad (8)$$

Eq. (4) describes the objective function of training, where  $f_{E \rightarrow D}$  is the parameter to be optimized. It can encode image  $I_p$ , and then reconstruct it to original images.

#### D. Multi-scale Loss function

In order to ensure that Cas-PRED can be effectively trained, a multi-scale loss function is introduced in this paper. Unlike the commonly used L1 or L2 loss functions, the multi-scale loss function measures the distance between layer-by-layer features. Since the traditional loss function can only calculate the error between the final output and the target result, the features of different depths extracted by each layer of Cas-PRED

cannot be fully utilized. The multi-scale loss can measure the feature error of each layer by calculating the layer-by-layer feature distance between the target and the output to guide the network better to learn the correct semantic information. This approach is similar to the idea proposed by Hamid Rezaatofghi *et al.* [49]. The computation of multi-scale loss functions requires both network inference output features and target features by a pre-trained encoder-decoder network. The loss function includes three structures: 1. The encoder multi-scale loss function computes encoder output features. 2. The decoder multi-scale loss function computes decoder output features. 3. the encoder-decoder multi-scale loss function that comprehensively computes encoder-decoder output features. The multi-scale loss function calculation process is shown in Fig. 4. Taking the encoder-decoder structure loss function as an example, its calculation formula is:



- (a)  $\text{Loss}(x, y) = \text{Multiscale}_{\text{encoder}}(x, y) + \text{pixel\_loss}(x, y)$
- (b)  $\text{Loss}(x, y) = \text{Multiscale}_{\text{decoder}}(x, y) + \text{pixel\_loss}(x, y)$
- (c)  $\text{Loss}(x, y) = \text{Multiscale}_{\text{en-decoder}}(x, y) + \text{pixel\_loss}(x, y)$

Fig. 4. The calculation process of multi-scale loss function. (a) represents the total loss function of  $\text{Multiscale}_{\text{encoder}}(x, y)$ . (b) represents the total loss function of  $\text{Multiscale}_{\text{decoder}}(x, y)$ . (c) represents the total loss function of  $\text{Multiscale}_{\text{en-decoder}}(x, y)$ .

$$L(x, y) = \lambda_0 \sum_{i=1}^N L_{MSE}(f'_i(x), f'_i(y)) + \lambda_1 L_{MSE}(x, y) + \lambda_2 L_{SSIM}(x, y) + \lambda_3 L_{MSE}(H(x), H(y)) \quad (9)$$

where  $f_i$  is the  $i$ -th layer feature extractor, and the last three parts include: the  $L_{MSE}$  error loss function between the output image and the target image; the  $L_{SSIM}$  between the output image and the target image; the high-frequency feature loss function  $L_{MSE}(H)$  between the output image and the target image.  $\lambda_0, \lambda_1, \lambda_2$  and  $\lambda_3$  are hyper-parameters that control the proportion of each loss function, thus ensuring that different loss function values can be adjusted to a similar scale. In this study, the values of  $\lambda_0, \lambda_1, \lambda_2$  and  $\lambda_3$  are 1.5, 1, 1000 and 0.2, respectively. Setting the hyper-parameters in this way can help the network to converge as soon as possible.

The calculation formula of the multi-scale loss function of the other two structures is similar. The difference is that the number of features and levels involved are different.

Features at different levels have different granularities of abstraction. Low-dimensional features are generally image color, grayscale, edge, corner, and other features. As the depth deepens, the features are gradually abstract and complex. Loss functions of different scales, numbers, and levels may prompt the network to focus on learning features at different levels of abstraction and then express this difference in the final result. In order to allow the network to learn the maximum number of layers and features, Cas-PRED uses the encoder-decoder structure loss function as the multi-scale loss function.

### E. Network design

The whole limited-angle CT image processing algorithm can be divided into three stages: data preparation stage, training stage and testing stage. The stage of data preparation is mainly to reconstruct by FDK method. The main work of the training phase is to use the data in the training set to train the designed network architecture. The main architecture of the network is a two-stage residual network, supplemented by a small encoder network that provides prior information. In the test phase, the test set data is sent to the trained network, and processed layer by layer to obtain the cone beam limited angle CT reconstruction results. The process is shown in Fig. 5 and as outlined in Algorithm 1.

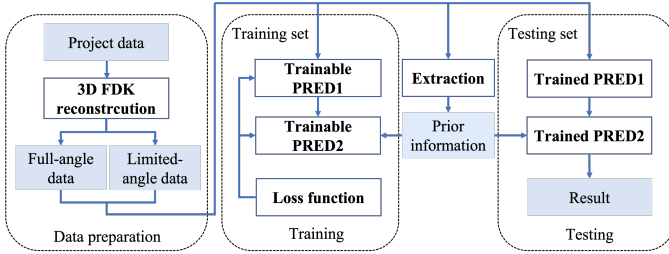


Fig. 5. The flowchart of Cas-PRED.

---

#### Algorithm 1 Cas-PRED

---

**Require:** The limited angle projection data  $l$ .  
 $x = FDK(l)$  Calculate image domain data.  
**repeat**  
 $x_1 = PRED1(x)$   
 $f = Registration(x)$  Select full angle data  $f$  that is closest to  $x$ .  
 $p = Prior(f)$  Extract prior features from  $f$ .  
 $x_{final} = PRED2(x_1, p)$  Combine feature  $p$  to enhance the reconstruction.  
 $loss = Calculate_{lossfunction}(x_{final}, y)$   
**until** converged

---

The Cas-PRED consists of a two-stage encoder-decoder network. The first stage network takes the cone beam limited-angle CT image generated by the FDK algorithm as input. The second stage network takes the output of the first stage network as the input. The prior characteristics of the full-angle images selected from the prior database can be integrated into the Cas-PRED. The training objectives of the whole network are:

$$\theta_1^* = \underset{\theta_1^*}{\operatorname{argmin}} L(f_I^{\theta_1^*}(I_{input}), I_{target}) \quad (10)$$

$$\theta_2^* = \underset{\theta_2^*}{\operatorname{argmin}} L(f_{II}^{\theta_2^*}(I_I, I_{prior}), I_{target}) \quad (11)$$

In Eq. (5),  $f_I^{\theta_1^*}$  and  $f_{II}^{\theta_2^*}$  indicate the parameters in the first level and the second level that need to optimize.  $[I_{input}, I_{target}]$  is a pair of input and target images.  $I_I$  and  $I_{prior}$  respectively represent the output result of the first stage and the complete angle CT image selected from the priori database for extracting prior features.  $L$  is the loss function to train the network. The two-level network uses the same loss function, namely:

$$L(x, y) = pixel\_loss(x, y) + multiscale_{en-decoder}(x, y) \quad (12)$$

$$pixel\_loss(x, y) = L_{MSE}(x, y) + \lambda_1 L_{SSIM}(x, y) + \lambda_2 L_{MSE}(H(x), H(y)) \quad (13)$$

$$multiscale_{en-decoder}(x, y) = \lambda_0 \sum_{i=1}^N L_{MSE}(f'_i(x), f''_i(y)) + L_{MSE}(x, y) + \lambda_1 L_{SSIM}(x, y) + \lambda_2 L_{MSE}(H(x), H(y)) \quad (14)$$

where  $H$  is a high-frequency information extraction operator, which is used to extract high-frequency information in image  $x$  and image  $y$ , such as edges, corners, etc., so that more high-frequency features can be retained and generated in the final processing results.

## III. RESULTS AND DISCUSSION

### A. Experimental conditions

We use three CT datasets in our experiments. The lung data are obtained from the AAPM dataset (from the low-dose CT challenge) and the LIDC-IDRI dataset (from the National Cancer Institute). The leg data are obtained from real patients in hospitals. The dataset consists of 23 human lung scans and 42 human leg scans. Among them, 50 sets are used as training sets, 10 sets are used as test sets, and 5 set is used as prior image data sets as shown in Table. I. The resolution of each data is  $512 \times 512 \times 512$ . For the evaluation, we conduct five-fold cross-validation. In order to verify the effectiveness of the proposed algorithm in cone beam CT limited-angle reconstruction, we generate cone beam projection data of the two datasets. The scanning geometry parameters include: the detector size is  $960 \times 960$ , each detector unit is  $0.45 \times 0.45 mm^2$ . On the basis of collecting the same number of projection data, the research goal is to obtain the highest quality limited-angle image. The full scan angle is  $360^\circ$ . A total of 600 projections are collected. We simulated and generated CBCT images reconstructed by several different limited-angle scanning methods. Taking the  $90^\circ$  limited angle as an example, we collect 150 sets of projections evenly spaced. The reconstruction data size is  $512 \times 512 \times 512$ . The algorithm is compared and verified on  $[0, 150^\circ]$ ,  $[0, 120^\circ]$ ,

TABLE I  
DETAILS OF TRAINING AND TESTING DATA SETS.

Category	Training/slices	Testing/slices	Prior/slices
Lung	8704	2560	512
Leg	16896	2560	2048

$[0, 90^\circ]$  limited-angles. In order to reduce the burden of GPU training, the image is trained by the slice.

All experiments are implemented using Tensorflow on one NVIDIA Tesla K80 GPU. FDK is completed with MATLAB on NVIDIA GeForce GTX1080-Ti GPU. Limited by the available GPU memory, the batch size is set to 16. The total epoch is set to 100. When the data set is enlarged, the initial value of the learning rate can be appropriately increased. In this study, the initial learning rate is set to  $10^{-3}$  based on experience. The learning rate decreases slowly as the training epoch increases from  $10^{-3}$  to  $10^{-5}$ . The Adam optimizer is used for training with setting  $\beta_1 = 0.9$ ,  $\beta_2 = 0.999$ . The filter used by the FDK algorithm is the Hann filter. In order to verify the effectiveness of the algorithm in this paper, the experimental results are compared qualitatively and quantitatively with FDK, total variation (TV), DDNet [50], and DCR [51]. DDNet is an encoder-decoder network. The encoder is a DenseNet and the decoder is a simple deconvolution layer. The DCR algorithm consists of two parts: first preliminary processing with the UNet network, and then reprocessing with the TV algorithm. For a fair comparison, all the competing methods were tuned according to the original papers to reach the best performance.

The quantitative comparison refers to three indicators: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), and root mean square error (RMSE). PSNR is used to measure the difference between two data and calculated as below:

$$MSE(x, y) = \frac{1}{M} \sum_{m=1}^M (x_m - y_m)^2 \quad (15)$$

$$PSNR(x, y) = 10 \log(L^2 / MSE) \quad (16)$$

where  $MSE$  means squared error of image  $x$  and image  $y$ , and  $L$  is the maximum gray value of the image pixel.

SSIM is used to measure the similarity of two data in brightness, contrast, and structure. Its formula is defined below:

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \quad (17)$$

where  $\mu$  means average value,  $\sigma^2$  means variance value,  $\sigma$  means covariance value, and  $c$  means variables to stabilize the division with a weak denominator.

RMSE is the algorithmic square root of MSE. Among the three indicators, the larger the result value of PSNR and SSIM, the higher the image quality. RMSE is the opposite.

### B. Evaluation for reconstruction result

For limited-angle reconstruction, a lot of effective information in CT images will be damaged by artifacts and noise,

and the original texture information will be lost. We use five algorithms (FDK, TV, DDNet, DCR, Cas-PRED) to process three cone beam CT reconstruction results with limited-angle. The processing results of some test data are shown in Fig. 6, Fig. 7, and Fig. 8. In order to more intuitively compare the effects of different algorithms, we calculate the mean and variance of PSNR, SSIM, and RMSE indicators of test results for each algorithm. The quantitative calculation results are shown in Table. II and Table. III. Regarding indicators, the processing results of any limited-angle FDK method are the worst. TV method is based on the traditional iterative method. Compared with the FDK method, it can retain the image details to a large extent. The processing results of DDNet can generally obtain high SSIM values. The DCR method first uses the UNet network for processing and then the TV iterative method for secondary processing of the processed results. The down-sampling in UNet better suppresses noise artifacts and loses some practical information. Therefore, DCR achieves relatively close results with the TV method regarding indicators. Among the several methods, the Cas-PRED has the best processing effect. In the  $[0, 150^\circ]$  limited-angle scanning scenario, compared with the FDK algorithm, the PSNR is increased by about 13dB, the SIMM is increased by about 0.2, and the RMSE is reduced by about 200HU. In other scenarios, although it is not superior to  $[0, 150^\circ]$ , it still achieves the best reconstruction effect among the several algorithms. It is proved that our algorithm effectively solves the problem of cone beam limited-angle reconstruction.

1) *Evaluation for reconstruction result of  $[0, 150^\circ]$* : Fig. 6 shows the processing results of  $150^\circ$  limited-angle data (a1)-(f1). (a2)-(f2) are the full-angle images. By comparing (a1)-(f1) and (a2)-(f2), it can be found that the location of artifacts is related to the location of missing projection data. For the reconstruction problem with a scanning angle of  $[0, 150^\circ]$ , the limited-angle artifacts mainly exist in the lower right corner of the image. Among several algorithms, the reconstruction effect of FDK is the worst. The image edge in the lower right corner is blurred and divergent. The reconstruction algorithm based on TV has some improvement effects. The image contrast is clearer. The edge divergence artifact has a certain inhibition effect but still cannot get a clear edge. Several algorithms based on deep learning can restore image edges. The DDNet method in Fig. 6(d1) can reconstruct clear edges and restore the bone tissue area covered by artifacts to a certain extent, although some divergent artifacts are still near the edges. The DCR method in Fig. 6(e1) can effectively suppress the divergent artifacts near the edges from the perspective of the artifact suppression effect. The artifacts covered on the pulmonary blood vessels have also been effectively suppressed, which can clearly show the details of blood vessels.

TABLE II  
QUANTITATIVE INDICATORS OF FIVE ALGORITHMS WITH DIFFERENT LIMITED-ANGLES.

Angle	Indicator	FDK	TV	DDNet	DCR	Cas-PRED
150°	PSNR/dB	20.58±0.76	24.73±0.34	28.93±0.17	29.49±0.55	33.69±0.90
	SSIM	0.7831±0.0083	0.8595±0.1090	0.9060±0.0020	0.8130±0.0220	0.9557±0.0029
	RMSE/HU	221.93±2.48	69.52±1.51	68.16±1.33	63.96±4.44	22.73±1.56
120°	PSNR/dB	17.96±0.76	21.11±0.40	20.26±1.11	22.35±0.39	25.63±0.81
	SSIM	0.6775±0.0131	0.7722±0.0132	0.7738±0.0148	0.7095±0.0126	0.8325±0.0050
	RMSE/HU	249.20±3.34	102.32±2.70	117.71±3.09	108.78±1.78	50.79±1.62
90°	PSNR/dB	15.86±0.60	18.75±0.54	18.27±0.94	18.34±0.44	24.96±0.82
	SSIM	0.5681±0.0101	0.6985±0.0134	0.6993±0.0175	0.5930±0.0204	0.8069±0.0077
	RMSE/HU	278.77±4.11	134.81±4.19	167.75±3.59	147.78±1.69	59.67±3.73

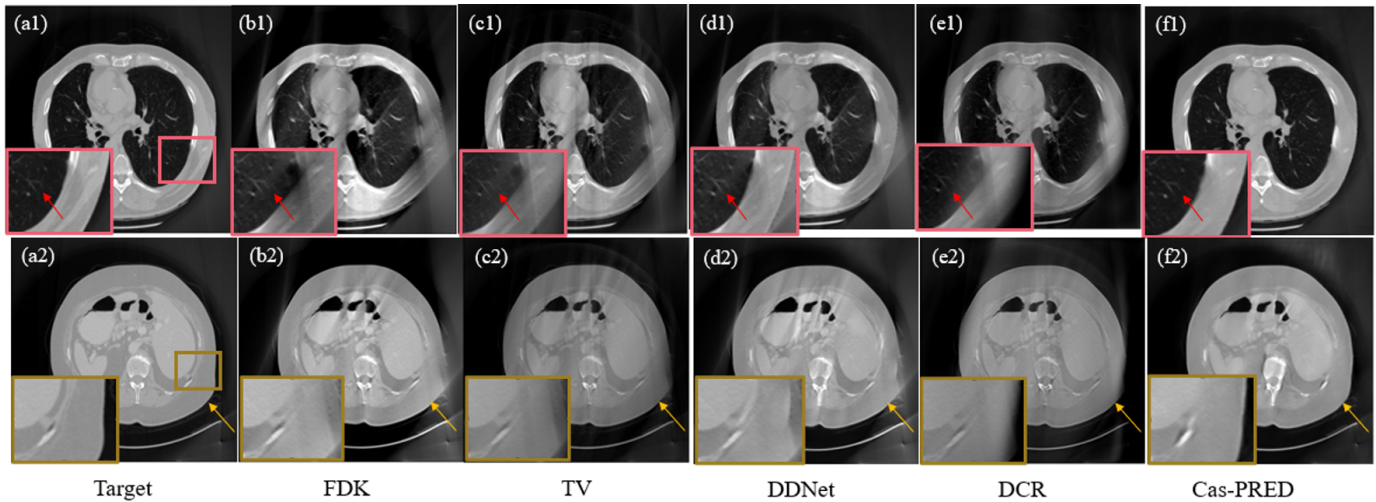


Fig. 6. Reconstruction result of  $[0, 150^\circ]$ . (a1) and (a2) are the target images. (b1) and (b2) are the FDK reconstruction results. (c1) and (c2) are the TV reconstruction results. (d1) and (d2) are the DDNet reconstruction results. (e1) and (e2) are the DCR reconstruction results. (f1) and (f2) are our Cas-PRED reconstruction results.

However, the internal edges are not clear. The bone tissue has not been effectively recovered, which exists in the form of brighter artifacts in the image. Fig. 6(f1) and Fig. 6(f2) are the results of the Cas-PRED algorithm proposed in this paper. The reconstructed image is richest in detail. The lung blood vessels have obtained processing results comparable to the reference image. The severely damaged bone tissue has also been effectively restored. The image edges are sharper compared to other algorithms. Fig. 6(f1) is from the LIDC-IDRI dataset, which is the processing result of human lungs. Fig. 6(f2) is from the AAPM data set, the human upper abdomen processing result. Comparing the results of the two test sets in Fig. 6(f1) and Fig. 6(f2), it can be shown that Fig. 6(f2) has a relatively better effect. This is because most of the samples in the training data are lung scanning images, while there are few abdominal images in the training samples. Our proposed Cas-PRED algorithm is the best comprehensive performance among the four methods. Several issues in finite-angle reconstruction: bright artifacts, edge divergence artifacts, and vitreous lung artifacts, can be resolved.

2) *Evaluation for reconstruction result of  $[0, 120^\circ]$* : Further, we processed the scanning reconstruction results with fewer limited-angle, i.e.,  $[0, 120^\circ]$ . The smaller the scan angle, the less projection data is obtained. Due to the severe loss of

practical information in the projection data, the reconstruction processing of the algorithm becomes more difficult. As shown in Fig. 7(b1), the FDK reconstruction results contain severe noise artifacts, especially in the lower left and upper right corners, ribs, and surrounding muscle tissues.

Fig. 7 shows the reconstruction results of  $[0, 120^\circ]$  limited-angle. Compared with the case of  $150^\circ$ , the semantic information is severely missing, and the reconstruction results of  $120^\circ$  have more serious artifacts and noise. The bones at the lower right are almost completely covered up. The lungs also have serious artifacts. Several methods other than FDK can restore image information to some extent. Compared with the FDK algorithm, although the TV method in Fig. 7(b1) requires repeated forward and back projection and takes a long time, the image quality is greatly improved, and some divergent artifacts can be removed. Processing algorithms based on deep learning perform relatively well. DDNet shows certain advantages in edge processing, but the effect of artifact suppression is worse than other deep learning algorithms. The DCR method showed good suppression of lung artifacts but could not solve the problem of high-density artifacts. In addition, since DCR needs to use the TV method for reprocessing, there are similar time-consuming drawbacks to the TV method.

Reconstruction methods usually only measure pixel-by-

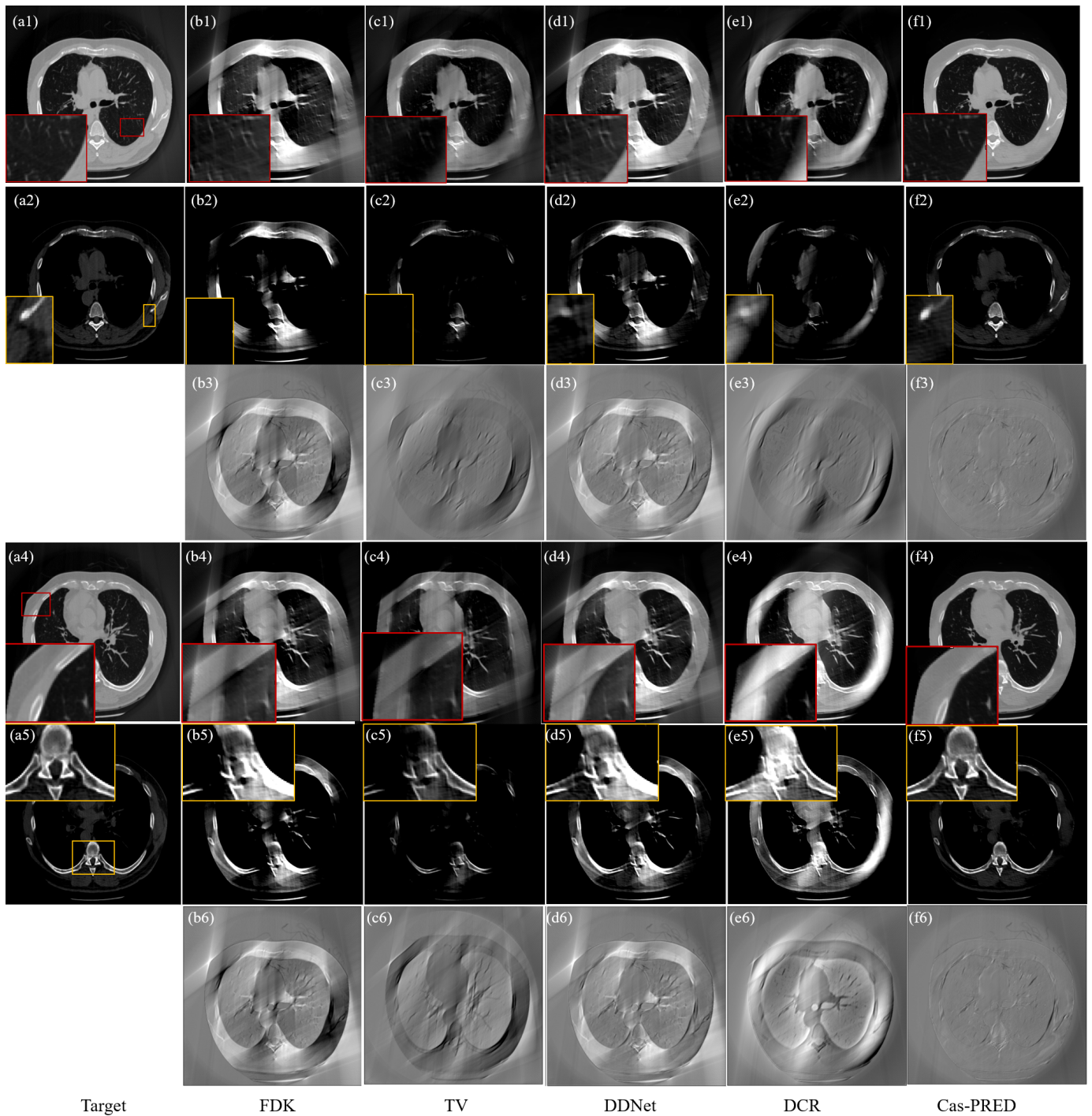


Fig. 7. Reconstruction result of  $[0, 120^\circ]$ . (a1)-(a5) are the target images. (b1)-(b6) are the FDK reconstruction results. (c1)-(c6) are the TV reconstruction results. (d1)-(d6) are the DDNet reconstruction results. (e1)-(e6) are the DCR reconstruction results. (f1)-(f6) are our Cas-PRED reconstruction results. The second and fifth rows are each algorithm's output results under the bone window width and window level. The third and sixth rows are the residual map between the output of each algorithm and the target image.

TABLE III  
QUANTITATIVE INDICATORS WITH THE  $[0, 60^\circ]$  LIMITED-ANGLE FOR LEG DATA.

Angle	Indicator	FDK	DDNet	DCR	Cas-PRED
60°	PSNR/dB	21.92±0.23	36.93±1.72	37.75±1.69	37.73±1.27
	SSIM	0.4975±0.0106	0.9342±0.0270	0.9419±0.0530	0.9524±0.0130
	RMSE/HU	160.40±4.13	29.12±7.03	26.43±5.36	26.25±4.02

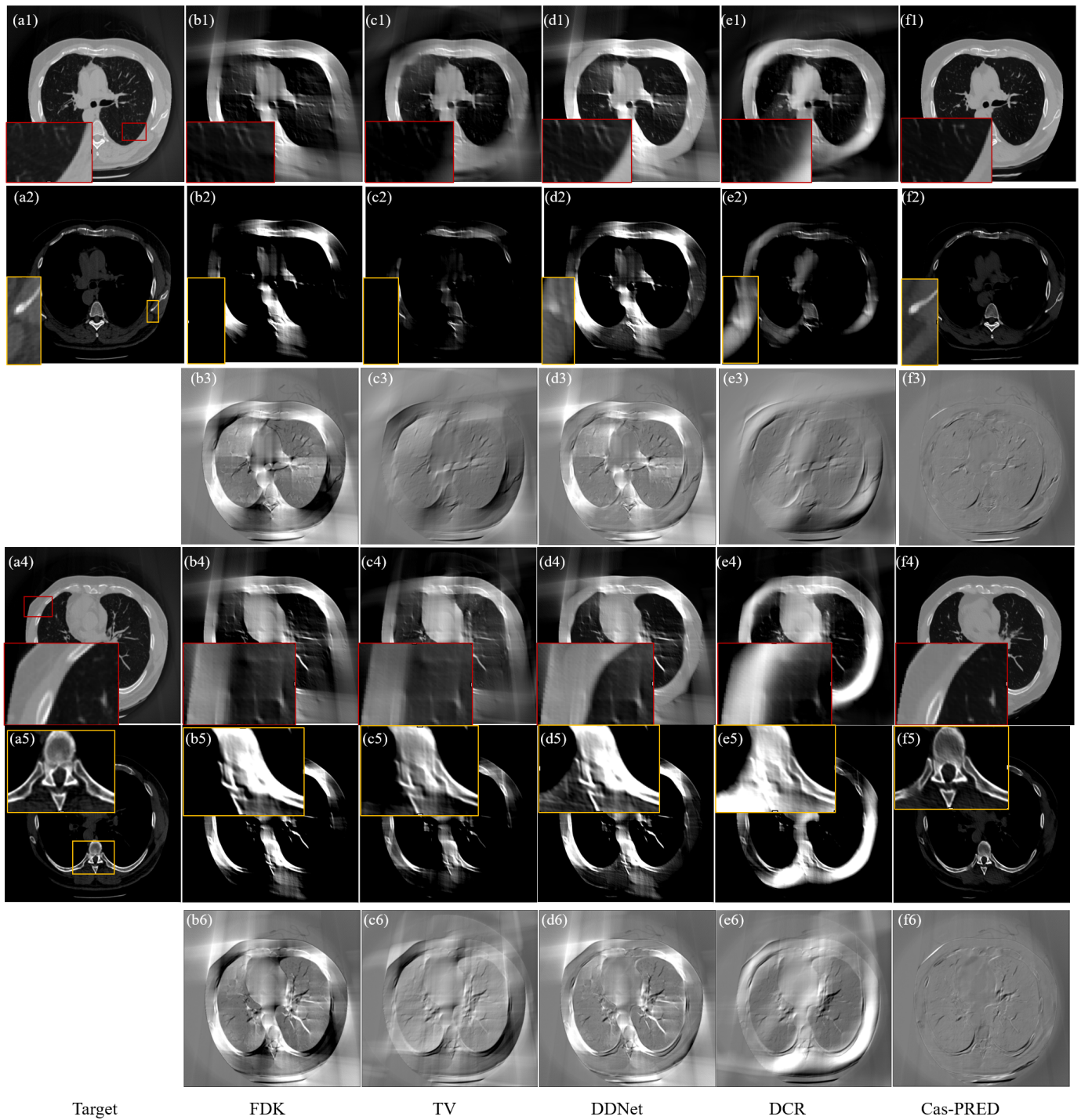


Fig. 8. Reconstruction result of  $[0, 90^\circ]$ . (a1)-(a5) are the target images. (b1)-(b6) are the FDK reconstruction results. (c1)-(c6) are the TV reconstruction results. (d1)-(d6) are the DDNet reconstruction results. (e1)-(e6) are the DCR reconstruction results. (f1)-(f6) are our Cas-PRED reconstruction results. The second and fifth rows are the output results of each algorithm under the bone window width and window level. The third and sixth rows are the residual map between the output of each algorithm and the target image.

TABLE IV  
QUANTITATIVE INDICATORS OF DIFFERENT NETWORK STRUCTURES.

Indicator	FDK	RED1	Cas-RED	Cas-PRED
PSNR/dB	15.86±0.60	23.74±0.77	25.18±0.66	24.96±0.82
SSIM	0.5681±0.0101	0.7368±0.015	0.8064±0.013	0.8069±0.0077
RMSE/HU	278.77±4.11	69.03±5.82	58.49±4.27	59.67±3.73

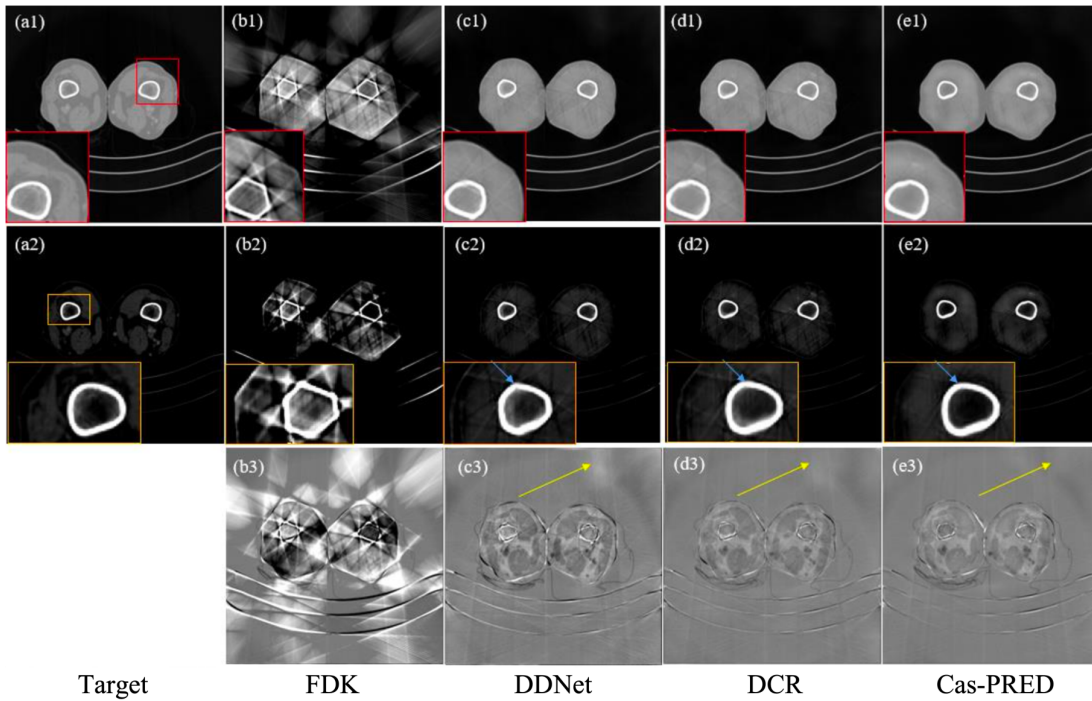


Fig. 9. Reconstruction leg result of  $[0, 60^\circ]$ . (a) is the target image. (b) is the FDK reconstruction result. (c) is the DDNet reconstruction result. (d) is the DCR reconstruction result. (e) is the Cas-PRED reconstruction result. The second row is the output results of each algorithm under the bone window width and window level. The third row is the residual map between the output of each algorithm and the target image.

TABLE V  
QUANTITATIVE INDICATORS OF DIFFERENT NETWORK STRUCTURES.

	PSNR/dB	SSIM	RMSE/HU
FDK	15.86±0.60	0.5681±0.0101	278.77±4.11
L2-loss	24.86±1.45	0.8055±0.0208	60.91±3.96
encoder-loss	24.61±1.32	0.8064±0.0155	63.62±3.45
decoder-loss	24.90±1.29	0.8089±0.0212	60.22±3.27
encoder-decoder-loss	24.96±0.82	0.8069±0.0077	59.67±3.73

TABLE VI  
COMPUTATIONAL COMPLEXITY OF DIFFERENT DEEP LEARNING METHODS.

Method	DDNet	DCR	Cas-PRED
Time(s)	0.04	0.06	0.12
Parameters(M)	0.16	2.76	0.74
FLOPs(G)	41.3	117.1	101.2

pixel errors, resulting in low generalization, while Cas-PRED can use both pixel-by-pixel errors and feature errors to make it easier for the network to learn semantic information with better universality. Among several processing methods, the Cas-PRED algorithm has the best effects. The pulmonary vascular structure can be completely restored without the glass artifacts. Outlines are sharp and correct, without high-density over-brightness artifacts in the image. The inner and outer boundaries of the chest are clear. The texture of the blood vessels in the lungs is similar to the reference image. Fig. 7(a2)-(f2) shows the processing results of the algorithm under the same slice bone window width and window level. It can be seen that Cas-PRED has a specific inhibition effect on the high-density artifacts near the bone structure in Fig. 7(f2).

In the enlarged ROI on Fig. 7(f2), Cas-PRED has achieved a better noise suppression effect than other reconstruction algorithms. The bones covered by artifacts can be restored to a certain extent, while the other methods can only get a mass of fuzzy highlight artifacts.

3) *Evaluation for reconstruction result of  $[0, 90^\circ]$* : Fig. 8 shows the reconstruction results of  $[0, 90^\circ]$ . This fewer reconstruction angles result in more noise and severe loss of image information. According to the reconstruction results of the FDK algorithm, it can be seen that the boundary artifacts are serious, especially in the upper left corner and the lower right corner, and the clear boundary cannot be seen. The muscle part has high-density artifacts caused by the influence of bones in a large range. The TV processing method can partially

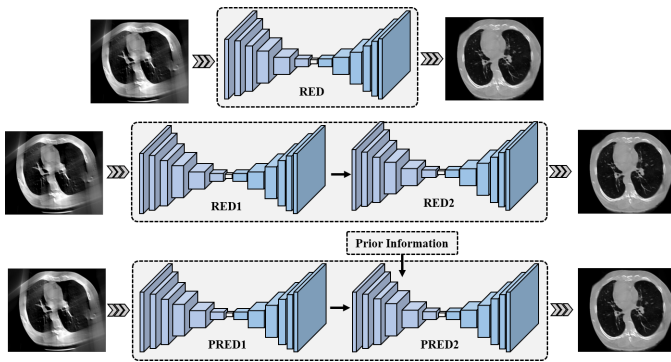


Fig. 10. The three different structures of network. (a) represents the structure of RED network. (b) represents the structure of Cas-RED network. (c) represents the structure of Cas-PRED network.

reconstruct the image boundary and suppress the high-density artifacts to a certain extent but cannot restore the covered anatomical structure. The boundaries of the processing results of DDNet are relatively clear. However, the artifacts around the border still need to be fixed. DDNet also fails to address lung artifacts. Although the DCR method can reconstruct the boundary, a certain gap exists between the reconstructed boundary and the real results. Because parts of the muscle tissue are wrongly considered artifacts by the DCR method, the result of the DCR method has a boundary contraction. Cas-PRED, which incorporates prior information and is trained with multi-scale loss, has better noise suppression effect. There are no obvious streak artifacts in the foreground, and boundaries are most consistent with the target image.

The muscle and bone texture of the FDK reconstruction results wholly disappeared under the bone window width and window level. DDNet and DCR methods can only recover blurred blobs. Among the five methods, the Cas-PRED performs best, especially the skeletal part can be reconstructed. However, because the gray values of different tissues in the muscle are too close, it is difficult for the Cas-PRED method to reconstruct the utterly correct muscle texture. As shown in the enlarged ROI in Fig. 8(f2), there is still some detail gap between the muscle texture around the bone and the real texture. Fig. 8(b3)-(f3), Fig. 8(b6)-(f6) are the residual maps between the processing results of each method and the target image. The grayscale residual effect of Cas-PRED is better. Only the upper left and lower right corners have minor mistakes in the reconstructed images of Cas-PRED.

4) *Evaluation for reconstruction result of  $[0, 60^\circ]$* : Fig. 9 shows the reconstruction results of  $[0, 60^\circ]$ . Fig. 9(b1) and (b2) depict the outcomes of the FDK reconstruction algorithm at both normal window width and window level, and bone window width and window level within the same slice. In Fig. 9(b3), we observe the disparity between the  $60^\circ$  limited angle reconstruction and the full angle reconstruction. In this context, the standard outer contour of the leg data transforms into an approximately hexagonal shape. Noticeably, severe artifacts manifest in the image's foreground and background, along with the appearance of hexagonal bright spots near the bones. Fig. 9(c1) and (c2) illustrate the results achieved with DDNet. Here, we can discern that the leg's outline

has been successfully restored and appears relatively clear. Furthermore, artifacts in the foreground and background have been effectively suppressed, though a noticeable background artifact remains present in the upper right corner. Fig. 9(d1) and (d2) showcase the results of the DCR algorithm. The DCR algorithm demonstrates superior background artifact removal compared to the other algorithms. However, a minor presence of strip artifacts can be noted in the foreground. Fig. 9(e1) and (e2) represent the results obtained using Cas-PRED, exhibiting the most effective artifact suppression. Within the same scenario, the training results from Cas-PRED surpass those from other algorithms.

### C. Ablation Experiments

As mentioned above, some innovative components are utilized, including cascaded network structure (Cas), prior-based residual encoder-decoder network structure (PRED), and multi-scale loss function. To validate the effectiveness of each component, we implement a series of ablation experiments on Cas-PRED in different settings. The ablation experiments are divided into two categories: the network structure experiments and the loss function experiments. The network structure experiments are shown in Fig. 10, Fig. 11, and Table. IV. The loss function experiments are shown in Fig. 12 and Table. V.

1) *Evaluation for network structure*: This section uses limited-angle data of  $[0, 90^\circ]$ . The three structures for comparison are the first-stage residual encoder-decoder (RED) network structure only, the cascaded residual encoder-decoder (Cas-RED) network structure without introducing prior information, and the complete Cas-PRED network structure. The flow of each structure is shown in Fig. 10. Fig. 10(a) only performs the first processing stage. Fig. 10(b) is a two-level cascaded residual encoder-decoder network removing the step of introducing prior information. Fig. 10(c) is the complete Cas-PRED network structure.

As shown in Fig. 11(c1), the RED network processing effect is relatively ambiguous. It can be seen that a RED network can roughly reconstruct the outline of the image. There are still block artifacts in the lower right corner and the background. In addition, the image details are blurred. For example, the bone tissue damaged by the artifact is restored as a blurred bright spot. The effect of the two-level cascaded network shown in Fig. 11(d1) is significantly improved. The outline of the image is clear and clump artifacts are removed. Some vascular details in the lungs can also be basically reconstructed. However, the skeletal tissue in the muscle in the lower left corner is relatively vague and has a large gap from the original structure. As shown in the enlarged skeleton in Fig. 11(d2), the shape of the middle part is blurred and the muscle texture is completely erased. After adding prior information, the visual effect of the result shown in Fig. 11(e1) is closer to the original image. The bone shape is closest to the actual reconstruction result. The muscle texture can be partially reconstructed. In the enlarged image of the ROI area in Fig. 11(e2), the processing result of the entire Cas-PRED network is closer to the actual image than the processing results of other network structures.

PSNR, SSIM, and RMSE metrics of different network structures are computed and shown in Table. IV. The recon-

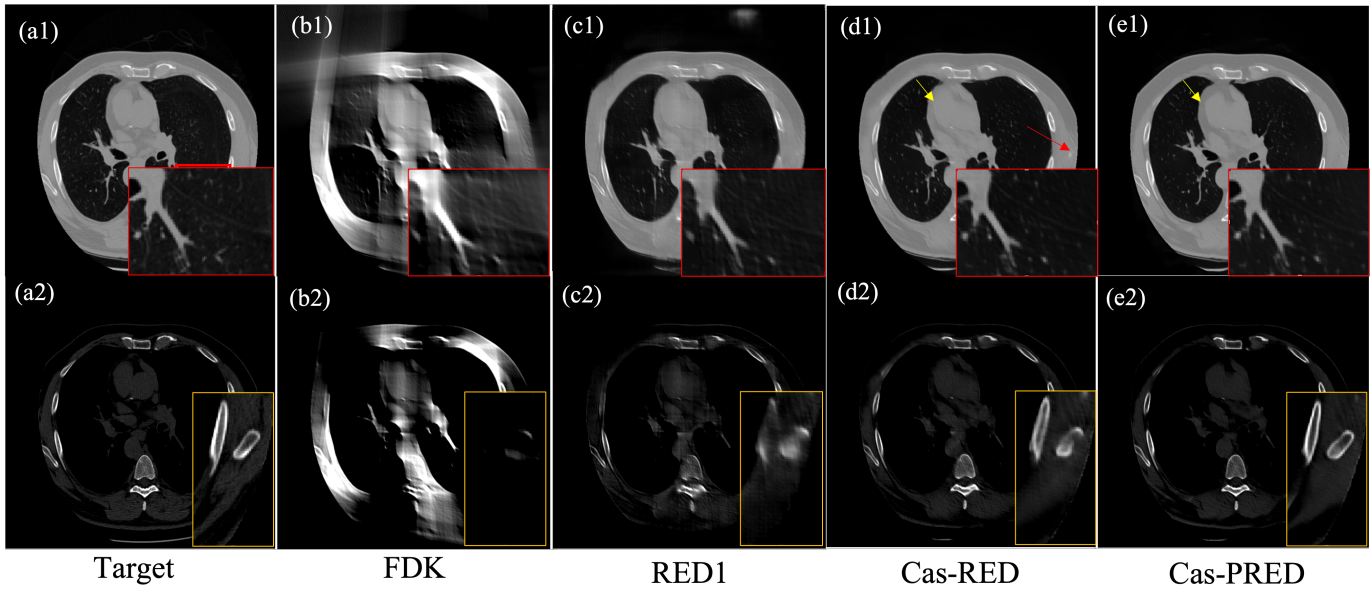


Fig. 11. Reconstruction result of different network structures with  $[0, 90^\circ]$ . The second rows is the output results of each algorithm under the bone window wide window level. The box in the lower right corner is the enlarged image of the ROI area.

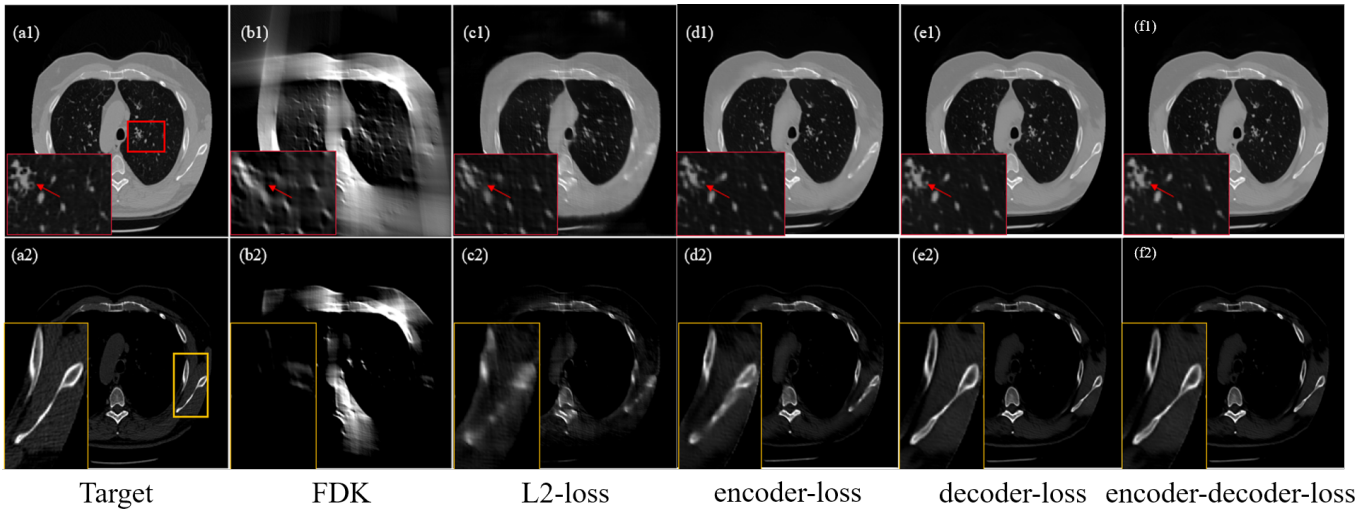


Fig. 12. Reconstruction result of different loss functions with  $[0, 90^\circ]$ . The second rows is the output results of each algorithm under the bone window wide window level. The box in the lower left corner is the enlarged image of the ROI area.

struction results of RED are greatly improved to that of the FDK algorithm. Although the processed images have problems such as unclear edges and clump artifacts, compared with the results of FDK, most artifacts can be suppressed. The two cascaded structures, Cas-RED and Cas-PRED, achieve better results than the RED network. It shows that the cascaded structure can improve the processing effect of the algorithm. The Cas-RED achieves slightly better statistical results on both metrics, while the Cas-PRED method achieves a higher SSIM metric. The difference in PSNR is about 0.2dB. The difference in RMSE is about 1HU. Cas-PRED introduces high-dimensional semantic information using prior information. High-dimensional semantic information can help images achieve better visual effects. For pixel-by-pixel PSNR and

RMSE metrics, introducing high-dimensional semantic information improves slightly. Since SSIM measures the structural similarity between images, Cas-PRED achieves more stable and better computational results on this metric.

2) *Evaluation for loss function:* The multi-scale loss functions of three different structures include encoder-loss (multi-scale loss function integrated into encoder features), decoder-loss (multi-scale loss function integrated into decoder features), encoder-decoder-loss (multi-scale loss function integrated into encoder-decoder features). In order to verify the effect of multi-scale loss functions with different structures, we perform verification on the  $[0, 90^\circ]$  limited-angle reconstruction results.

The experimental results are shown in Fig. 12. From the

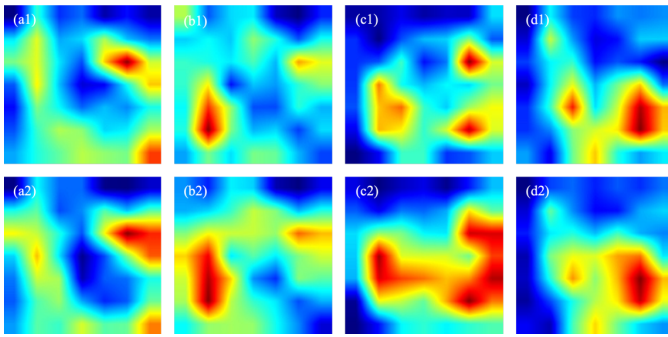


Fig. 13. Multi-scale feature heat maps between Cas-PRED and ground truth in  $[0, 90^\circ]$  scenario. The first row is the multi-scale loss output result of Cas-PRED. The second line is the multi-scale loss output result of ground truth. The first two columns are the encoder outputs of the multi-scale loss. The last two columns are the decoder outputs of the multi-scale loss.

overall visual effect of the image, the effects of Fig. 12(c)-(f) are relatively close. However, for the enlarged ROI area, it can be found that the reconstruction effect of only using the L2 loss function is not ideal without using the multi-scale loss function. As shown in the enlarged ROI area in Fig. 12(f), muscle tissue appears absent as circular holes. The processing results of training with several multi-scale loss functions can obtain the outer contour consistent with the target image. As shown by the ROI region pointed by the arrow, the muscle texture can be partially reconstructed using the multi-scale loss function. In contrast, the texture is blurred in the simple network processing result without multi-scale loss.

Fig. 13 illustrates the distinction between Cas-PRED results and Ground Truth when calculating multi-scale losses. The feature heat map shows that after introducing multi-scale loss, the network can measure the error between the network output result and the target result in the multi-scale feature dimension. With the same network architecture as a foundation, the network's ability to suppress artifacts is enhanced. For instance, within a scene constrained to a  $90^\circ$  angle, there is minimal differentiation between the Cas-PRED outcome and the feature heat map from Ground Truth during multiscale loss calculations. This substantiates that the utilization of multi-scale loss leads to more convergent results in network training.

In order to further compare the reconstruction effects of different loss functions, various index values are calculated and shown in Table. V. The decoder-loss incorporating the decoder features achieves the highest SSIM. The encoder-decoder-loss incorporating the encoder-decoder features achieves the highest PSNR and the smallest RMSE. Since the multi-scale loss function measures the error between features, the advantage of the multi-scale loss function lies in processing semantic information to obtain the most comprehensive features.

#### D. Computational Complexity

The average computational times required for different deep learning are summarized in Table. VI, along with the parameters counts and Floating-point Operations (FLOPs). The computational time includes the I/O operations. Due to the large gap between the reconstruction effect of traditional algorithms and deep learning algorithms, traditional algorithms

have not been included in the comparison of computational complexity. The Cas-PRED can bring good results within an appropriate time. In real-time experimental comparisons, Cas-PRED, while slightly more time-consuming, consistently delivers superior reconstruction performance. Notably, Cas-PRED can effectively meet real-time requirements when utilized on computing setups that are comparable in performance to the experimental configuration. It should be noted that the table only includes the time spent on the network calculations, but did not include the time spent on the additional forward projection and reconstruction processes of the proposed method.

#### IV. SUMMARY

CBCT technology boasts numerous advantages, including swift imaging, precise reconstruction of anatomical structures, and a non-invasive approach. It has unequivocally become an indispensable part of medical diagnostics. Nevertheless, certain scenarios present challenges in collecting a sufficient amount of projection data to produce clear images. Factors such as radiation dose constraints or the need to image dynamic organs often restrict us to acquiring projection data from limited angles. In the realm of CBCT reconstruction, restrictions related to radiation dose and the physical scanning environment frequently confine scans to angles typically below  $180^\circ$ . This limitation inevitably gives rise to pronounced artifacts in the resulting reconstructed images, rendering the task of reconstructing precise anatomical information a formidable one and, consequently, impairing diagnostic accuracy. Existing methods are constrained by their learning and representation capabilities, which hinder a complete resolution of this issue. In contrast, Cas-PRED employs a cascaded prior-based residual encoder-decoder architecture to learn from an expansive solution space through continuous training, thereby achieving high-quality limited-angle CBCT reconstructions. Cas-PRED uses full-angle CT images from a prior database to extract useful prior information. It can borrow the existing full-angle CT image features to guide the reconstruction of limited-angle tasks. The residual cascaded encoder-decode network structure used by Cas-PRED enables progressive image restoration for the limited-angle reconstruction task. The multi-scale loss function is introduced. The multi-scale loss function can measure the layer-by-layer feature error in the encoding and decoding to force the Cas-PRED to focus on more semantic information. Since the existing clinical limited-angle data is difficult to obtain the patient's full-angle data at the same time, the limited-angle reconstruction data by the method in this paper cannot be directly compared with the real full-angle reconstruction data. But in terms of the visibility of the reconstruction result, the reconstruction result has rich textures and details. The effectiveness of the Cas-PRED has also been verified in various limited-angle experiments such as  $[0, 150^\circ]$ ,  $[0, 120^\circ]$ ,  $[0, 90^\circ]$ , and  $[0, 60^\circ]$ . In general, the algorithm proposed in this paper can achieve limited-angle reconstruction, and have a good prospect in clinical applications.

Although the proposed method has significant advantages compared with other advanced methods, some problems still

need to be addressed. 1) Although the similarity between CT images is greater than that of natural images, there is still a large gap between CT slices of different people and different locations. In the future, we are committed to extracting more general and abstract prior information, which can be combined with convolutional features to facilitate network training. 2) Due to the limited-angle projection data obtained by scanning is the most direct information about the actual anatomical structure of the object, we will focus on integrating the projection data into the algorithm more fully while considering the calculation pressure of cone beam scanning. 3) Cone beam CT reconstruction is a three-dimensional reconstruction problem. Each CT data contains multiple slices, and there are rich semantic relationships between different slices. However, limited by the constraints of computing performance, this paper only studies the processing of two-dimensional images. Decomposing the computational pressure and realizing a three-dimensional processing network is a problem worthy of further research. 4) In future work, we will focus on collecting clinical limited-angle data corresponding to full-angle data to improve our method.

#### REFERENCES

- [1] Y. Song, S. Zheng, L. Li, X. Zhang, X. Zhang, Z. Huang, J. Chen, R. Wang, H. Zhao, Y. Chong *et al.*, "Deep learning enables accurate diagnosis of novel coronavirus (covid-19) with ct images," *IEEE/ACM transactions on computational biology and bioinformatics*, vol. 18, no. 6, pp. 2775–2780, 2021.
- [2] S. Wang, H. Yu, Y. Xi, C. Gong, W. Wu, and F. Liu, "Spectral-image decomposition with energy-fusion sensing for spectral ct reconstruction," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–11, 2021.
- [3] W. Wu, D. Hu, K. An, S. Wang, and F. Luo, "A high-quality photon-counting ct technique based on weight adaptive total-variation and image-spectral tensor factorization for small animals imaging," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2021.
- [4] M. Su, Y. Liu, Y. Xue, C. Qu, P. Wang, and Y. Zhao, "Detection method of pile foundation on subway lines based on cross-hole resistivity computed tomography," *Journal of Performance of Constructed Facilities*, vol. 34, no. 6, p. 04020103, 2020.
- [5] Y. Li, X. Zhao, Y. Han, and P. Chen, "Improving ct quality for complex objects with the novel autoexposure imaging of stepped voltage scanning," *IEEE Transactions on Instrumentation and Measurement*, 2022.
- [6] B. Liu and L. Zeng, "Parallel sart algorithm of linear scan cone-beam ct for fixed pipeline," *Journal of X-Ray Science and Technology*, vol. 17, no. 3, pp. 221–232, 2009.
- [7] T. Ito and S. Fujimura, "Simultaneous measurement of temperature and absorption of distributed medium by using infrared emission ct," in *Conference Proceedings. 10th Anniversary. IMTC/94. Advanced Technologies in I & M. 1994 IEEE Instrumentation and Measurement Technology Conference (Cat. No. 94CH3424-9)*. IEEE, 1994, pp. 257–260.
- [8] P. Chen, Y. Han, and Y. Li, "X-ray multispectrum ct imaging by projection sequences blind separation based on basis-effect decomposition," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–8, 2020.
- [9] M. H. Hesamian, W. Jia, X. He, and P. Kennedy, "Deep learning techniques for medical image segmentation: achievements and challenges," *Journal of digital imaging*, vol. 32, no. 4, pp. 582–596, 2019.
- [10] C.-H. Chiang, C.-L. Weng, and H.-W. Chiu, "Automatic classification of medical image modality and anatomical location using convolutional neural network," *Plos one*, vol. 16, no. 6, p. e0253205, 2021.
- [11] B. Newman and M. J. Callahan, "Alara (as low as reasonably achievable) ct 2011—executive summary," *Pediatric radiology*, vol. 41, no. 2, pp. 453–455, 2011.
- [12] X. Zhu, J. Shao, J. Shi, Z. You, S. Li, and Y. Xue, "The relationship between image quality and ct dose index of multi-slice low-dose chest ct," *Chinese Journal of Radiology*, vol. 37, no. 10, pp. 945–950, 2003.
- [13] C. Gong, L. Zeng, and C. Wang, "Image reconstruction model for limited-angle ct based on prior image induced relative total variation," *Applied Mathematical Modelling*, vol. 74, pp. 586–605, 2019.
- [14] T. Humphries, J. Winn, and A. Faridani, "Superiorized algorithm for reconstruction of ct images from sparse-view and limited-angle polyenergetic data," *Physics in Medicine & Biology*, vol. 62, no. 16, p. 6762, 2017.
- [15] L. A. Feldkamp, L. C. Davis, and J. W. Kress, "Practical cone-beam algorithm," *Josa a*, vol. 1, no. 6, pp. 612–619, 1984.
- [16] Z. Zhang, B. Chen, D. Xia, E. Y. Sidky, and X. Pan, "Directional-tv algorithm for image reconstruction from limited-angular-range data," *Medical Image Analysis*, vol. 70, p. 102030, 2021.
- [17] S. Xie, W. Huang, T. Yang, D. Wu, and H. Liu, "Compressed sensing based image reconstruction with projection recovery for limited angle cone-beam ct imaging," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 1307–1310.
- [18] W. Wang, X.-G. Xia, C. He, Z. Ren, and J. Lu, "A model-based deep network for limited-angle computed tomography image reconstruction," *Displays*, vol. 73, p. 102166, 2022.
- [19] D. Hu, Y. Zhang, J. Liu, C. Du, J. Zhang, S. Luo, G. Quan, Q. Liu, Y. Chen, and L. Luo, "Special: single-shot projection error correction integrated adversarial learning for limited-angle ct," *IEEE Transactions on Computational Imaging*, vol. 7, pp. 734–746, 2021.
- [20] C. Gong and L. Zeng, "Self-guided limited-angle computed tomography reconstruction based on anisotropic relative total variation," *IEEE Access*, vol. 8, pp. 70465–70476, 2020.
- [21] D. S.-C. Jin and J.-C. Chen, "Low-dose graphic-processing-unit based limited-angle ct reconstruction algorithm development for a home-designed dual modality micro-ft/ct system," in *2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC)*. IEEE, 2013, pp. 1–7.
- [22] X. Jin, L. Li, Z. Chen, L. Zhang, and Y. Xing, "Anisotropic total variation for limited-angle ct reconstruction," in *IEEE Nuclear Science Symposium & Medical Imaging Conference*. IEEE, 2010, pp. 2232–2238.
- [23] L. I. Rudin, S. Osher, and E. Fatemi, "Nonlinear total variation based noise removal algorithms," *Physica D: nonlinear phenomena*, vol. 60, no. 1–4, pp. 259–268, 1992.
- [24] W. Wu, D. Hu, K. An, S. Wang, and F. Luo, "A high-quality photon-counting ct technique based on weight adaptive total-variation and image-spectral tensor factorization for small animals imaging," *IEEE Transactions on Instrumentation and Measurement*, vol. 70, pp. 1–14, 2020.
- [25] Z. Liang, Y. Guan, G. Liu, R. Bian, X. Zhang, Y. Xiong, and Y. Tian, "Reconstruction of limited-angle and few-view nano-ct image via total variation iterative reconstruction," in *X-Ray Nanoimaging: Instruments and Methods*, vol. 8851. SPIE, 2013, pp. 156–162.
- [26] L. Zeng, J. Guo, and B. Liu, "Limited-angle cone-beam computed tomography image reconstruction by total variation minimization and piecewise-constant modification," *Journal of Inverse and Ill-Posed Problems*, vol. 21, no. 6, pp. 735–754, 2013.
- [27] W. Yu, C. Wang, and M. Huang, "Edge-preserving reconstruction from sparse projections of limited-angle computed tomography using l0-regularized gradient prior," *Review of Scientific Instruments*, vol. 88, no. 4, p. 043703, 2017.
- [28] M. Cao and Y. Xing, "Limited angle reconstruction with two dictionaries," in *2013 IEEE Nuclear Science Symposium and Medical Imaging Conference (2013 NSS/MIC)*. IEEE, 2013, pp. 1–4.
- [29] J. A. Tropp and A. C. Gilbert, "Signal recovery from random measurements via orthogonal matching pursuit," *IEEE Transactions on information theory*, vol. 53, no. 12, pp. 4655–4666, 2007.
- [30] M. Xu, D. Hu, F. Luo, F. Liu, S. Wang, and W. Wu, "Limited-angle x-ray ct reconstruction using image gradient l0-norm with dictionary learning," *IEEE Transactions on Radiation and Plasma Medical Sciences*, vol. 5, no. 1, pp. 78–87, 2020.
- [31] G. Litjens, T. Kooi, B. E. Bejnordi, A. A. A. Setio, F. Ciompi, M. Ghafoorian, J. A. Van Der Laak, B. Van Ginneken, and C. I. Sánchez, "A survey on deep learning in medical image analysis," *Medical image analysis*, vol. 42, pp. 60–88, 2017.
- [32] G. Wang, W. Li, M. A. Zuluaga, R. Pratt, P. A. Patel, M. Aertsen, T. Doel, A. L. David, J. Deprest, S. Ourselin *et al.*, "Interactive medical image segmentation using deep learning with image-specific fine tuning," *IEEE transactions on medical imaging*, vol. 37, no. 7, pp. 1562–1573, 2018.

- [33] S. Huang, M. Huang, Y. Zhang, J. Chen, and U. Bhatti, "Medical image segmentation using deep learning with feature enhancement," *IET Image Processing*, vol. 14, no. 14, pp. 3324–3332, 2020.
- [34] B. D. De Vos, F. F. Berendsen, M. A. Viergever, H. Sokooti, M. Staring, and I. Išgum, "A deep learning framework for unsupervised affine and deformable image registration," *Medical image analysis*, vol. 52, pp. 128–143, 2019.
- [35] A. Lin, B. Chen, J. Xu, Z. Zhang, G. Lu, and D. Zhang, "Ds-transunet: Dual swin transformer u-net for medical image segmentation," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–15, 2022.
- [36] H. Zhang, L. Li, K. Qiao, L. Wang, B. Yan, L. Li, and G. Hu, "Image prediction for limited-angle tomography via deep learning with convolutional neural network," *arXiv preprint arXiv:1607.08707*, 2016.
- [37] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.
- [38] Y. S. Han, J. Yoo, and J. C. Ye, "Deep residual learning for compressed sensing ct reconstruction via persistent homology analysis," *arXiv preprint arXiv:1611.06391*, 2016.
- [39] Q. Zhang, Z. Hu, C. Jiang, H. Zheng, Y. Ge, and D. Liang, "Artifact removal using a hybrid-domain convolutional neural network for limited-angle computed tomography imaging," *Physics in Medicine & Biology*, vol. 65, no. 15, p. 155010, 2020.
- [40] Z. Huang, J. Zhang, Y. Zhang, and H. Shan, "Du-gan: Generative adversarial networks with dual-domain u-net-based discriminators for low-dose ct denoising," *IEEE Transactions on Instrumentation and Measurement*, vol. 71, pp. 1–12, 2021.
- [41] Z. Li, W. Zhang, L. Wang, A. Cai, N. Liang, B. Yan, and L. Li, "A sinogram inpainting method based on generative adversarial network for limited-angle computed tomography," in *15th International Meeting on Fully Three-Dimensional Image Reconstruction in Radiology and Nuclear Medicine*, vol. 11072. SPIE, 2019, pp. 345–349.
- [42] S. Xie, H. Xu, and H. Li, "Artifact removal using gan network for limited-angle ct reconstruction," in *2019 Ninth International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, 2019, pp. 1–4.
- [43] Y. Zhang, D. Hu, T. Lyu, J. Zhu, G. Quan, J. Xiang, G. Coatrieux, S. Luo, and Y. Chen, "Pie-arnet: Prior image enhanced artifact removal network for limited-angle dect," *IEEE Transactions on Instrumentation and Measurement*, 2022.
- [44] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
- [45] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.
- [46] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 4700–4708.
- [47] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 2472–2481.
- [48] H. Chen, Y. Zhang, M. K. Kalra, F. Lin, Y. Chen, P. Liao, J. Zhou, and G. Wang, "Low-dose ct with a residual encoder-decoder convolutional neural network," *IEEE transactions on medical imaging*, vol. 36, no. 12, pp. 2524–2535, 2017.
- [49] H. Rezaatofghi, N. Tsoi, J. Gwak, A. Sadeghian, I. Reid, and S. Savarese, "Generalized intersection over union: A metric and a loss for bounding box regression," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2019.
- [50] Z. Zhang, X. Liang, X. Dong, Y. Xie, and G. Cao, "A sparse-view ct reconstruction method based on combination of densenet and deconvolution," *IEEE transactions on medical imaging*, vol. 37, no. 6, pp. 1407–1417, 2018.
- [51] Y. Huang, A. Preuhs, M. Manhart, G. Lauritsch, and A. Maier, "Data consistent ct reconstruction from insufficient data with learned prior images," *arXiv preprint arXiv:2005.10034*, 2020.