Low-Dose CT Reconstruction Based on Fused State-Space Modelling

Yucong Liu
School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou, Henan, China
Collaborative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou, Henan, China
Zhe Zhao
School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou, Henan, China
Collaborative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou, Henan, China
Dandan Xu
Collaborative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou, Henan, China
School of Computer and Artificial Intelligence, Zhengzhou University, Zhengzhou, Henan, China
Yusong Lin[*] YSLIN@HA.EDU.CN
School of Cyber Science and Engineering, Zhengzhou University, Zhengzhou, Henan, China
Collaborative Innovation Center for Internet Healthcare, Zhengzhou University, Zhengzhou, Henan, China

Editors: Nianyin Zeng, Ram Bilas Pachori and Dongshu Wang

Abstract

Low-dose CT is widely used in medical imaging, but reducing the radiation dose introduces noise that affects image quality. To this end, we propose a low-dose CT reconstruction method based on fused state-space modelling, which uses the FuseSSM module to extract contextual information in the spatial and channel domains, balances short-range and long-range sensitivities, and introduces the Axial Attention mechanism to reduce the computational complexity, while enhancing the remote-dependent modelling and global texture consistency. The experiments validate the model on the Mayo-2016 dataset, which outperforms the comparative methods in PSNR, SSIM and RMSE metrics, showing good potential for clinical applications.

Keywords: Deep Learning, Medical Imaging, Low Dose CT.

1. Introduction

Low-dose CT is widely used in medical imaging, and the radiation dose is usually reduced by lowering the X-ray tube current and voltage or using tube current modulation, but this introduces significant amounts of noise, which affects the quality of the image (Zhang et al., 2024). In recent years, deep learning methods have achieved excellent results in low-dose CT reconstruction tasks. Chen et al. (2017) proposed RED-CNN, which demonstrated excellent performance in the field of low-dose CT reconstruction. Pan et al. (2022) proposed a multi-domain fusion Swin Transformer network, which for the first time verified the feasibility and advantages. The CNN method suppresses CT image noise through local convolution, but is limited by the receptive field, which makes it difficult to capture global noise patterns and leads to blurred edges in the reconstructed image; Transformer has global modelling capability, but is weak in processing local information, which is prone to lead to the loss of local texture, and has a high computational overhead, which is limited in high-resolution medical image processing (Huang et al., 2024).

The distribution of texture information of low-dose CT images in different tissue regions differs significantly. Therefore, models need to pay attention not only to local details but also to global consistency when reconstructing them, otherwise the global morphology of the reconstructed images may be affected. We propose a low-dose CT reconstruction method based on fused state-space modelling to capture spatial and channel context information in CT images and suppress noise without sacrificing local accuracy.

2. Methodology

In order to capture fine-grained remote dependencies in medical images and improve the learning ability of the model for different anatomical structures, we improve RED-CNN (Chen et al., 2017) and propose a low-dose CT reconstruction method based on fused state-space modelling (LC-SSM). LC-SSM balances short-range and long-range sensitivities by extracting contextual information through the FuseSSM block in the spatial domain and the channel domain (Şaban Öztürk et al., 2024). Axial Attention is then introduced to reduce computational complexity while enhancing remote dependency modelling and global texture consistency (Ho et al., 2019). The model adopts a 10-layer residual encoder-decoder architecture (5 convolutional and 5 inverse convolutional layers), where the encoder uses convolutional layers with ReLU to progressively suppress the noise, with no pooling operation to preserve the key structural information, and the decoder reconstructs the CT details progressively by inverse convolution with ReLU, as shown in Figure 1.

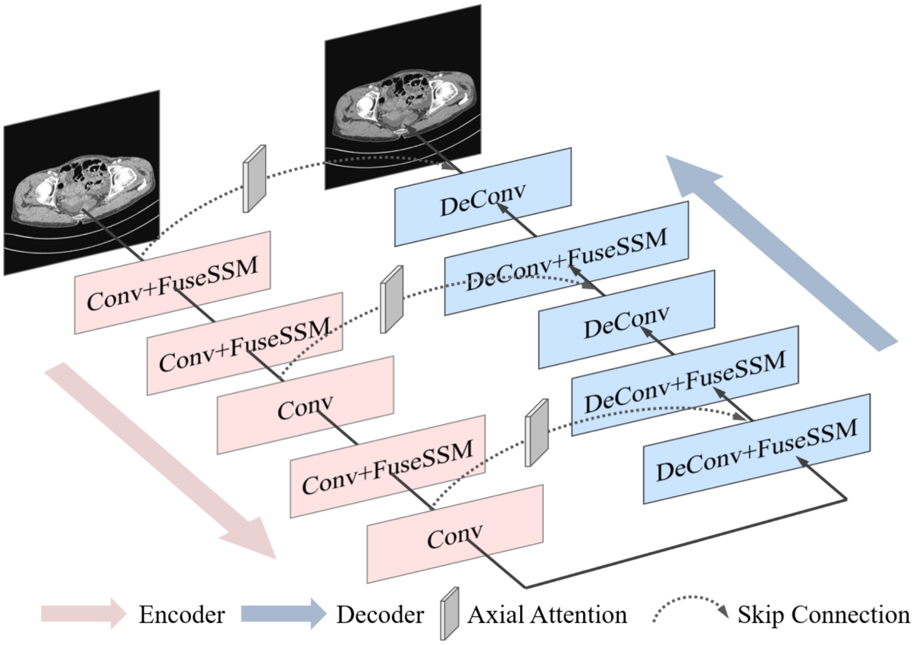
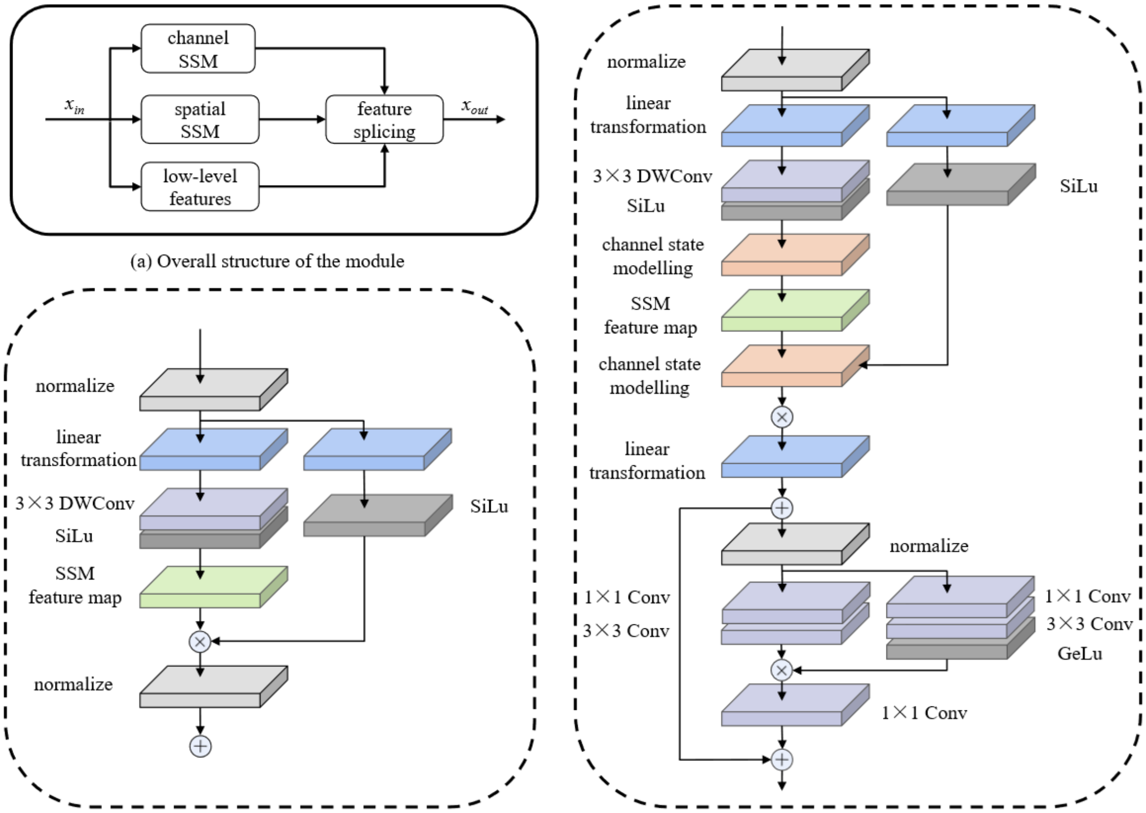


Figure 1: Schematic diagram of LC-SSM model structure.

2.1. Fused state space module

The Fused State-Space Module (FuseSSM) mainly consists of a spatial-SSM module and a channel-SSM module. The spatial-SSM module is used to capture the contextual representation in the spatial domain and is responsible for modelling the spatial dimension with remote dependent information. The channel-SSM is used to capture the contextual representation in the channel domain and is responsible for feature association in the channel dimension. On this basis, the important structural information in the CT image is preserved by directly transferring the low-level features through residual summation, as shown in Figure 2.

Assume that the input feature of the nth stage is $Z_i = x^n \in \mathbb{R}^{H' \times W' \times C'}$, the FuseSSM block first computes the contextual representation by projecting this input feature through three parallel paths. Outputs z_s, z_c and lower level features z_i from the two SSM branches are spliced in the channel dimension to form a richer feature representation, Z_{concat}. Z_{concat} undergoes a series of convolution and feature fusion operations to generate an enhanced feature representation that takes into account both low-level spatial information and multi-scale context.



(a) Overall structure of the module

(b) Channel SSM

(c) Spatial SSM

Figure 2: Schematic diagram of FuseSSM module.

2.2. Axial Attention

Skip Connection is widely used to pass low-level features across layers in the en-coder-decoder architecture. The low-level features extracted by the encoder usually contain background noise and low-frequency components, and passing these unfiltered low-level features directly through the skip connection will lead to the propagation of noise, which will affect the effectiveness of subsequent denoising. To minimise this effect, we embed an axial attention mechanism at the skip connections of the LC-SSM network model to adaptively screen the low-level features before they enter the decoder, avoiding excessive injection of noise components into the decoding process, and at the same time enhancing the texture consistency of the reconstructed image.Axial attention is a special variant of self-attention that performs attention computation on only a single axis of multidimensional data, while keeping the other axes information independent. This computation significantly reduces the computational complexity. Axial attention avoids the computational bottleneck caused by directly computing the global attention matrix by decomposing the attention computation of a high-dimensional tensor into different axes and performing local feature interactions on an axis-by-axis basis. Then, it makes the computational complexity reduce from $O(N^2)$ of traditional self-attention to $O(N^{(d-1)/d})$. Where the tensor dimension is ‘d’ and ‘N’ is the total number of pixels. The structure of axial attention is shown in Figure 3.

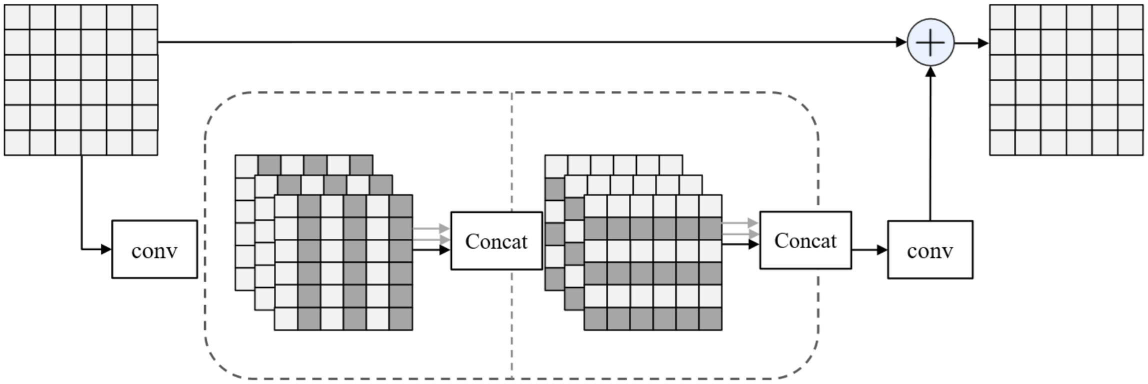


Figure 3: Schematic diagram of axial attention.

2.3. Loss function

The loss function of LC-SSM consists of MSE Loss and SSIM Loss. The overall loss function is as follows,where λ_1 and λ_2 are weight parameters.

$$\mathcal{L}_{\text{total}} = \lambda_1 \mathcal{L}_{\text{MSE}} + \lambda_2 \mathcal{L}_{\text{SSIM}} \tag{1}$$

3. Design of experiments

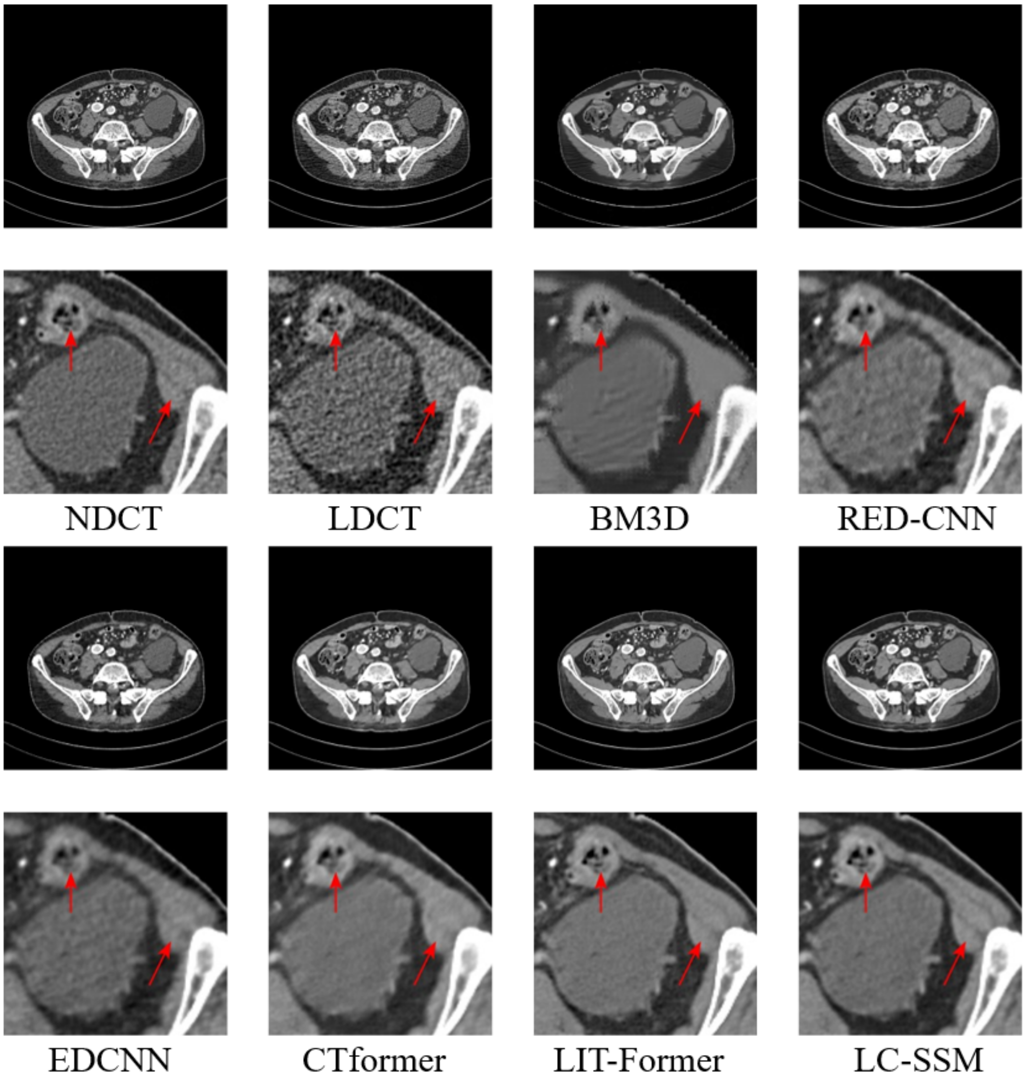
The dataset we used was licensed by Mayo Clinics and was used in the 2016 NIH-AAPM-Mayo Clinic Low Dose CT Grand Challenge. The dataset consists of full-dose images and corresponding quarter-dose images from 10 anonymised patients. We selected 4800 pairs images for training, and 1136 pairs for testing.

Our proposed LC-SSM model is trained using Adam optimiser for a total of 100 Epochs. The Batch Sizeof the training data during the iterations of the model is set to 4, and the learning rate has an initial value of 10-4 and slowly decreases to 10-5. The GPU used for the experiments is NVIDIA Tesla4 (8GB).

4. Experimental results and analysis

4.1. Qualitative evaluation

Figure 4 shows the results of reconstruction of randomly selected CT slices from the test part of the dataset. EDCNN has slightly better noise suppression compared to RED-CNN, and the reconstructed tissue texture is more uniform than the former. The CTformer based only on the Transformer architecture recovers a visually appealing CT image, but loses some local texture information, and the contrast is reduced com-pared to NDCT. The CT image reconstructed by LIT-Former retains both the overall and local texture, and it also performs well in terms of denoising, which may be attributed to the fact that it employs a convolutional and Transformer Combined hybrid architecture that better combines the advantages of both. Our proposed LC-SSM model reconstruction performs the best among all the compared methods.LC-SSM reduces noise while preserving texture details without blurring, and the contrast of the reconstructed image is closer to that of NDCT, especially the structures marked with arrows in Fig. 4


NDCT LDCT BM3D RED-CNN

EDCNN CTformer LIT-Former LC-SSM

Figure 4: Results of different methods on Mayo-2016 dataset.

4.2. Quantitative evaluation

The performance of different models on Mayo-2016 dataset is demonstrated in Table 1. The experimental results on the Mayo-2016 dataset show that our proposed LC-SSM model improves all three image quality metrics compared to BM3D. Compared with RED-CNN, LC-SSM has an average improvement of 0.4936 dB in the PSNR metric and the remaining two metrics, which indicates that our introduced FuseSSM block and axial attention mechanism improves the model's ability to capture the noise in CT, which is beneficial to improve the quality of reconstructed images. LC-SSM, despite its slightly lower SSIM than the CTformer, it has an average improvement of about 0.42 dB in PSNR and an improvement in RMSE.

Table 1: Comparison of the performance of different models on the Mayo-2016 dataset

Method	Evaluation indicators (average)		
	PSNR (dB)	SSIM	RMSE
BM3D (Fumene Feruglio et al., 2010)	39.9938	0.9573	0.0086
RED-CNN (Chen et al., 2017)	43.8739	0.9735	0.0064
EDCNN (Liang et al., 2020)	43.3725	0.9673	0.0070
CTformer (Wang et al., 2023)	43.9471	**0.9742**	0.0069
LIT-Former (Chen et al., 2024)	44.1830	0.9617	0.0067
LC-SSM	44.3675	0.9738	0.0063

5. Conclusion

We propose a low-dose CT reconstruction model based on fused state space modelling. The results of the three metrics of PSNR, SSIM and RMSE of the model in the publicly available dataset were 44.3675, 0.9738 and 0.0063, respectively, which proved the clinical application value of the model.

Acknowledgments

This paper is supported by the National Natural Science Foundation of China under Grant No. 82441022, and Collaborative Innovation Major Project of Zhengzhou under Grant No.20XTZX05015.

References

Hu Chen, Yi Zhang, Mannudeep K. Kalra, Feng Lin, Yang Chen, Peixi Liao, Jiliu Zhou, and Ge Wang. Low-dose ct with a residual encoder-decoder convolutional neural network. *IEEE Transactions on Medical Imaging*, 36(12):2524–2535, 2017. doi: 10.1109/TMI.2017.2715284.

Z. Chen, C. Niu, Q. Gao, G. Wang, and H. Shan. Lit-former: Linking in-plane and through-plane transformers for simultaneous ct image denoising and deblurring. *IEEE Trans Med Imaging*, 43 (5):1880–1894, 2024. doi: 10.1109/tmi.2024.3351723.

P. Fumene Feruglio, C. Vinegoni, J. Gros, A. Sbarbati, and R. Weissleder. Block matching 3d random noise filtering for absorption optical projection tomography. *Phys Med Biol*, 55(18): 5401–15, 2010. doi: 10.1088/0031-9155/55/18/009.

Jonathan Ho, Nal Kalchbrenner, Dirk Weissenborn, and Tim Salimans. Axial attention in multidimensional transformers. *CoRR*, abs/1912.12180, 2019.

Wenli Huang, Ye Deng, Siqi Hui, Yang Wu, Sanping Zhou, and Jinjun Wang. Sparse self-attention transformer for image inpainting. *Pattern Recogn.*, 145(C), January 2024. doi: 10.1016/j.patcog.2023.109897.

Tengfei Liang, Yi Jin, Yidong Li, Tao Wang, Songhe Feng, and Congyan Lang. Edcnn: Edge enhancement-based densely connected network with compound loss for low-dose ct denoising. *2020 15th IEEE International Conference on Signal Processing (ICSP)*, 1:193–198, 2020.

J. Pan, H. Zhang, W. Wu, Z. Gao, and W. Wu. Multi-domain integrative swin transformer network for sparse-view tomographic reconstruction. *Patterns (N Y)*, 3(6):100498, 2022. doi: 10.1016/j.patter.2022.100498.

D. Wang, F. Fan, Z. Wu, R. Liu, F. Wang, and H. Yu. Ctformer: convolution-free token2token dilated vision transformer for low-dose ct denoising. *Phys Med Biol*, 68(6), 2023. doi: 10.1088/1361-6560/acc000.

Ju Zhang, Weiwei Gong, Lieli Ye, Fanghong Wang, Zhibo Shangguan, and Yun Cheng. A review of deep learning methods for denoising of medical low-dose ct images. *Comput. Biol. Med.*, 171 (C), March 2024. doi: 10.1016/j.compbiomed.2024.108112.

Şaban Öztürk, Oguz Can Duran, and Tolga Çukur. Denomamba: A fused state-space model for low-dose ct denoising. *ArXiv*, abs/2409.13094, 2024.