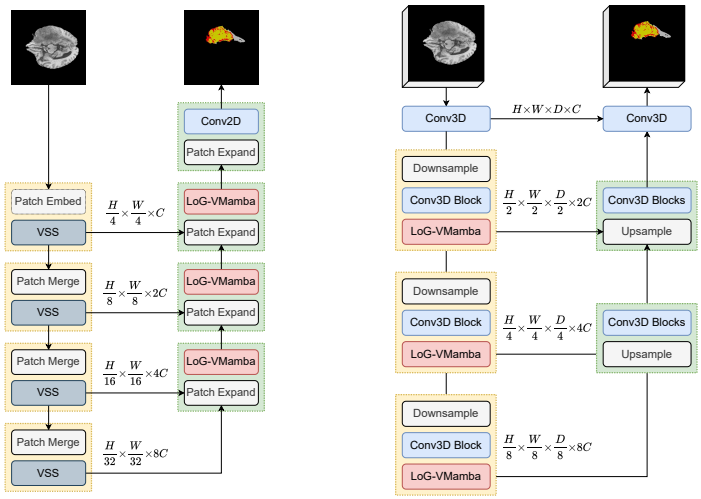
Supplementary Material



(a) 2D segmentation model **(b)** 3D segmentation model

Fig. S1: The overview of our 2D and 3D segmentation models. In 3D models (and U-Mamba baselines), we used max pooling at the start of U-Mamba to massively reduce the number of computations. These illustrations serve as a conceptual representation. Following [34, 38], the number of building blocks differs across datasets. The detailed configurations are shown in Tab. 1.

S1 More Analysis

Impact of Each Component. We hypothesize that G-VMamba is more robust than L-VMamba in handling "long" sequences, where the forgetting of distant tokens and global context in Mamba is more severe. Therefore, the global tokens from GTX are highly crucial for recalling overall information of long sequences. This hypothesis was validated by our experiments in Table 6, where the impact of G-VMamba was more pronounced on BraTS 3D inputs than L-VMamba. One can observe that 3D scans from BraTS produce **8.5 times longer** input sequences compared to the 2D images from EndoVis. Compared to G-VMamba, L-VMamba is expected to be good at learning local information and fine-grained details, which should be crucial for accurate segmentation of numerous classes in EndoVis 2D images. This might explain why L-VMamba showed stronger performance on the Endoscopy dataset.

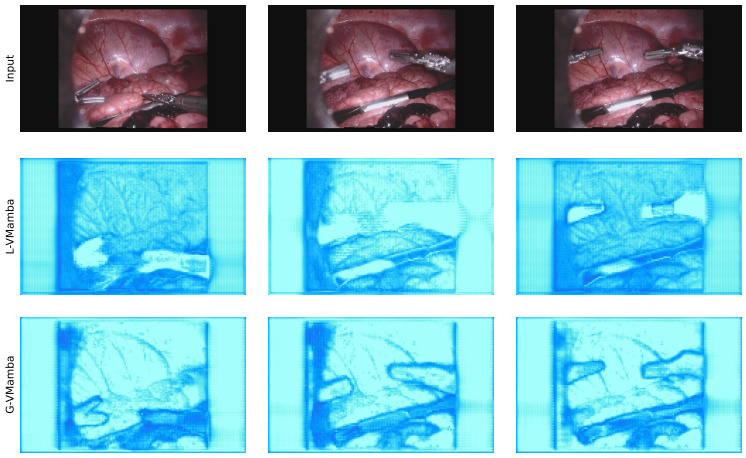


Fig. S2: The activation of features in L-VMamba and G-VMamba when running on the test set of the Endoscopy dataset

Concatenation of Local and Global Tokens. In general, the "Interleaved" strategy showed the best performance. This is intuitively reasonable because it keeps reminding the SSM module of the global context, which reduces the possibility of forgetting information due to the sequential nature of Mamba. Although we applied concatenation strategies to global tokens, similar benefits could be observed when mixing prediction tokens between local tokens, as reported in [50, 59].

S2 Limitations and Broader Impact

Limitations. Firstly, the Mamba block in our LoG-VMamba may result in exploding gradients during training, due to the inherent recurrent nature of SSM. When such issues happen, it is recommended to lower the learning rate. Secondly, as our study focuses on the problem of medical image segmentation, we did not verify the effectiveness of the proposed LoG-VMamba module on natural images or in other vision tasks such as image classification. Lastly, the evaluation of this approach could be broadened by validating on multiple different anatomical regions or comparing to a wider range of deep learning baselines.

Broader Impact. The proposed LoG-VMamba is a generic neural module that can be inserted into different architectures and applied on different vision tasks, e.g., semantic segmentation, image classification, or vision-language multimodal learning. It has no direct negative social impact. The potential malicious uses of LoG-VMamba as a general-purpose neural module are beyond the scope of our study.

Table S1: Performance comparisons on the BraTS test set for each class. The best results are highlighted in bold while the second-best ones are underlined.

Method	Dice score (%) ↑				HD95 (*mm*) ↓			
	ET	TC	WT	Avg	ET	TC	WT	Avg
UNet3D [27]	$83.1_{\pm 0.2}$	$86.1_{\pm 0.3}$	$90.4_{\pm 0.2}$	$86.5_{\pm 0.1}$	$3.8_{\pm 0.4}$	$5.9_{\pm 0.3}$	$6.5_{\pm 0.6}$	$5.4_{\pm 0.4}$
nnUNet [25]	$84.1_{\pm 0.3}$	$\underline{87.2_{\pm 0.5}}$	$91.3_{\pm 0.1}$	$87.5_{\pm 0.3}$	$3.3_{\pm 0.5}$	$5.7_{\pm 0.6}$	$5.5_{\pm 0.3}$	$4.8_{\pm 0.4}$
UNETR [18]	$82.3_{\pm 0.2}$	$82.0_{\pm 0.4}$	$90.0_{\pm 0.1}$	$84.8_{\pm 0.2}$	$4.0_{\pm 0.3}$	$7.4_{\pm 0.2}$	$6.2_{\pm 0.4}$	$5.9_{\pm 0.2}$
Swin-UNETR [17]	$84.1_{\pm 0.2}$	$85.7_{\pm 0.4}$	$90.9_{\pm 0.1}$	$86.9_{\pm 0.2}$	$3.7_{\pm 0.3}$	$6.4_{\pm 0.2}$	$6.0_{\pm 0.2}$	$5.4_{\pm 0.2}$
NestedFormer [53]	$83.5_{\pm 0.1}$	$85.4_{\pm 0.1}$	$91.2_{\pm 0.1}$	$86.7_{\pm 0.1}$	$4.4_{\pm 0.4}$	$7.4_{\pm 0.4}$	$6.6_{\pm 0.4}$	$6.1_{\pm 0.4}$
EoFormer [45]	$82.5_{\pm 0.2}$	$84.8_{\pm 0.4}$	$91.3_{\pm 0.0}$	$86.2_{\pm 0.2}$	$3.7_{\pm 0.2}$	$6.4_{\pm 0.4}$	$5.9_{\pm 0.3}$	$5.3_{\pm 0.2}$
U-Mamba-Bot [38]	$84.1_{\pm 0.2}$	$87.0_{\pm 0.3}$	$\underline{91.5_{\pm 0.1}}$	$87.5_{\pm 0.2}$	$\underline{2.9_{\pm 0.1}}$	$\underline{5.0_{\pm 0.3}}$	$\mathbf{5.0}_{\pm 0.2}$	$\underline{4.3_{\pm 0.2}}$
U-Mamba-Enc [38]	$83.7_{\pm 0.1}$	$86.2_{\pm 0.2}$	$91.2_{\pm 0.1}$	$87.0_{\pm 0.1}$	$3.1_{\pm 0.2}$	$\underline{5.0_{\pm 0.2}}$	$\underline{5.1_{\pm 0.1}}$	$4.4_{\pm 0.1}$
SegMamba [52]	$\mathbf{85.0}_{\pm 0.2}$	$86.7_{\pm 0.3}$	$91.2_{\pm 0.1}$	$\underline{87.6_{\pm 0.2}}$	$3.2_{\pm 0.3}$	$5.8_{\pm 0.4}$	$5.2_{\pm 0.1}$	$4.7_{\pm 0.2}$
Ours	$\underline{84.7_{\pm 0.2}}$	$\mathbf{87.9}_{\pm 0.3}$	$\mathbf{91.6}_{\pm 0.1}$	$\mathbf{88.1}_{\pm 0.1}$	$\mathbf{2.4}_{\pm 0.1}$	$\mathbf{4.5}_{\pm 0.1}$	$\mathbf{5.0}_{\pm 0.2}$	$\mathbf{4.0}_{\pm 0.0}$

Table S2: Performance comparisons on the ACDC test set for each class. The best results are highlighted in bold while the second-best ones are underlined.

Method	Dice score (%) ↑				HD95 (*mm*) ↓			
	RV	MYO	LV	Avg	RV	Myo	LV	Avg
UNet3D [27]	$90.2_{\pm 0.1}$	$89.3_{\pm 0.1}$	$93.3_{\pm 0.1}$	$90.9_{\pm 0.0}$	$\underline{1.3_{\pm 0.0}}$	$\underline{1.1_{\pm 0.0}}$	$\underline{1.2_{\pm 0.0}}$	$\underline{1.2_{\pm 0.0}}$
nnUNet [25]	$91.4_{\pm 0.0}$	$89.9_{\pm 0.0}$	$\mathbf{94.3}_{\pm 0.1}$	$\underline{91.9_{\pm 0.0}}$	$\mathbf{1.2}_{\pm 0.0}$	$\mathbf{1.0}_{\pm 0.0}$	$1.3_{\pm 0.2}$	$\underline{1.2_{\pm 0.1}}$
UNETR [18]	$85.0_{\pm 0.2}$	$84.7_{\pm 0.2}$	$89.9_{\pm 0.2}$	$86.5_{\pm 0.1}$	$2.8_{\pm 0.1}$	$2.1_{\pm 0.1}$	$2.6_{\pm 0.1}$	$2.5_{\pm 0.1}$
Swin-UNETR [17]	$87.9_{\pm 0.2}$	$87.5_{\pm 0.2}$	$92.1_{\pm 0.3}$	$89.2_{\pm 0.2}$	$2.8_{\pm 0.3}$	$1.4_{\pm 0.1}$	$2.3_{\pm 0.2}$	$2.2_{\pm 0.2}$
NestedFormer [53]	$89.2_{\pm 0.1}$	$88.3_{\pm 0.1}$	$92.9_{\pm 0.1}$	$90.1_{\pm 0.1}$	$1.6_{\pm 0.2}$	$1.6_{\pm 0.4}$	$2.5_{\pm 0.6}$	$1.9_{\pm 0.3}$
EoFormer [45]	$89.9_{\pm 0.0}$	$89.8_{\pm 0.0}$	$93.7_{\pm 0.2}$	$91.1_{\pm 0.1}$	$\underline{1.3_{\pm 0.0}}$	$\mathbf{1.0}_{\pm 0.0}$	$\underline{1.2_{\pm 0.1}}$	$\underline{1.2_{\pm 0.0}}$
U-Mamba-Bot [38]	$\underline{91.6_{\pm 0.1}}$	$\underline{90.2_{\pm 0.0}}$	$94.0_{\pm 0.1}$	$\underline{91.9_{\pm 0.1}}$	$\mathbf{1.2}_{\pm 0.0}$	$1.3_{\pm 0.2}$	$1.4_{\pm 0.2}$	$1.3_{\pm 0.1}$
U-Mamba-Enc [38]	$91.1_{\pm 0.4}$	$89.9_{\pm 0.3}$	$93.9_{\pm 0.2}$	$91.6_{\pm 0.3}$	$\mathbf{1.2}_{\pm 0.0}$	$\mathbf{1.0}_{\pm 0.0}$	$\mathbf{1.1}_{\pm 0.0}$	$\mathbf{1.1}_{\pm 0.0}$
SegMamba [52]	$89.6_{\pm 0.2}$	$89.0_{\pm 0.1}$	$93.6_{\pm 0.0}$	$90.7_{\pm 0.1}$	$\underline{1.3_{\pm 0.0}}$	$\underline{1.1_{\pm 0.0}}$	$\underline{1.2_{\pm 0.1}}$	$\underline{1.2_{\pm 0.0}}$
Ours	$\mathbf{92.0}_{\pm 0.1}$	$\mathbf{90.3}_{\pm 0.0}$	$\underline{94.2_{\pm 0.1}}$	$\mathbf{92.2}_{\pm 0.0}$	$\mathbf{1.2}_{\pm 0.0}$	$\mathbf{1.0}_{\pm 0.0}$	$\mathbf{1.1}_{\pm 0.0}$	$\mathbf{1.1}_{\pm 0.0}$