# VesMamba: 3D Pulmonary Vessel Segmentation from CT images via Mamba with Structural Perception and Scale-aware Filtering

Supplementary Material

## 1. Efficiency Analysis

The limited hardware environment restricts the application of models in clinical practice. Therefore, we further compared the FLOPs (T), inference memory (M), and inference speed (iterations/s) of VesMamba with advanced methods. As shown in Table 1, LKM-UNet exhibits the highest computational complexity, with a value of 10.75 TFLOPs. This is attributed to the employment of pixel-level SSM and patch-level SSM in LKM-UNet, and a bidirectional scanning strategy, which collectively leads to a significant surge in computational overhead. 3D-UNet presents the highest inference memory (IM) requirements, reaching 11,132 MiB, due to its extensive use of convolutional operations and the retention of numerous intermediate features. DSCNet shows the slowest inference speed (IS) at only 1.13 iterations /s. This is attributed to its integration of dynamic snake convolution and a multi-view feature fusion strategy, which are complex operations with low resource utilization. Specifically, dynamic snake convolution requires the adaptive adjustment of kernel shapes based on the characteristics of tubular structures, thereby increasing computational complexity. Meanwhile, multi-view feature fusion involves the integration of more feature information, which imposes heavier burdens and idle waits, resulting in slow inference speed. Finally, Segformer3D exhibits minimal computational cost, inference memory, and inference speed. By leveraging a hierarchical Transformer structure combined with a lightweight MLP decoder, SegFormer3D significantly reduces computational overhead. However, such a simple decoder results in a significant loss of performance. Our method is designed with a focus on lightweight model requirements. Comparisons with other state-of-the-art methods demonstrate that our method achieves optimal performance while maintaining acceptable efficiency.

Table 1. Comparison of VesMamba with state-of-the-art methods in terms of inference memory (IM) and inference speed (IS) on the Parse22 dataset.

| Method | Dice ↑ | FLOPs (T) ↓ | IM (M) ↓ | IS (iter/s) ↑ |
|---|---|---|---|---|
| 3D-UNet [1] | 85.25 | 2.22 | 11132 | 6.15 |
| nnUNet [3] | 84.20 | 4.25 | 3705 | 32.11 |
| DSCNet [9] | 76.63 | 2.22 | 7862 | 1.13 |
| SegFormer3D[8] | 77.69 | **0.01** | **3508** | **128.8** |
| UNETR++ [10] | 85.27 | 0.25 | 5074 | 9.71 |
| Swin-UNETR [12] | 79.96 | 0.91 | 6869 | 8.82 |
| UMamba [7] | 85.94 | 2.69 | 4738 | 14.09 |
| SegMamba [15] | 85.23 | 1.81 | 3973 | 8.88 |
| LKM-UNet [13] | 86.21 | 10.75 | 10026 | 4.66 |
| Ours | **86.65** | 2.85 | 7538 | 9.35 |

Table 2. Ablation studies for the number of network layers of VesMamba on the Parse22 dataset.

| Method | Dice↑ | clDice↑ | HD95↓ | NSD↑ |
|---|---|---|---|---|
| 3 layers | 84.21 | 85.11 | 6.65 | 89.91 |
| 4 layers | 85.10 | 86.03 | 4.36 | 91.98 |
| 5 layers | 85.88 | 86.60 | 3.72 | 92.69 |
| 6 layers | 85.94 | **86.64** | 3.69 | 92.74 |
| 7 layers | **85.96** | 86.60 | **3.55** | **92.83** |

Table 3. Ablation studies for the scanning strategy in SSP on the Parse22 dataset.

| Method | Dice↑ | clDice↑ | HD95↓ | NSD↑ |
|---|---|---|---|---|
| Z-shape | 86.27 | 87.19 | 3.43 | 92.81 |
| S-shape | 86.25 | 87.16 | 3.50 | 92.67 |
| Diagonal | 86.28 | 87.15 | 3.54 | 92.78 |
| Bi-Z-shape | **86.42** | **87.38** | **3.25** | **92.93** |
| Bi-S-shape | 86.35 | 87.23 | 3.30 | 92.88 |
| Bi-diagonal | 86.38 | 87.35 | 3.33 | 92.89 |

## 2. Addtional Albation Studies

To validate the effectiveness of our network architecture and module design, we further perform the following ablation studies on the Parse22 dataset. The ablation studies focus on the number of network layers and the Mamba scanning strategy in the SSP module.

**Number of Network Layers:** The default network architecture of UMamba consists of six layers. To further lightweight the model while maintaining performance, we perform ablation studies on the number of layers. As shown in Table 2, as the number of layers decreases, most evaluation metrics progressively decrease. Too few layers fail to adequately extract features, leading to significant drops in all metrics. Conversely, excessive layers yield minimal performance gains, while the five-layer architecture achieves

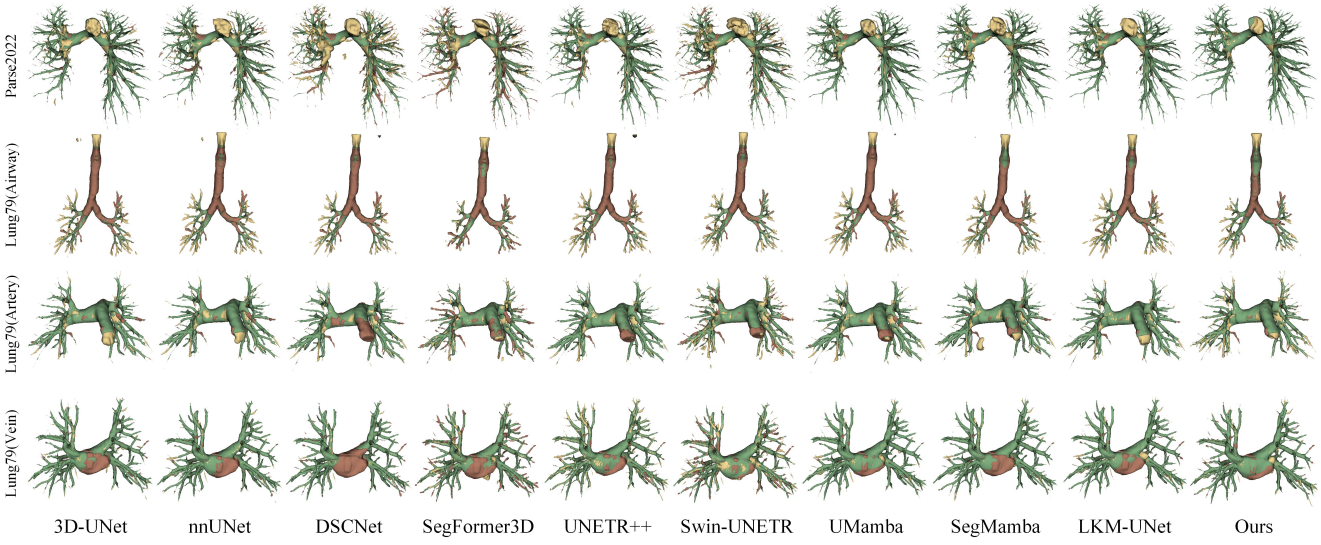


Figure 1. More visual comparison of challenging cases in the Parse2022 and Lung79 datasets with state-of-the-art methods. The green, yellow, and red represent true positives, false positives, and false negatives, respectively.

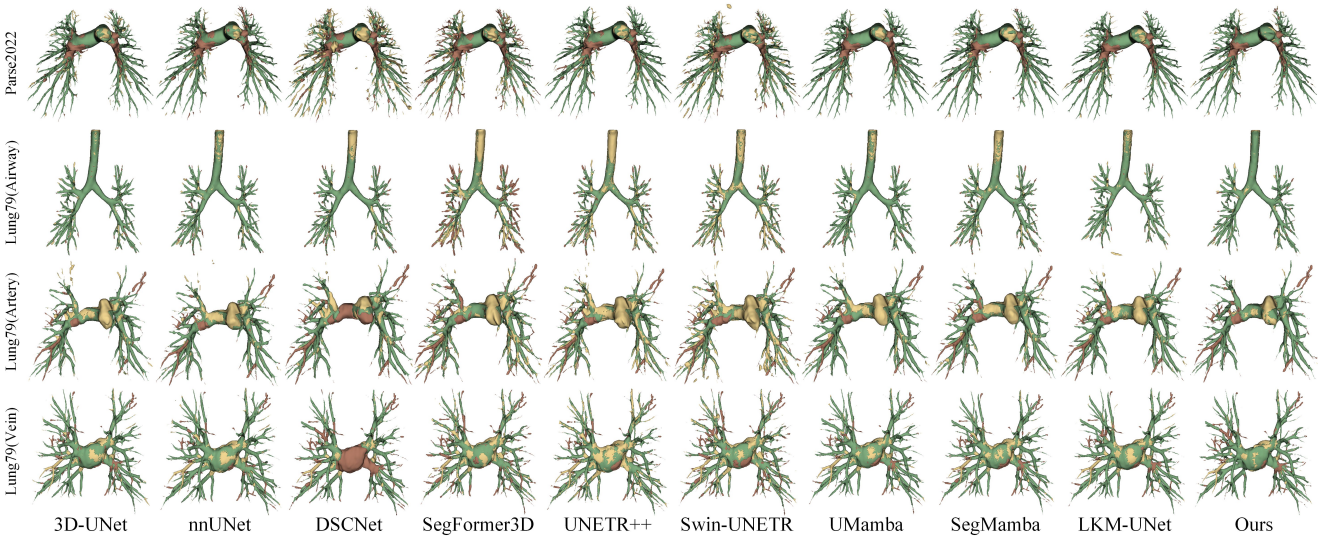


Figure 2. More visual comparison of challenging cases in the Parse2022 and Lung79 datasets with state-of-the-art methods. The green, yellow, and red represent true positives, false positives, and false negatives, respectively.

both lightweight design and guaranteed performance.

**Mamba Scanning Strategy in SSP:** Several studies [4, 5, 11] propose modified Mamba scanning strategies to enhance its suitability for visual tasks. As shown in Table 3, we compare three unidirectional scanning strategies: Z-shape, S-shape, and Diagonal, along with their corresponding bidirectional variants: Bi-Z-shape, Bi-S-shape, and Bi-diagonal, on the Parse22 dataset. The performance differences among the three unidirectional scanning strategies are not significant, as fixed preset scanning strategies cannot guarantee the continuity of all adjacent spatial information. Bidirectional scanning strategies demonstrate improvements over their corresponding unidirectional counterparts, since Mamba is fundamentally a state space model with unidirectional information flow. Bidirectional scanning strategies can simultaneously integrate information from both the front and back of the sequence, thereby alleviating the limitations of unidirectional modeling in Mamba vision tasks. However, while the combined use of more scanning methods can further enhance Mamba's local perception capabilities, it leads to a multiple increase in computational complexity [2, 14, 16]. Among these, the Bi-Z-

shape scanning strategy is relatively superior to the other two strategies and is simpler to implement. Therefore, in SSP, we selected the simple and effective Bi-Z-shape scanning method.

## 3. More Visual Results

As shown in Figure 1 and Figure 2, we provide more visual comparisons with different state-of-the-art methods on the Parse2022 [6] and Lung79 datasets. For these two datasets, we select challenging cases and display the segmentation targets separately. Clearly, on these two datasets, our method is better than other methods at focusing on segmentation targets of different scales, with fewer false positives and false negatives, demonstrating the robustness of our model.

## References


[1] Özgün Çiçek, Ahmed Abdulkadir, Soeren S. Lienkamp, Thomas Brox, and Olaf Ronneberger. 3d u-net: Learning dense volumetric segmentation from sparse annotation. In *Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 424–432, Cham, 2016. Springer International Publishing. 1

[2] Hang Guo, Yong Guo, Yaohua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, and Yawei Li. Mambairv2: Attentive state space restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 28124–28133, 2025. 2

[3] Fabian Isensee, Paul F Jaeger, Simon AA Kohl, Jens Petersen, and Klaus H Maier-Hein. nnu-net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, 2021. 1

[4] Hui Liu, Chen Jia, Fan Shi, Xu Cheng, and Shengyong Chen. Scsegamba: Lightweight structure-aware vision mamba for crack segmentation in structures. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 29406–29416, 2025. 2

[5] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, Jianbin Jiao, and Yunfan Liu. Vmamba: Visual state space model. In *Advances in Neural Information Processing Systems (NeurIPS)*, pages 103031–103063. Curran Associates, Inc., 2024. 2

[6] Gongning Luo, Kuanquan Wang, Jun Liu, Shuo Li, Xinjie Liang, Xiangyu Li, Shaowei Gan, Wei Wang, Suyu Dong, Wenyi Wang, Pengxin Yu, Enyou Liu, Hongrong Wei, Na Wang, Jia Guo, Huiqi Li, Zhao Zhang, Ziwei Zhao, Na Gao, Nan An, Ashkan Pakzad, Bojidar Rangelov, Jiaqi Dou, Song Tian, Zeyu Liu, Yi Wang, Ampatishan Sivalingam, Kumaradevan Punithakumar, Zhaowen Qiu, and Xin Gao. Efficient automatic segmentation for multi-level pulmonary arteries: The parse challenge, 2024. 3

[7] Jun Ma, Feifei Li, and Bo Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation, 2024. 1

[8] Shehan Perera, Pouyan Navard, and Alper Yilmaz. Segformer3d: An efficient transformer for 3d medical image segmentation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, pages 4981–4988, 2024. 1

[9] Yaolei Qi, Yuting He, Xiaoming Qi, Yuan Zhang, and Guanyu Yang. Dynamic snake convolution based on topological geometric constraints for tubular structure segmentation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 6070–6079, 2023. 1

[10] Abdelrahman Shaker, Muhammad Maaz, Hanoona Rasheed, Salman Khan, Ming-Hsuan Yang, and Fahad Shahbaz Khan. Unetr++: Delving into efficient and accurate 3d medical image segmentation. *IEEE Transactions on Medical Imaging (TMI)*, 43(9):3377–3390, 2024. 1

[11] Abdelrahman Shaker, Syed Talal Wasim, Salman Khan, Juergen Gall, and Fahad Shahbaz Khan. Groupmamba: Efficient group-based visual state space model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14912–14922, 2025. 2

[12] Yucheng Tang, Dong Yang, Wenqi Li, Holger R. Roth, Bennett Landman, Daguang Xu, Vishwesh Nath, and Ali Hatamizadeh. Self-supervised pre-training of swin transformers for 3d medical image analysis. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 20730–20740, 2022. 1

[13] Jinhong Wang, Jintai Chen, Danny Chen, and Jian Wu. Lkm-unet: Large kernel vision mamba unet for medical image segmentation. In *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, pages 360–370. Springer, 2024. 1

[14] Chaodong Xiao, Minghan Li, zhengqiang ZHANG, Deyu Meng, and Lei Zhang. Spatial-mamba: Effective visual state space models via structure-aware state fusion. In *International Conference on Representation Learning (ICLR)*, pages 73777–73795, 2025. 2

[15] Zhaohu Xing, Tian Ye, Yijun Yang, Guang Liu, and Lei Zhu. Segmamba: Long-range sequential modeling mamba for 3d medical image segmentation. In *Medical Image Computing and Computer Assisted Intervention (MICCAI)*, pages 578–588, Cham, 2024. Springer Nature Switzerland. 1

[16] Qinfeng Zhu, Yuan Fang, Yuanzhi Cai, Cheng Chen, and Lei Fan. Rethinking scanning strategies with vision mamba in semantic segmentation of remote sensing imagery: An experimental study. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 17:18223–18234, 2024. 2