Supplemental Document for Submission #286: "LKM-UNet: Large Kernel Vision Mamba UNet for Medical Image Segmentation"

Jinhong Wang[1,2,5], Jintai Chen[3(✉)], Danny Chen[4], and Jian Wu[5,6]

[1] College of Computer Science and Technology, Zhejiang University, China
[2] Institute of Wenzhou, Zhejiang University, China
[3] Computer Science Department, University of Illinois Urbana-Champaign, USA
[4] Department of Computer Science and Engineering, University of Notre Dame, USA
[5] State Key Laboratory of Transvascular Implantation Devices of The Second Affiliated Hospital, Zhejiang University School of Medicine, China
[6] School of Public Health Zhejiang University, China
wangjinhong@zju.edu.cn, jtchen721@gmail.com, dchen@nd.edu, wujian2000@zju.edu.cn

1 More Experiments on Abdomen CT Dataset.

Here, we show more kernel size analysis and ablation study results on the Abdomen CT Dataset to further demonstrate the effectiveness of our method and its components.

1.1 Kernel Size Analysis on Abdomen CT Dataset.

As the same as the main article, we explore `LKM-UNet`'s performance in different kernel size settings on on Abdomen CT Dataset, whose stage number is 6. Table 1 shows the results. Comparing the performances of the three kernel-size settings, a similar conclusion can be found that `LKM-UNet` with larger kernel sizes achieves better performances. This indicates that large receptive fields are critical for both 2d and 3d medical image segmentation.

Table 1. Performances of `LKM-UNet` in three different kernel size settings. The kernel size sequence indicates the kernel size in each stage.

Kernel size	(5, 7, 6), (5, 7, 6), (5, 7, 6), (5, 7, 6), (5, 7, 6), (5, 7, 6)]	[(10, 14, 12), (10, 14, 12), (10, 14, 12), (5, 7, 6), (5, 7, 6), (5, 7, 6)]	[(20, 28, 24), (20, 28, 24), (10, 14, 12), (10, 14, 12), (5, 7, 6), (5, 7, 6)]
DSC	86.18	86.45	**86.82**
NSD	89.75	89.89	**90.02**

✉: Corresponding Author.

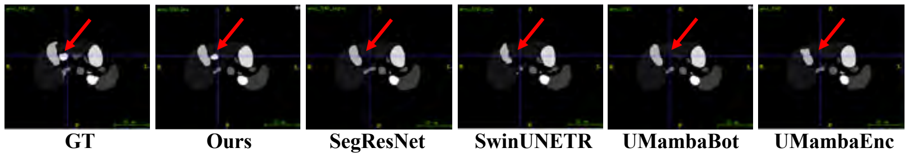

Fig. 1. Qualitative segmentation visualization of previous methods and our LKM-UNet.

Table 2. Performances of `LKM-UNet` with different sub-modules. PiM = Pixel-level SSM. PaM = Patch-level SSM. BiM = Bidirectional Mamba.

Method	Baseline	Only PiM	Only PaM	PiM + BiM	PaM + BiM	PiM + PaM	PiM + PaM + BiM
DSC	86.15	86.54	86.42	86.70	86.60	86.73	**86.82**
NSD	89.72	89.85	89.78	89.99	89.95	90.00	**90.02**

1.2 Ablation Study on Abdomen CT Dataset.

We also conduct ablation experiments on the Abdomen CT dataset to show the importance of each component of `LKM-UNet`. Table 2 shows the results. We can also find that both PiM and PaM provide improvements for `LKM-UNet` over the baseline model, while PiM gains more improvements than PaM, suggesting that enlarging the receptive field of local feature modeling is a key to improving model performance. After introducing BiM, the performance of `LKM-UNet` further improves, which shows the importance of bidirectional Mamba for location-aware sequence modeling. Finally, `LKM-UNet` with all the components also achieves the best performance, which further demonstrates the effectiveness of our method and its components.

2 Qualitative Segmentation Results

To show a more detailed performance of segmentation, we show the qualitative segmentation visualization of previous methods and our LKM-UNet in Fig. 1. It can be seen that our LKM-UNet can recognize the small organ and segment it well which shows LKM-UNet is stronger in not only global modeling but also local details, further demonstrating the effectiveness of LKM-UNet by its large kernel Mamba design.