PROCEEDINGS OF SCIENCE

Sparse view CT reconstruction based on fusion learning in hybrid domain

Tian Haolai[a,b], Ling li[c], Yu Hu[a*], XiaoMeng Qiu[d], Tijian Deng[a], Gang Li[a], Fazhi Qi[a]

[a]Institute of High Energy Physics, Chinese Academy of Sciences,
Beijing 100049, China

[b]Spallation Neutron Source Science Center,
Dongguan 523803, China

[c]University of Chinese Academic of Sciences,
Beijing 100049, China

[d]Zhengzhou university,
Zhengzhou 450001, China

E-mail: huyu@ihep.ac.cn

In the synchrotron radiation tomography experiment, sparse-view sampling is capable of reducing severe radiation damages of samples from X-ray, accelerating sampling rate and decreasing total volume of experimental dataset. Consequently, the sparse-view CT reconstruction has been a hot topic nowadays. Generally, there are two types of traditional algorithms for CT reconstruction, i.e., the analytic and iterative algorithms. However, the widely used analytic CT reconstruction algorithms usually lead to severe stripe artifacts in the sparse-view reconstructed images, due to the Nyquist rule is not satisfied. While the more accurate iterative algorithms often result in prohibitively high computational costs and difficulty in selecting production parameters. In this paper, we propose a new hybrid domain method based on fusion learning which contain the image domain and projection domain. In the image domain, we propose a UNet-like network TransCovUNet which contains the Transformer module to consider the global correlation of the extracted features. In the projection domain, we employ a modified Laplacian Pyramid network to recover unmeasured data in the sinogram, which progressively reconstructs the sub-band residuals and can reduce the quantity of network parameters. Subsequently, we employ a deep fusion network to fuse the two reconstruction results at a feature-level, which can merge the useful information of the two reconstructed images. We also compared the performances of those single-domain methods and the hybrid domain method. Experimental results indicate that the proposed method is practical and effective for reducing the artifacts and preserving the quality of the reconstructed image.

International Symposium on Grids & Clouds 2022 (ISGC 2022)
21 - 25 March, 2022
Online, Academia Sinica Computing Centre (ASGC), Taipei, Taiwan

[*] Speaker

1. Introduction

The High Energy Photon Source (HEPS) is a new light source in China with high energy and high brightness [1], which is located in Beijing, 80 km from the institute of high energy physics, CAS. This project was officially approved in Dec. 2017 with construction beginning in late 2018 and completion in middle of 2025. The storage ring with electron energy of 6 GeV and emittance lower than 0.06nm×rad, would provide the synchrotron beam which will brilliance higher than 1×10^{22} phs/s/mm^2/mrad^2/0.1%BW. So, the HEPS experiments can generate massive amounts of data in a short time after it starts running. As an example, the hard X-ray imaging beamline of HEPS(HEPS-B7), can collect 10k projections (each 10k × 10k) in 100s. The data rate will reach 30 GB/s, which will make it easily to generate petabytes of measurement data. So, advanced methods are urgently needed that can reduce the amount of data collected, or feedback timely and permit real time determination of whether specific data are useful. On the other hand, for some experiment, for example, the solution reactions require fast detection, biological materials need to maintain in vivo indicators and rapidly detect, and the radiation dose received of light sensitive materials needs to be reduced. Sparse-view computed tomography (CT) can reduce the radiation dose in experimental sample, speeding up the data acquisition and reduce the total volume of measurement dataset. So sparse-view CT is a promise method to overcome the above difficulties on the future HEPS-B7. However, insufficient projection views in sparse-view CT will bring severe stripe artifacts in the conventionally analytic filtered back projection (FBP) [2] reconstruction. On the other hand, the iterative approaches [3-6] show an excellent noise reduction performance, but those approaches are often computationally expensive because of repetitive projections and back-projections during the iterative update procedure.

In recent years, artificial intelligence has developed rapidly and achieved great success in many fields, such as image classification [7], image segmentation [8], super-resolution [9], and image denoising [10], etc. In CT applications, deep learning are attracting more and more attention. Because of its excellent performance in solving inverse problems, it is increasingly applied to the ill-posed inverse problem of CT reconstruction. In [11], Würfl et al. mapped the FBP algorithm to an artificial neural network (ANN), which can reconstruct CT data with limited views and show consistent improvement over the FBP method with the same computational complexity. Jin et al. [12] proposed a CT image reconstruction strategy named “FBPConvNet”, which is based on modified UNet [13] and residual learning and using FBP reconstruction of sparse view CT projection as input. Experimental results show that this method has better imaging performance than the iterative reconstruction method based on total variation constraints. UNet can extract features through continuous downsampling with the encoder, and then use the decoder to gradually upsample the features output from the encoder through skip connections, so that the network can obtain features of different granularities. In view of the powerful decoding and encoding capabilities of UNet, many new models designed for CT image reconstruction or segmentation problems are based on the UNet structure, and have achieved good performance, such as UNet++[14], Res-UNet[15], Attention U- Net [16], ResAtt-UNet [17], etc. The models described above all rely on the convolutional neural network (CNN) structure. CNNs have dominated a series of medical imaging tasks. However, due to the inherent inductive bias, each convolution kernel can only focus on a sub-region in the whole image, which makes it lose the global context connection and cannot build long-range dependencies.

Recently, a novel artificial neural network structure Transformer [19] was proposed. This model is designed for sequence-to-sequence modeling in natural language processing (NLP) tasks. Transformer's Multi-headed Self-attention (MSA) can effectively establish global connections between sequences. Transformers have revolutionized most NLP tasks such as machine translation, named entity recognition, and question answering systems. The great success of Transformers in natural language has prompted researchers to explore their applicability in computer vision. But it faces great challenges when transferring its efficient performance in the natural language domain to the vision domain. Pixels in images have much higher resolution than words in text passages. Vision tasks such as image segmentation require pixel-level dense predictions. The computational complexity of Transformer self-attention is quadratic of the image size, which makes it difficult to handle high-resolution images. To reduce the computational complexity, a hierarchical neural network model Swin Transformer [20] is proposed based on the structure of windowed multi-head self-attention layer (W-MSA) and shifted window multi-head self-attention layer (SW-MSA). The model surpasses previous state of the art (SOTA) methods in dense prediction tasks such as image classification, object detection, and semantic segmentation. Subsequently, the U-shaped network Swin-UNet [21] with Swin Transformer as the basic unit of encoding and decoding was proposed for medical image segmentation, showing good performance and generalization ability. Although Transformer has made significant progress in the field of images, it still has some limitations. Transformer does not provide an up-sampling method similar to deconvolution and rely on other interpolation methods. Transformer couldn’t share the weight as CNN, and is at a disadvantage in computational overhead. Transformer uses a block patch method to deal with image problems, ignoring pixel-level internal structural features in blocks.

One the other hand, some approaches attempt to solve the ill-posed inverse problem of CT image reconstruction from the projection domain [22,23]. Those approaches often use a pre-defined upsampling operator, e.g., bicubic interpolation, to upsample the sparse view sinogram to the desired full view before applying the network for prediction. This preprocessing step increases unnecessary computational cost and often results in visible reconstruction artifacts.

To overcome these problems, we propose a new hybrid domain method based on fusion learning. In the image domain, we propose a UNet-like network (TransCovUNet) which contains the Transformer module to consider the global correlation of the extracted features. In the projection domain, we employ a Laplacian Pyramid network inspired by LapSRN [24] to recover unmeasured data in the sinogram, which progressively reconstructs the sub-band residuals and can reduce the quantity of network parameters. At last, we employ a deep fusion network DenseFuse [25] to fuse the two reconstruction results at a feature-level, which can merge the useful information of the two reconstructed images.

The structure of this paper is as follows. We first introduce the proposed neural network structure and its basic unit modules. Then the experimental method used to study the performance of this network model is illustrated, and the performance of other existing advanced network models is compared, and finally a summary and discussion.

2. Method

2.1 Overview

The framework of the proposed method is shown in Fig. 1, which consists of three components in total: image domain part, projection domain part and the fusion part. In the image domain, the sparse view CT projection sinogram will first be reconstructed by the FBP algorithm to obtain a reconstructed image with stripe artifacts. The image with artifacts will then be processed by a pre-trained deep learning neural network to get pure artifact image. Finally, the pure artifact image will be subtracted from the reconstructed image with artifacts to obtain the final corrected high-definition reconstructed image. In the projection part, the sparse view CT projection sinogram will first be interpolated by a pre-trained deep learning neural network to get full view CT projection sinogram. The estimated full view CT projection sinogram will then be reconstructed by the FBP algorithm obtain the final corrected high-definition reconstructed image. At last, these two reconstructions will be fused by a pre-trained deep learning neural network. Neural networks require supervised training on the training dataset in advance.

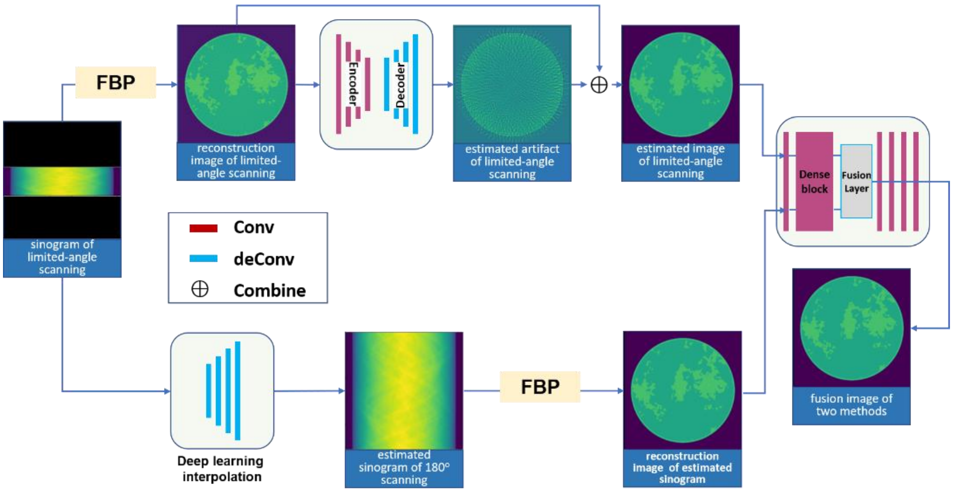
FBP
reconstruction image of limited-angle scanning
Encoder
Decoder
estimated artifact of limited-angle scanning
⊕
estimated image of limited-angle scanning
Dense block
Fusion Layer
sinogram of limited-angle scanning
Conv
deConv
⊕ Combine
fusion image of two methods
Deep learning interpolation
estimated sinogram of 180° scanning
FBP
reconstruction image of estimated sinogram

Figure 1: The overview of the hybrid domain sparse-view CT reconstruction method.

2.2 The archtecture of TransCovUNet

The network structure of proposed TransCovUNet is shown in Fig. 2. TransCovUNet consists of encoder, decoder and skip connections. TransCovUNet first divides the (W×H) input image into multiple non-overlapping patches. In our implementation, the patch size is set to 4×4. Each patch can be regarded as a 'token', so the feature dimension of each 'token' (patch) is 4×4=16, and the image segmentation layer will get (W/4×H/4) 'tokens' (patchs). A linear embedding layer then maps the features of each 'token' to the desired dimension (denoted as C). The transformed 'tokens' (patchs) are then successively passed through a sequence of Swin Transformer modules and patch merging layers, resulting in a hierarchical feature representation. Among them, the Swin Transformer module is responsible for feature representation learning, and the block merging layer is responsible for downsampling. Inspired by the UNet network, we adopt a decoder structure that is symmetric with the encoder. The difference is that the basic unit of the decoder is composed of several convolutional layers module and a deconvolutional layer. The convolutional layers are responsible for feature representation learning, and the deconvolutional layer is responsible for the upsampling operation, which reshapes the feature map into a large feature map with twice the resolution. The contextual features extracted by the decoder are fused with the multi-scale features of the encoder through skip connections to compensate for the loss of spatial information caused by the downsampling process. Finally, two times of upsampling is performed through two convolution and deconvolution operations, and the resolution of the feature map is restored to the resolution of the input image (W×H).

We employed the Mean Squared Error (MSE) as the loss function. The MSE is used to measure the difference between the network output and the artifact label. The MSE loss function is generally defined as:

$$\underset{\theta}{min}L = \|I_{AF} - F_{\theta}(I_{LD})\|_2^2 \tag{1}$$

where the I_{AF} is the the pure artifact image, I_{LD} is the FBP reconstructed image from a sparse view sinogram, F is the nueral network, θ present the parameters of network, $\|\ \|_2$ is the ℓ_2 norm.

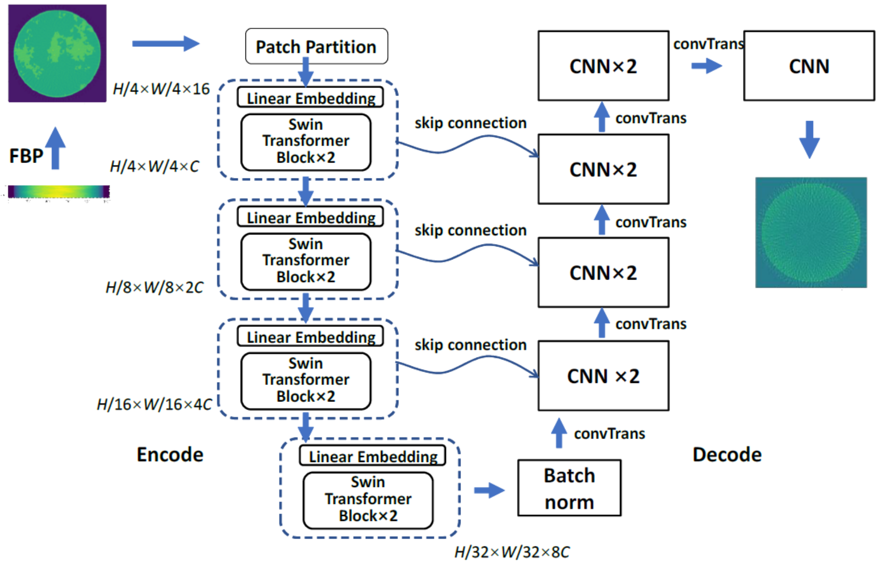

Figure 2: The network architecture of the TransCovUNet.

2.3 The archtecture of modified LapSRN

The network structure of proposed modified LapSRN is shown in Fig. 3. The network is constructed based on the Laplacian pyramid framework. The model takes an sparse view CT projection sinogram as input (rather than an upscaled version of the sparse view CT projection sinogram) and progressively predicts residual images at $\log_2 S$ levels where S is the scale factor in the y direction. For example, the network consists of 3 sub-networks for expanding an sparse view CT projection sinogram at a scale factor of 8 in the y direction. This model has two branches: (1) feature extraction and (2) image reconstruction.

On the feature extraction branch, at each level, the feature extraction branch consists of several convolutional layers and one transposed convolutional layer to upsample the extracted features by a scale of 2 in the y direction. The output of each transposed convolutional layer is connected to two different layers: (1) a convolutional layer for reconstructing a residual image at the same level, and (2) a convolutional layer for extracting features at the next level. The modified LapSRN perform the feature extraction at the coarse resolution and generate feature maps at the finer resolution with only one transposed convolutional layer. In contrast to other projection approach that perform all feature extraction and reconstruction at the fine resolution, the modified LapSRN design significantly reduces the computational complexity.

On the image reconstruction branch, at each level, the input image is upsampled by a scale of 2 in the y direction with a transposed convolutional (upsampling) layer. This layer is initialized with the bilinear kernel and allow to be jointly optimized with all the other layers. The upsampled image is then combined (using element-wise summation) with the predicted residual image from the feature extraction branch to produce a high-resolution output image. The output HR image at each level is then fed into the image reconstruction branch of next level. The entire network is a cascade of CNNs with a similar structure at each level.

The loss funtion is defined as:

$$\min_{\theta} L = \sum_{s=1}^{L} \rho(\hat{I}_s - I_s) \tag{2}$$

Where I_s is the ground truth of each layer, which is uniform downsampled from full view sinogram with corresponding scale. $\hat{I}_s$ is predicted high-resolution sinogram. $\rho(x)=\sqrt{x^2+\varepsilon^2}$ is the Charbonnier penalty function (a differentiable variant of ℓ_1 norm), the ε is empirically set to 1e-3. L is the number of level in the pyramid.

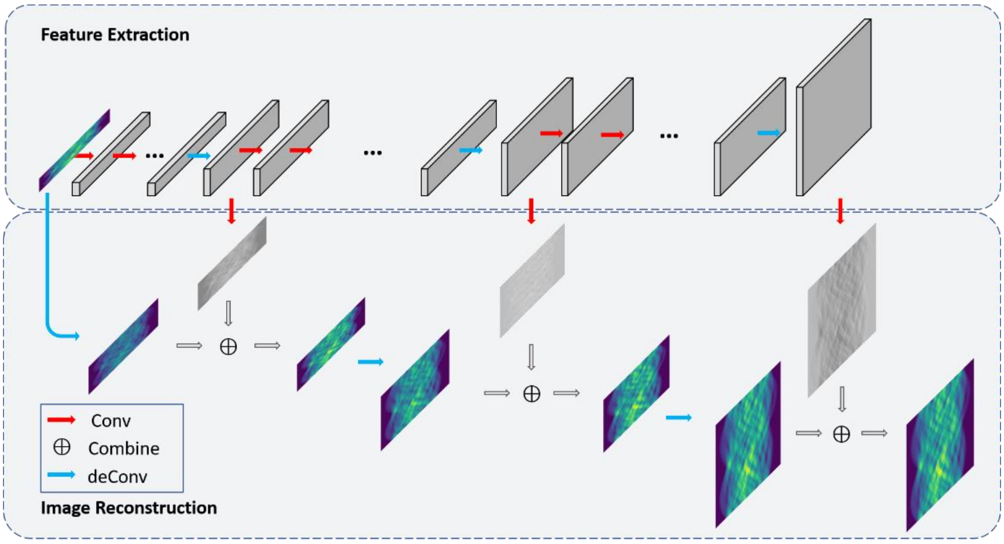

Figure 3: The network archtecture of the modified LapSRN.

2.4 The archtecture of DenseFuse

We employ the DenseFuse to fuse the reconstruction from the image domain and projection domain. As shown in Fig. 4, the DenseFuse consist of three parts: encoder, fusion layer, and decoder. The encoder are utilized to extract deep features which contains two parts: C1 and DenseBlock. The first layer C1 contains 3 × 3 filters to extract rough features. The dense block contains three convolutional layers which also contain 3 × 3 filters. Each layer's output is cascaded as the input of the next layer. For each convolutional layer in encoding network, the input channel number of feature maps is 16. For the encoder, the filter size and stride of convolutional operation are 3×3 and 1, respectively, which make the input image can be any size.

The dense block architecture can preserve deep features as much as possible in encoding network which can make sure all the salient features are used in fusion strategy.

The Fusion Layer adopt the ℓ_1-Norm Strategy, which is based on ℓ_1-Norm and soft-max operation. Here we denote the features maps by ϕ_i^m , the activity level map $\hat{C}_i$ will be calculated by ℓ_1-norm and block-based average operator. The fused feature maps are denoted as f^m . the initial activity level map C_i is calculated by:

$$C_i(x,y) = \left\|\phi_i^{1:M}(x,y)\right\|_1 \tag{3}$$

Then the final activity level map is calculated by block-based average operator :

$$\hat{C}_i(x,y) = \frac{\sum_{a=-r}^{r}\sum_{b=-r}^{r} C_i(x+a,y+b)}{(2r+1)^2} \tag{4}$$

where r determines the block size and is set to 1 in our strategy.

Then the f^m is calculated by:

$$f^m(x,y) = \sum_{i=1}^{k} \omega_i(x,y) \times \phi_i^m(x,y), \tag{5}$$

$$\omega_i(x,y) = \frac{\hat{C}_i(x,y)}{\sum_{n=1}^{k} \hat{C}_n(x,y)}$$

The final fused image will be reconstructed by decoder in which the fused feature maps f^m as the input.

The decoder contains four convolutional layers with 3×3 filters, respectively, which is used to reconstruct the final fused image. The output of fusion layer f^m will be the input of decoder.

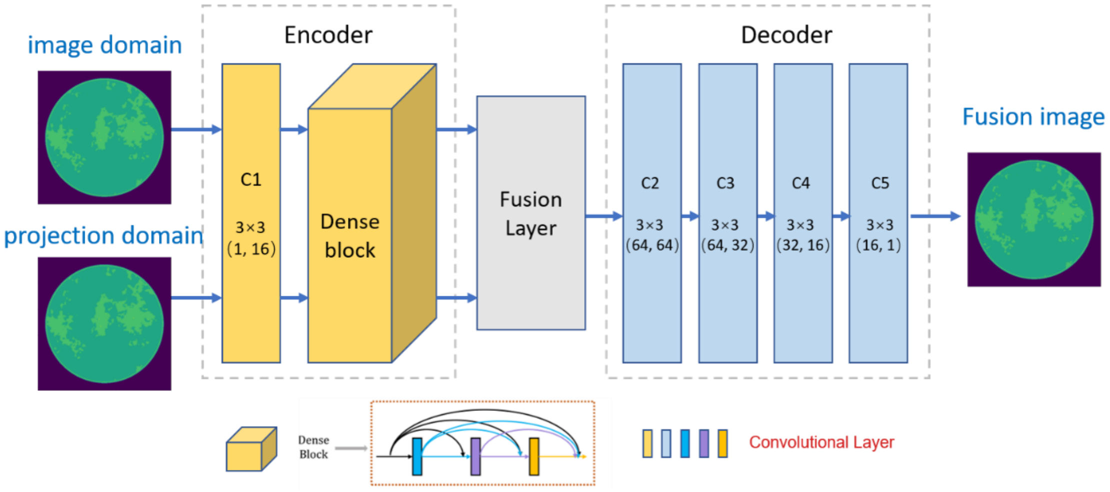

Figure 4: The network archtecture of the DenseFuse.

3. Implementation

3.1 Dataset

In this study, we use two datasets: simulated breast CT dataset and simulated foam CT dataset.

The simulated breast CT dataset were collected from DL-sparse-view CT challenge. The details of how the simulation images and data are generated are provided in Ref. [26]. The dataset contains 4000 cases where each case consists of the truth image, and the corresponding 128-view FBP image. This dataset is used to verify the effectiveness of the proposed TransCovUNet model.

The foam CT dataset is simulated by tomophantom [27], in which the phantom is a combinations of several geometrical objects. The dataset contains 2400 cases where each case consists of the truth image, the corresponding 1024-view sinogram, the corresponding 128-view sinogram, and the corresponding 128-view FBP image. This dataset is used to verify the effectiveness of the proposed hybrid domian method.

3.2 The implementation and trainning of the models.

During training, we emploied the Adaptive Momentum Estimation (Adam) to minimize the loss function and optimize the models' parameters. The exponential decay rates of the first and second moments are 0.9 and 0.999, respectively, ε is 1E-7. The network is trained for 400 epochs using the mini-batch training method with a batch size of 1. The initial learning rate is set to 0.0001 and then decays by 1% every 10 epochs.

The models are implemented based on the deep learning framework Pytorch, and trained on a graphic workstation with the CPU of Intel(R) Xeon(R) Silver 4114 @ 2.20 GHz, and GPU of NVIDIA Titan V 12G.

3.3 Metric

To objectively evaluate the performance of the models, this article employs Structural Similarity Index (SSIM), Peak Signal-to-Noise Ratio (PSNR), Root Mean Square Error (RMSE), the worst-case 25×25 pixel ROI-RMSE(WC_ROI-RMSE) to quantitatively analyze each model. The SSIM is defined as follows:

$$SSIM(I,K)=\frac{(2\mu_I\mu_K+c_1)(2\sigma_{IK}+c_2)}{(\mu_I^2+\mu_K^2+c_1)(\sigma_I^2+\sigma_K^2+c_1)} \tag{6}$$

where *K* represents the image with stripe artifacts; *I* represents the standard image; μ_I and μ_K are the averages of *I* and *K*, respectively; σ_I and σ_K are the variances of *I* and *K*, respectively, and σ_{IK} is the covariance of *I* and *K*; c_1 and c_2 are constants. The SSIM takes values in the range [0,1] which is to ascertain how similar a sparse view reconstruction is to the full view images. A value of 0 implies that there is no correlation between images, while avalue of 1 implies that two images are identical.

The PSNR is defined as follows:

$$\mathrm{PSNR(I,K)}=10log_{10}\left(\frac{MAX_K^2}{\|I-K\|_2^2}\right) \tag{7}$$

where the MAX_I represents the maximum value of the pixel in the image *I.* The PSNR represents the ratio between the maximum possible power of an image and the power of corrupting noise that affects the quality of its representation. The higher the PSNR means the better image has been reconstructed to match the original image.

The RMSE is defined as follows:

$$RMSE(I,K)=\sqrt{\|I-K\|_2^2} \tag{8}$$

The WC_ROI-RMSE is defined as follows:

$$WC_ROI - RMSE(I,K) = max\sqrt{\frac{\|b_c(I-K)\|_2^2}{m}} \qquad (9)$$

where b_c is a masking indicator function for the 25x25 ROI centered on coordinates c in the image, and m is the number of pixels in the local area.

4. Result

In image domain, to verify the effectiveness of TransCovUNet, we selected 100 sets of CT images for testing and compared with FBP and advanced deep learning methods such as Unet and Swin-Unet. The reconstructed CT image from the uniformly selected 128 views sinogram is used as the input of the models, and processed by the TransCovUNet, Unet and Swin-Unet respectively. In the experiment, the computing time for TransCovUNet, UNet, Swin-UNet and FBP to process a single image is 0.017s, 0.007s, 0.025s and 0.004s, respectively. It can be concluded that the operating efficiency of the classic UNet is better than the model with the Transformer module, because the Transformer module contains a large number of fully connected operations, resulting in increased computational overhead.

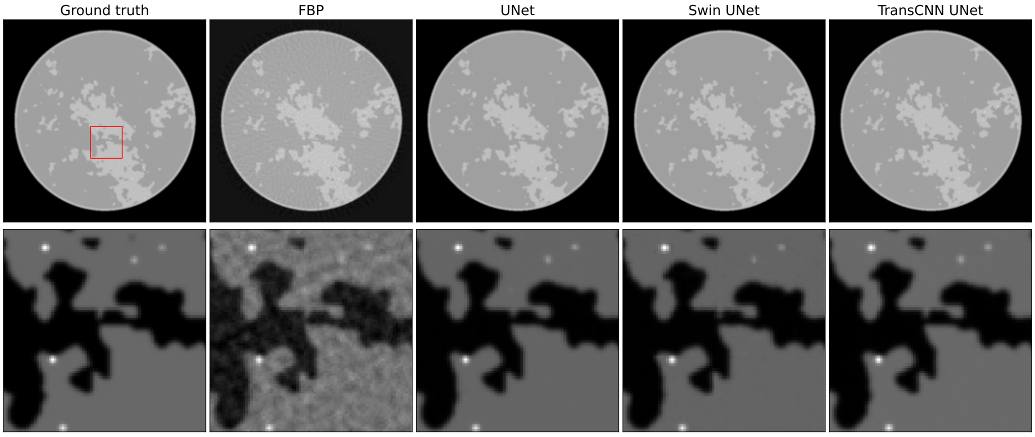

Figure 5: The reconstructed results of simulated breast datasets from 128 views sinogram using FBP, UNet, Swin Unet.

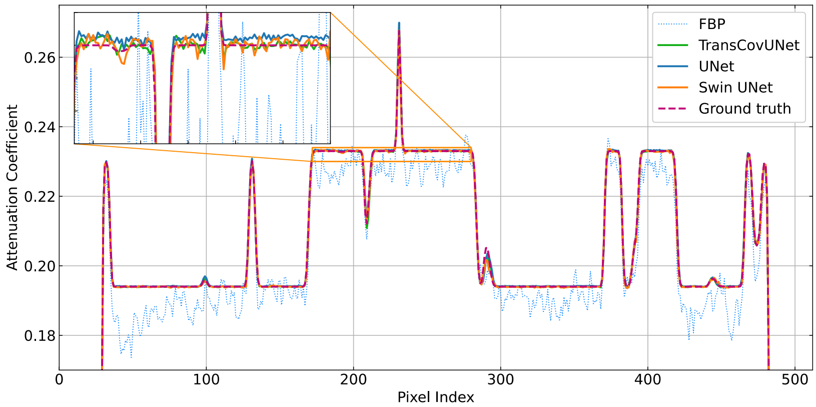

Figure 6: Quantitative line intensity profiles comparison. The line intensity profiles correspond to the central vertical lines in the CT images shown in Fig. 5.

Figures 5-6 show the comparison of the experimental results of the models on the test data. We can see that for CT image reconstruction with sparse view, the reconstructed slice images obtained by the FBP algorithm show severe striping artifacts. Neural network methods such as TransCovUNet, UNet, and Swin-UNet effectively remove these artifacts and remain enough image details to give visually indistinguishable results. The enlarged image shows that TransCovUNet has more complete reconstruction details and clearer image edges than the other two advanced neural network methods. Correspondingly, in the quantitative line intensity profiles, the reconstruction result of TransCovUNet is closer to the ground truth distribution, which is better than that of UNet and Swin-UNet, and the reconstruction result of FBP shows strong volatility and discrepancy. Overall, TransCovUNet's reconstructed images have higher accuracy in visual effects, effectively suppress stripe artifacts, while remain more image details and closer to groundtruth images.

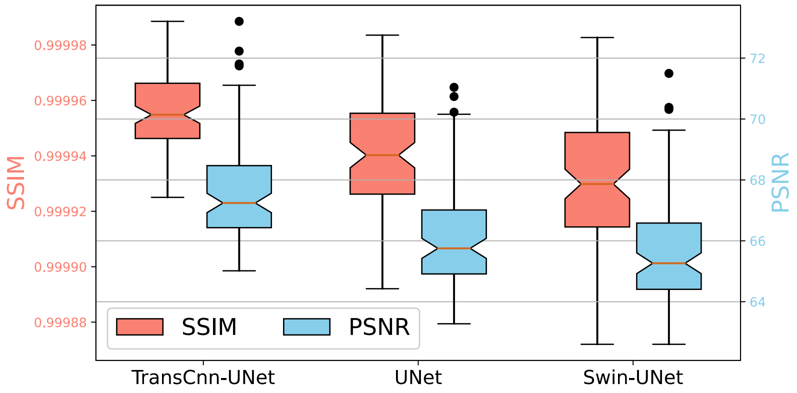

Figure 7: The SSIM and PSNR of UNet, Swin UNet, and TransCNN UNet method.

This study uses the reconstruction results of 100 sets of test images to calculate the metric of image quality. Figure 7 shows the box plots of the SSIM and PSNR values of the reconstruction results for each method. The box plot show that, compared with the classical UNet and Swin-Unet, the mean of SSIM and PSNR for TransCovUNet's result are both relatively high. Moreover, from the distribution shown by the box plot, the SSIM and PSNR value distribution of the TransCovUNet reconstruction results are more concentrated, which means the robustness is better. Table 1 show the average RMSE and the worst local RMSE of the reconstruction results of each model. We can see, compared to the FBP algorithm, UNet, and Swin-UNet, the average RMSE and worst local RMSE of the reconstruction results of TransCovUNet are lower.

Table 1: The mean of RMSE and the Worst-case 25x25 pixel ROI-RMSE of UNet, Swin UNet, and TransCNN UNet method.

Method	RMSN	WC_ROI-RMSN
FBP	0.0057	0.0105
UNet	0.00051	0.0026
Swin-UNet	0.00054	0.0025
TransCNNUNet	0.00042	0.0022

In projection domain, to verify the effectiveness of the modified LapSRN, we selected 100 sets of CT images for testing and compared with traditional method Bicubic. The uniformly selected 128 views sinogram is used as the input of the models, and processed by the modified LapSRN and Bicubic, respectively. The left plot of Fig. 8 show the comparison of the experimental results of the models on the test data. We can see that modified LapSRN can efficiently estimate the missing projection data, and show very small discrepancy with the ground truth. As comparation, the Bicubic couldn't achieve the same perfermance. In addition, the metrics result from SSIM, PSNR, RMSE and WC_ROI-RMSE as show in left plot of Fig. 9 and Table 2 also confirmed that the modified LapSRN improve the interpolation performance compared with Bicubic. Then the 128 views sinogram, and 1024 views sinogram estimated by modified LapSRN and Bicubic were reconstructed by FBP, as shown in right plot of Fig. 8. We can see that Bicubic not only couldn't remove the stripe artifacts effectively, instead of bring in additional artifacts. In contrast, the modified LapSRN can remove the stripe artifacts effectively. In addition, the metrics result from SSIM, PSNR, RMSE and WC_ROI-RMSE as show in right plot of Fig. 9 and Table 3 also confirmed that the modified LapSRN show improved results.

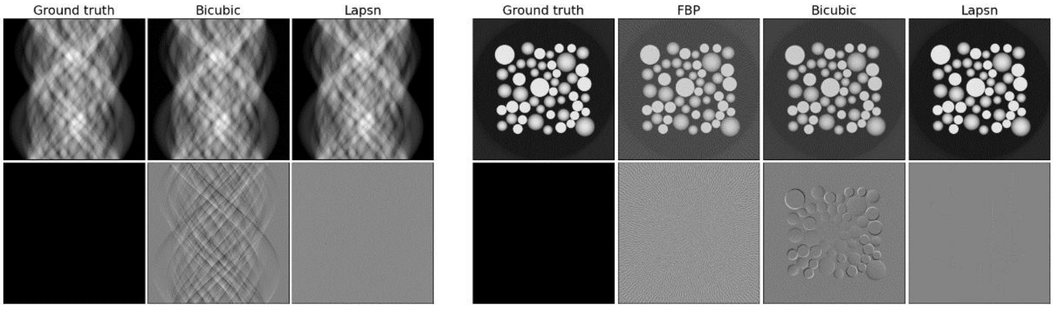

Figure 8: The interpolation results of simulated foam datasets from 128 view sinogram using Bicubic, modified LapSRN(left), and the reconstruction results of 128 views sinogram, 1024 views sinogram estimated by Bicubic and Lapsn(right).

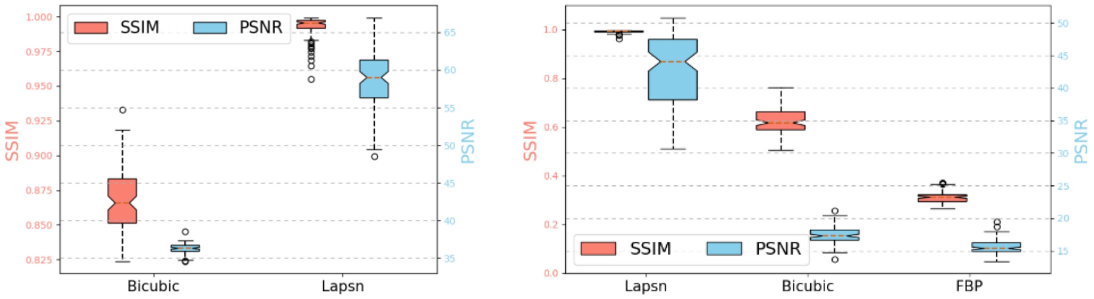

Figure 9: The SSIM and PSNR for interpolation result of Bicubic and modified LapSRN(left), and reconstruction of 128 views sinogram, 1024 views sinogram estimated by Bicubic and Lapsn(right).

Table 2: The mean of RMSE and the Worst-case 25x25 pixel ROI-RMSE for interpolation result of Bicubic and modified LapSRN.

Method	RMSN	WC_ROI-RMSN
Bicubic	8.5	43.3
modified LapSRN	0.2	3.3

Table 3: The mean of RMSE and the Worst-case 25x25 pixel ROI-RMSE for reconstruction of 128 views sinogram, 1024 views sinogram estimated by Bicubic and modified LapSRN.

Method	RMSN	WC_ROI-RMSN
FBP	0.099	0.163
Bicubic	0.085	0.427
modified LapSRN	0.004	0.028

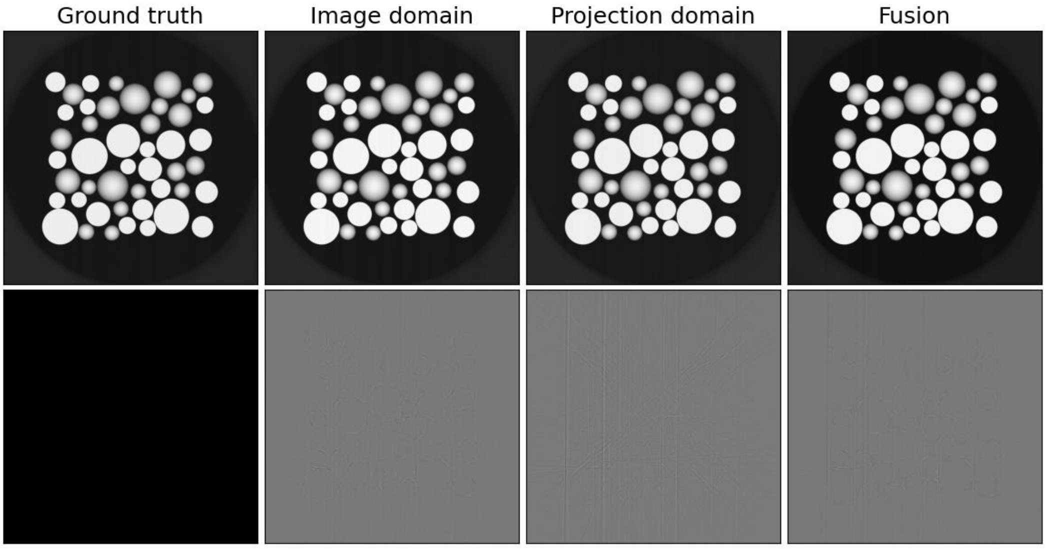

Figure 10: The reconstructed results of simulated foam datasets from 128 views sinogram in image domain, projection domain, and the fusion result.

Finaly, the reconstructions in image and projection domain were combined by the DenseFuse model. The fussion result is shown in Fig. 10. The fused image show no visiual difference with the reconstructions in image and projection domain. The average values of the quality metrics for 100 images which are obtained by image domain, projection domain and the fusion method are shown in Fig. 11 and Table 4. The fused method has the median in the quality metrics, which means that the fusion method can keep the structural information and features without bing in additional artifacts, and improve the robustness of the prediction.

5. Conclusion

In this paper, a new hybrid domain method based on fusion learning, is proposed to deal with the problem of CT image reconstruction with sparse views. The proposed TransCovUNet combines Transformer's long-range contextual information modeling capabilities and CNN's local structural feature extraction capabilities. Combining TransCovUNet with the traditional analytic algorithm FBP effectively solves the problem of stripe artifacts in the process of FBP sparse reconstruction. The experimental results show that, in terms of subjective evaluation, the reconstructed images obtained by this algorithm retain rich details. Compared with other algorithms, the edges, contours and textures are clearer. The proposed algorithm also achieves better results than other advanced algorithms in objective performance evaluation metrics. The proposed modified LapSRN progressively predicts the projection data of missing view in a coarse-to-fine manner, which reduces the computational complexity and alleviates issues with additional artifacts. The experimental results show that, the proposed modified LapSRN achieves better results than the troditional Bicubic method. In addition, we used the DenseFuse model to combine the reconstructions from image domain and projection domain, which can make the prediction robust.

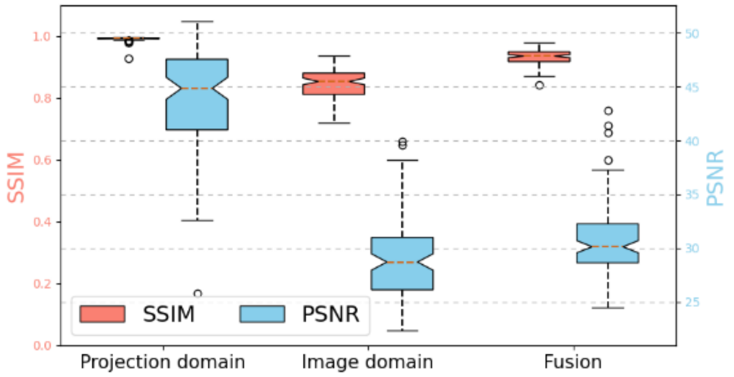

Figure 11: The SSIM and PSNR for the reconstruction results of simulated foam datasets from 128 views sinogram in image domain, projection domain, and the fusion result.

Table 4: The mean of RMSE and the Worst-case 25x25 pixel ROI-RMSE for the reconstruction results of simulated foam datasets from 128 views sinogram in image domain, projection domain, and the fusion result.

Method	RMSN	WC_ROI-RMSN
Image domain	0.0123	0.0577
Projection domian	0.0037	0.0233
Fusion	0.0065	0.0317

References

[1] Jiao Y, Xu G, Cui XH, et al. The HEPS project. J Synchrotron Radiat. 2018 Nov 1;25(Pt 6):1611-1618.

[2] Avinash Kak, Malcolm Slaney. Principles of Computerized Tomographic Imaging [M]. Society for Industrial and Applied Mathematics, Jan. 2001.

[3] Richard Gordon, Robert Bender, Gabor T.Herman.. Algebraic reconstruction techniques (ART) for three-dimensional electron microscopy and x-ray photography [J]. Journal of Theoretical Biology, 1970, 29(3): 471-481.

[4] A. H. Andersen and A. C. Kak. Simultaneous algebraic reconstruction technique (SART): A superior implementation of the ART algorithm [J]. Ultrasonic Imaging, 1984, 6(1): 81-94.

[5] A. P. Dempster, N. Laird, D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm [J]. Journal of the Royal Statistical Society Series B, 1977, 39(1): 1-38.

[6] Peter Gilbert. Iterative methods for the three-dimensional reconstruction of an object from projections[J]. Journal of Theoretical Biology, 1972, 36(1): 105-117.

[7] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton. Imagenet classification with deep convolutional neural networks [J]. Advances in Neural Information Processing Systems, 2012, 60(6): 1097–1105.

[8] Ross Girshick, Jeff Donahue, Trevor Darrell, et al. Rich feature hierarchies for accurate object detection and semantic segmentation [C]. CVPR 2014: 2014 IEEE Conference on Computer Vision and Pattern Recognition, 2014: 580–587.

[9] Jiwon Kim, Jung Kwon Lee, Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks [J]. CVPR 2016: 2016 IEEE Conference on Computer Vision and Pattern Recognition, 2016: 1646–1654.

[10] Kai Zhang, Wangmeng Zuo, Yunjin Chen, et al. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising[C]. IEEE Transactions on Image Processing, 2017, 26(7): 3142–3155.

[11] Tobias Würfl, Mathis Hoffmann, Vincent Christlein, et al. Deep Learning Computed Tomography: Learning Projection-Domain Weights From Image Domain in Limited Angle Problems [J]. IEEE Transactions on Medical Imaging, 2018, 37(6): 1454-1463.

[12] Kyong Hwan Jin, Michael T. McCann, Emmanuel Froustey, et al. Deep convolutional neural network for inverse problems in imaging [C]. IEEE Transactions on Image Processing, 2017, 26(9): 4509–4522.

[13] Olaf Ronneberger, Philipp Fischer, Thomas Brox. U-net: Convolutional networks for biomedical image segmentation [C]. MICCAI 2015: Medical Image Computing and Computer-Assisted Intervention. Springer, 2015: 234–241.

[14] Zhou Zongwei, Md Mahfuzur Rahman Siddiquee, Nima Tajbakhsh, et al. Unet++: A nested u-net architecture for medical image segmentation [C]. DLMIA 2018, ML-CDS 2018: Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision. Springer, 2018: 3–11.

[15] Xiao Xiao, Shen Lian, Zhiming Luo, et al. Weighted res-unet for high-quality retina vessel segmentation [C]. ITME: 9th international conference on information technology in medicine and education. IEEE, 2018: 327–331.

[16] Ozan Oktay, Jo Schlemper, Loic Le Folgoc, et al. Attention u-net: Learning where to look for the pancreas [OL]. MIDL 2018: Medical Imaging with Deep Learning, 2018.

[17] Ling Li, Yu Hu. Deep learning based low-dose synchrotron radiation CT reconstruction [OL]. EPJ Web of Conferences 2021, 251:03058.

[18] Daniël M Pelt, James A Sethian. A mixed-scale dense convolutional neural network for image analysis [J]. Proceedings of the National Academy of Sciences 2018, 115(2): 254-259.

[19] Ashish Vaswani, Noam Shazeer, Niki Parmar, et al. Attention is all you need [OL]. NIPS 2017: 31st Conference on Neural Information Processing Systems, 2017: 6000-6010.

[20] Ze Liu, Yutong Lin, Yue Cao, et al. Swin transformer: Hierarchical vision transformer using shifted windows [OL]. arXiv preprint arXiv:2103.14030, 2021.

[21] Hu Cao, Yueyue Wang, Joy Chen, et. al. Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation [OL]. arXiv preprint arXiv:2105.05537, 2021.

[22] Bellens, Simon et al. “Deep learning based sinogram interpolation applied to X-ray CT measurements of polymer additive manufacturing parts.” (2022).

[23] Dong, Xu et al. “Sinogram interpolation for sparse-view micro-CT with deep learning neural network.” Medical Imaging (2019).

[24] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang, "Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution", in IEEE Conference on Computer Vision and Pattern Recognition, 2017.

[25] H. Li and X. -J. Wu, "DenseFuse: A Fusion Approach to Infrared and Visible Images," in IEEE Transactions on Image Processing, vol. 28, no. 5, pp. 2614-2623, May 2019.

[26] E. Y. Sidky, I. Lorente, J. G. Brankov and X. Pan, "Do CNNs Solve the CT Inverse Problem?," in IEEE Transactions on Biomedical Engineering, vol. 68, no. 6, pp. 1799-1810, June 2021.

[27] D. Kazantsev et al. 2018, TomoPhantom, a software package to generate 2D-4D analytical phantoms for CT image reconstruction algorithm benchmarks, Software X, Volume 7, January–June 2018, Pages 150–155.