CT Image Segmentation Using Frequency Domain Fea-ture-Assisted Selective Long Memory State Space Model

Xueren Zhang[1] Xianghong Wang[3] Tianye Niu[1,2]*
1 School of Science, Sun Yat-sen University, Shenzhen, China
2 School of Information Science and Technology, University of Science and Technology of Chi-na, China
3 Shenzhen Bay Laboratory, China

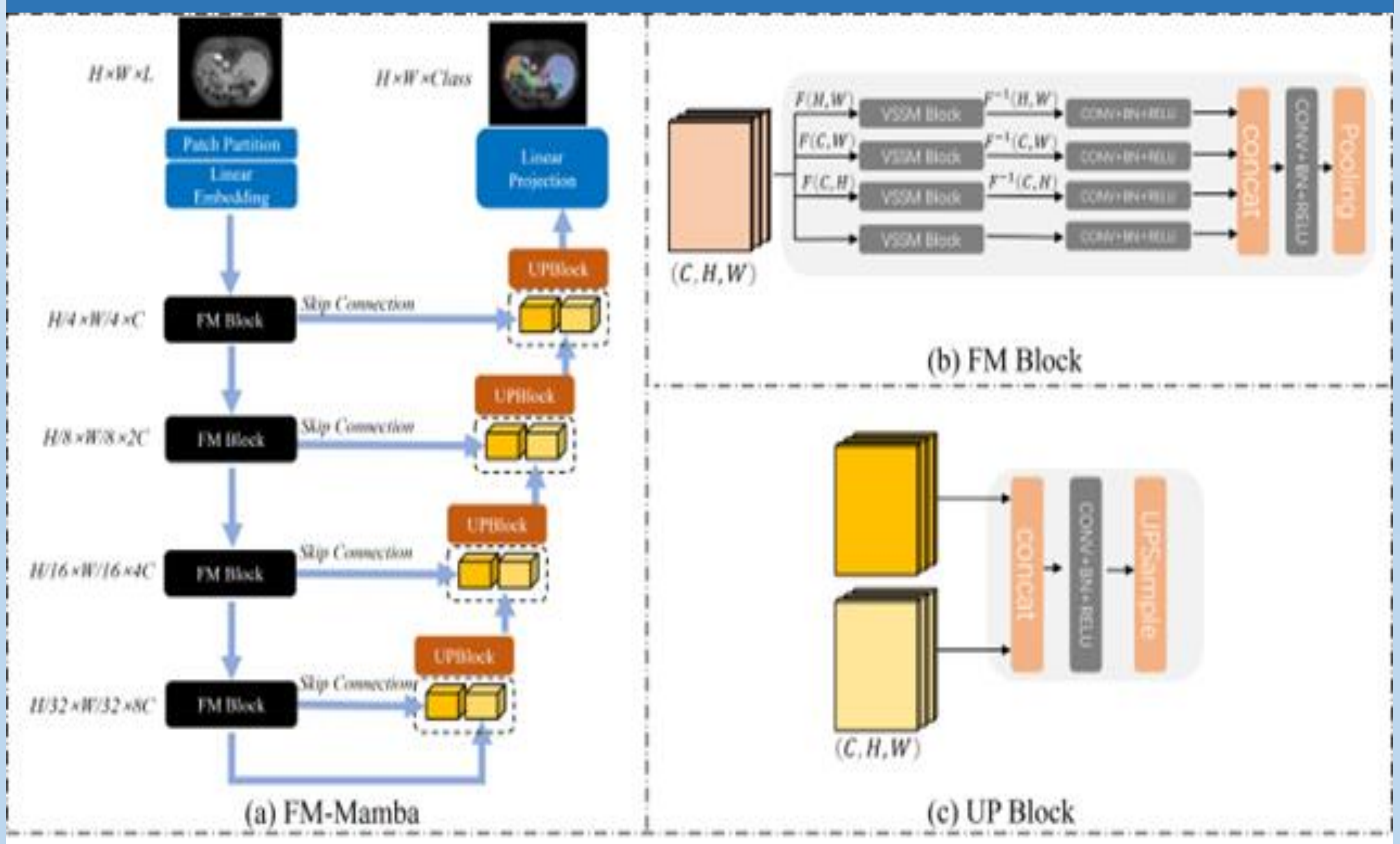
Abstract

Convolutional Neural Networks (CNNs) and Transformers are the most popular architectures for medical image segmentation. However, both have limitations in handling long-range dependencies due to their inherent locality and computational complexity, respectively. To address this challenge, we introduce FM-Mamba, a general-purpose network for CT image segmentation. Inspired by the State Space Sequence Models (SSMs), a new family of deep sequence models known for their strong capability in handling long sequences with fewer parameters, we de-sign a hybrid FM-SSM block that integrates the local frequency domain feature extraction with the abilities of SSMs for capturing the long-range dependency. This module ensures that the model captures global features with fewer parame-ters while reducing the difficulty of the model in identifying segmentation bound-aries. In addition, FM-Mamba employs an expansion technique for SSM's A-matrix, which enables it to guarantee excellent inference capabilities even when faced with large size images. We conducted experiments on the publicly available CT-ORG and Synapse multi-organ segmentation datasets. The results reveal that FM-Mamba outperforms state-of-the-art CNN-based and Transformer-based segmentation networks. This advancement opens new avenues for efficient long-range dependency modeling in medical image analysis.

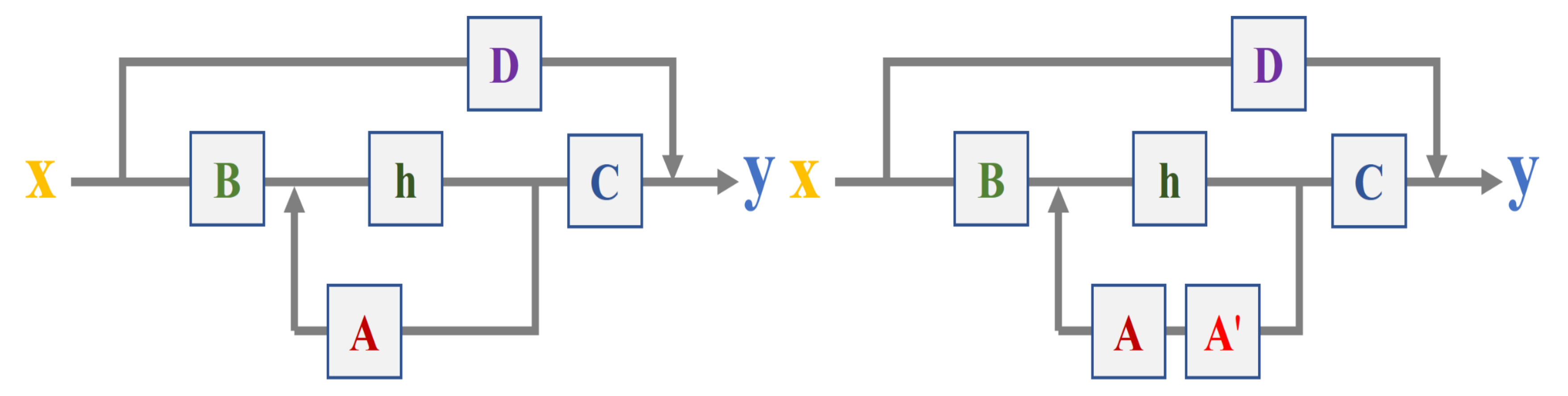
Introduction

CT image segmentation is crucial in medical image analysis. This paper proposes FM-Mamba, a Mamba-inspired model, to explore its feasibility for CT segmentation. The model uses a U-shaped encoder-decoder structure to retain features. To handle large-resolution CT images, we enhance S6's historical state memory via an aug-mented A-matrix, creating the selective long memory state space model (S6M). The encoder includes a frequency-domain feature enhancement module with S6M, lever-aging frequency transformations to improve boundary segmentation. The decoder uses CNNs and upsampling layers. Validated on the CT-ORG and Synapse datasets for multi-organ segmentation, FM-Mamba outperforms state-of-the-art CNN and Transformer-based networks, enabling efficient long-range dependency modeling in medical image analysis.

Methods and Materials



(a) FM-Mamba (b) FM Block (c) UP Block



(a) SSM for S6 (b) SSM for S6M

The state space model (SSM) functioning on S6 is simply formulated as:

$$h'(t) = \boldsymbol{A}h(t) + \boldsymbol{B}x(t) \quad (1)$$

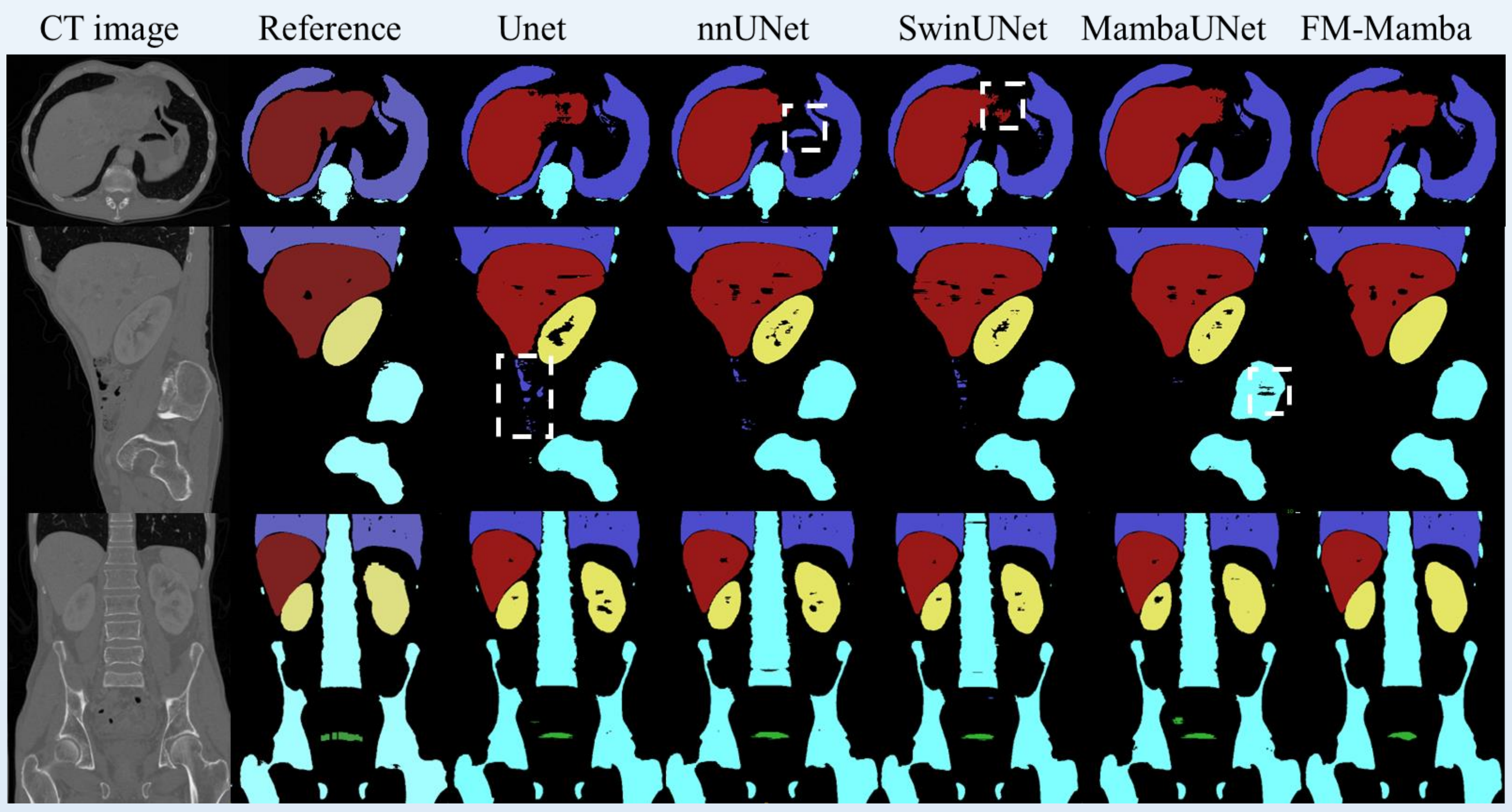
$$y(t) = \boldsymbol{C}h(t) \quad (2)$$

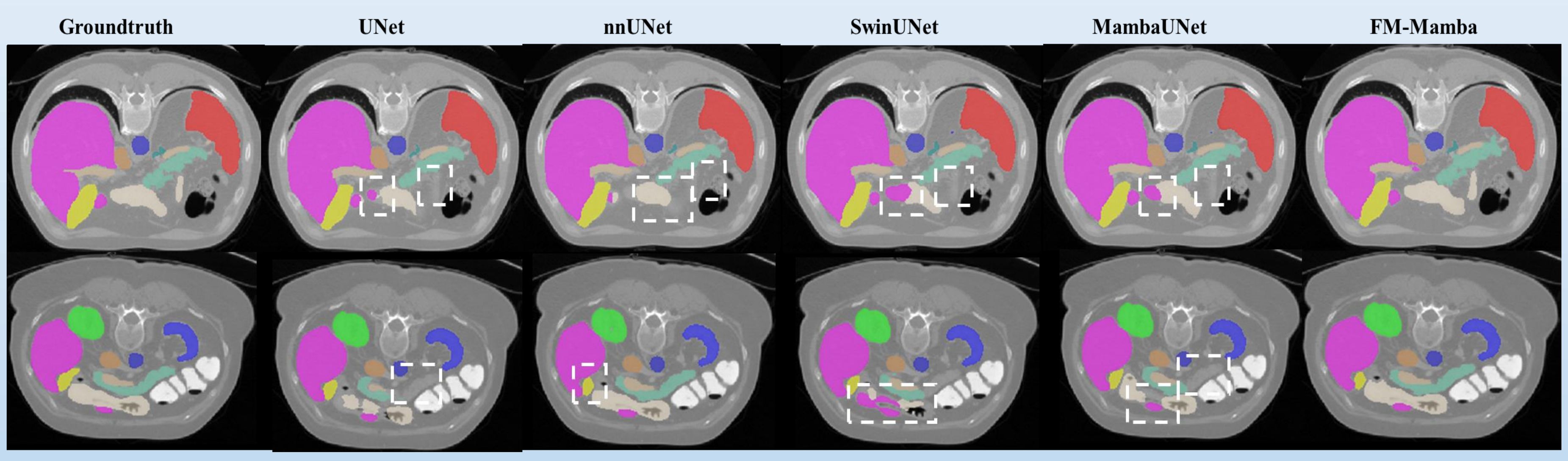
The model is able to map a one-dimensional sequence $x(t) \in \mathbb{R}^L$ to $y(t) \in \mathbb{R}^L$ by means of an intermediate potential state $h(t) \in \mathbb{R}^N$, where the system matrices $\boldsymbol{A} \in \mathbb{R}^{N\times N}$, $\boldsymbol{B} \in \mathbb{R}^{N\times 1}$ and $\boldsymbol{C} \in \mathbb{R}^{N\times 1}$ control the dynamic mapping and output mapping, respectively. We multiply the $\boldsymbol{A}'$-matrix of equal size with the $\boldsymbol{A}$-matrix, then Eq. (1) becomes:

$$h'(t) = \boldsymbol{A}\boldsymbol{A}'h(t) + \boldsymbol{B}x(t) \quad (3)$$

The form of multiplication of the two parameter matrices enables the A matrix to be augmented while also fitting more ways of feature compression.

Results





Conclusions

In this paper, we introduce FM-Mamba, a U-shaped CT image segmentation network that combines the selective long memory state space model (S6M) with frequency domain features. Experimental results show that FM-Mamba relies on S6M and fre-quency domain features to demonstrate its unique advantages in identifying the organ to which each pixel belongs. In the future, we aim to extend FM-Mamba to semi/weakly-supervised learning, as well as other medical image processing domains, in order to further promote the development of medical imaging.

References

1. Ronneberger, O., Fischer, P., & Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing(2015).
2. Xing, Z., Ye, T., Yang, Y., Liu, G., Zhu, L.: SegMamba: Long-Range Sequential Modeling Mamba for 3D Medical Image Segmentation. In: Linguraru, M.G., et al. Medical Image Computing and Computer Assisted Intervention – MICCAI 2024(2024).
3. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023).
4. Ruan, J., Xie, M., Xiang, S., Liu, T., & Fu, Y.: MEW-UNet: Multi-axis representation learning in frequency domain for medical image segmentation. arXiv preprint arXiv:2210.14007(2022).
5. Xiao, Y., Yuan, Q., Jiang, K., Chen, Y., Zhang, Q., & Lin, C. W.: Frequency-assisted mamba for remote sensing image super-resolution. IEEE Transactions on Multimedia(2024).