SAM-Mamba: Mamba Guided SAM Architecture for Generalized Zero-Shot Polyp Segmentation
Supplementary Material

Tapas Kumar Dutta[1], Snehashis Majhi[2], Deepak Ranjan Nayak[3], and Debesh Jha[4]
[1] University of Surrey, United Kingdom [2] Côte d'Azur University, France
[3] Malaviya National Institute of Technology Jaipur, India [4] University of South Dakota, USA

1. Additional Quantitative Comparison

Table 1. Quantitative results comparison of SAM-Mamba with SOTA methods on ETIS dataset (unseen).

Methods	mDice ↑	mIoU ↑	F_β^w ↑	S_α ↑	$E_\phi^{\max}$ ↑	MAE ↓
U-Mamba [2]	74.0	66.8	72.3	85.0	87.0	2.2
SAM-H [1]	65.1	60.6	81.2	76.7	76.7	2.0
SAM-L [1]	72.6	67.6	84.9	82.6	82.6	2.0
SAM-Mamba	84.8	78.2	85.5	91.6	93.3	1.0

In order to further assess the generalization capabilities of the model on unseen data, we perform experiments on the ETIS dataset, comparing our proposed SAM-Mamba against state-of-the-art (SOTA) methods such as U-Mamba and SAM. As shown in Table 1, SAM-Mamba consistently demonstrates superior performance across all metrics. Specifically, SAM-Mamba surpasses the next-best method, U=Mamba, by a significant margin of $+11.4\%$ on the mDice metric and $+11.4\%$ on mIoU. Similar improvements are observed for F_β^w, S_α, and $E_\phi^{\max}$, where SAM-Mamba achieves the highest scores of 85.5, 91.6, and 93.3, respectively. In addition, SAM-Mamba achieves the lowest Mean Absolute Error (MAE) with a value of 1.0, further validating its robustness. These results highlight the strength of our method in generalizing to unseen data in zero-shot scenarios, particularly in challenging medical image segmentation tasks.

To further evaluate the zero-shot generalization capability of SAM-Mamba, we conducted experiments on the CVC-300 and CVC-ColonDB datasets. As seen in Table 2, SAM-Mamba consistently outperforms all SOTA methods across both datasets. On the CVC-300 dataset, SAM-Mamba achieves the highest mDice and mIoU scores of 92.0% and 86.1%, respectively, outperforming U-Mamba by a margin of $+19.5\%$ on mDice and $+22.1\%$ on mIoU. Similar trends can be observed for F_β^w, S_α, and $E_\phi^{\max}$, where SAM-Mamba achieves the top scores across all metrics, including the lowest MAE of 0.6.

On the CVC-ColonDB dataset, SAM-Mamba also outperforms the other methods, achieving mDice and mIoU scores of 85.3% and 77.1%, respectively, representing an improvement of $+34.3\%$ and $+33.1\%$ over U-Mamba. SAM-Mamba also demonstrates superior performance in other metrics, including F_β^w, S_α, and $E_\phi^{\max}$, along with the lowest MAE of 1.7. These results further emphasize the robustness and generalization strength of SAM-Mamba in unseen medical image segmentation tasks.

To assess the performance of SAM-Mamba on seen datasets, we compared it against state-of-the-art (SOTA) methods on the Kvasir-SEG and CVC-ClinicDB datasets. As shown in Table 3, SAM-Mamba exhibits superior performance across all metrics on both datasets. On the Kvasir-SEG dataset, SAM-Mamba achieves the highest scores with mDice and mIoU of 92.4% and 87.3%, respectively, surpassing U-Mamba by $+9.9\%$ on mDice and $+12.3\%$ on mIoU. SAM-Mamba also leads in F_β^w, S_α, and $E_\phi^{\max}$, with the lowest MAE of 2.5.

On the CVC-ClinicDB dataset, SAM-Mamba again outperforms other methods, achieving mDice and mIoU scores of 94.2% and 88.7%, which represent improvements of $+11.2\%$ and $+12.7\%$ over U-Mamba. The method also excels in F_β^w, S_α, and $E_\phi^{\max}$, with the lowest MAE of 0.6. These results underscore the effectiveness and robustness of SAM-Mamba in segmentation tasks on seen datasets, demonstrating significant gains over existing SOTA methods.

2. Additional Qualitative Analysis

In this section we extend the notion of qualitative analysis provided in Figure 2 and 3 of main paper and visualize the segmentation results for additional samples of seen (*i.e., Kvasir-SEG and Clinic-DB*) and unseen (*i.e., CVC-300, CVC-ColonDB, and ETIS)*) datasets. In order to derive detailed analytical insights, we choose a diverse set of samples pertaining to various shape, structure and texture of polyp region across seen and unseen dataset. By carefully

Table 2. Quantitative results comparison of SAM-Mamba with SOTA methods on CVC-300 and CVC-ColonDB datasets (unseen). '↑' and '↓ ' represent that larger or smaller scores are better. 'Red' and 'Blue' color fonts indicate the best and second best scores.

Methods	CVC-300 (Unseen)						CVC-ColonDB (Unseen)					
	mDice ↑	mIoU ↑	F_β^w ↑	S_α ↑	$E_\phi^{\max}$ ↑	MAE ↓	mDice ↑	mIoU ↑	F_β^w ↑	S_α ↑	$E_\phi^{\max}$ ↑	MAE ↓
Umamba [2]	72.5	64.0	70.0	85.5	87.5	2.0	51.0	44.0	50.0	71.5	79.0	5.8
SAM-H [1]	65.1	60.6	81.2	76.7	76.7	2.0	44.1	39.6	43.4	67.6	58.7	5.6
SAM-L [1]	72.6	67.6	84.9	82.6	82.6	2.0	46.8	42.2	46.3	69.0	60.8	5.4
SAM-Mamba	92.0	86.1	88.8	94.6	98.1	0.6	85.3	77.1	85.6	89.8	93.3	1.7

Table 3. Quantitative results comparison of SAM-Mamba with SOTA methods on Kvasir-SEG and CVC-ClinicDB datasets (seen). '↑' and '↓ ' represent that larger or smaller scores are better. 'Red' and 'Blue' color fonts indicate the best and second best scores.

Methods	Kvasir-SEG (Seen)						CVC-ClinicDB (Seen)					
	mDice ↑	mIoU ↑	F_β^w ↑	S_α ↑	$E_\phi^{\max}$ ↑	MAE ↓	mDice ↑	mIoU ↑	F_β^w ↑	S_α ↑	$E_\phi^{\max}$ ↑	MAE ↓
U-Mamba [2]	82.5	75.0	80.1	86.0	90.0	5.3	83.0	76.0	82.0	89.2	95.0	1.8
SAM-H [1]	77.8	70.7	76.9	82.9	83.1	6.2	54.7	50.0	54.6	73.8	67.7	4.0
SAM-L [1]	78.2	71.0	77.3	83.2	83.6	6.1	57.9	52.6	56.3	74.4	68.5	3.8
SAM-Mamba	92.4	87.3	94.2	93.6	96.1	2.5	94.2	88.7	94.3	95.5	98.2	0.6

observing Figure 1 to 10, we are convinced that the SAM-Mamba produce best segmentation results for all the scenarios considered which corroborates the learning and generalization ability of our method in seen and unseen dataset.

References

[1] Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C Berg, Wan-Yen Lo, et al. Segment anything. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 4015–4026, 2023. 1, 2

[2] Jun Ma, Feifei Li, and Bo Wang. U-mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024. 1, 2

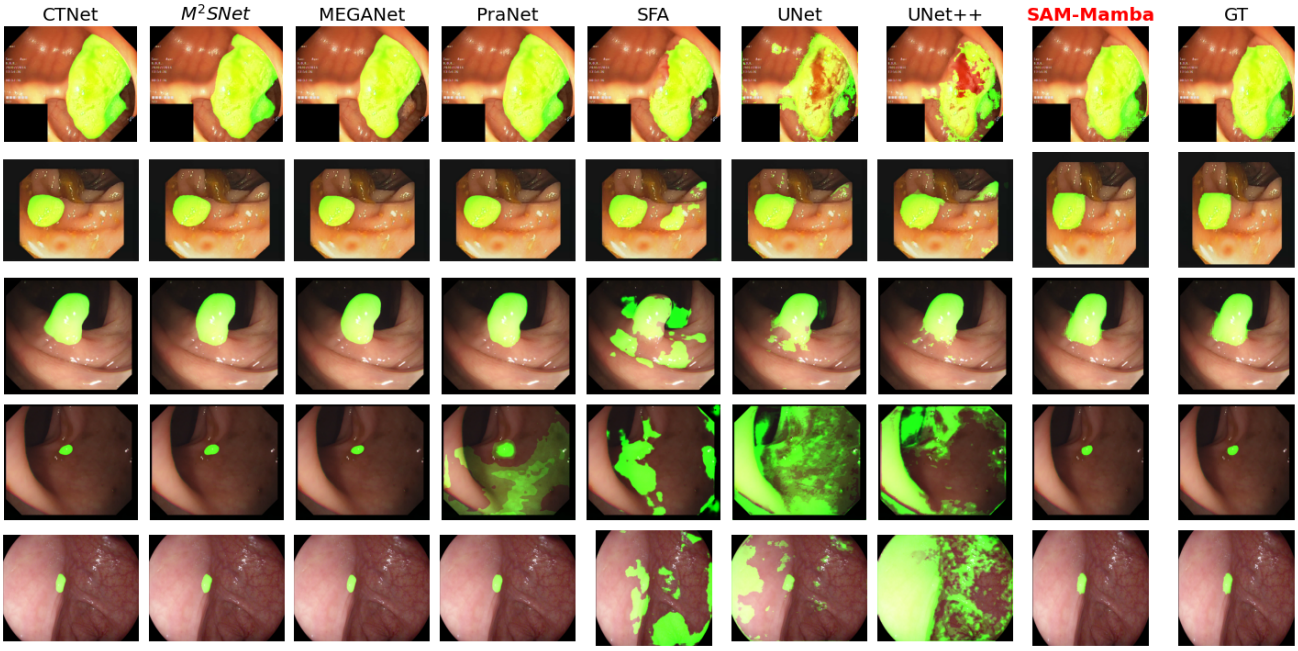

Figure 1. **Visualization and comparison** of segmentation results of state-of-the-art methods with SAM-Mamba (ours). Here, `row-1 and 2` corresponds to the samples from seen datasets *i.e.,* Kvasir-SEG and Clinic-DB. `Row-3,4 and 5` corresponds to the samples from unseen datasets *i.e.* CVC-300, CVC-ColonDB, and ETIS.

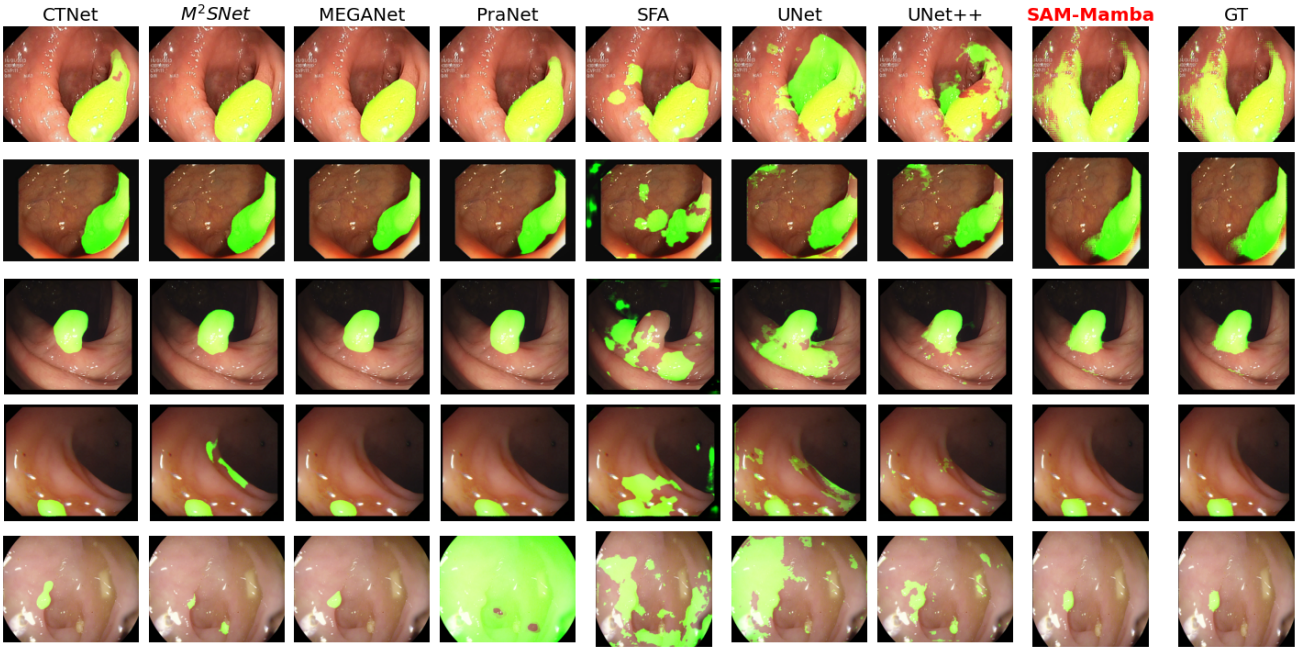

Figure 2. **Visualization and comparison** of segmentation results of state-of-the-art methods with SAM-Mamba (ours). Here, `row-1 and 2` corresponds to the samples from seen datasets *i.e.* Kvasir-SEG and Clinic-DB. `Row-3,4 and 5` corresponds to the samples from unseen datasets *i.e.* CVC-300, CVC-ColonDB, and ETIS.

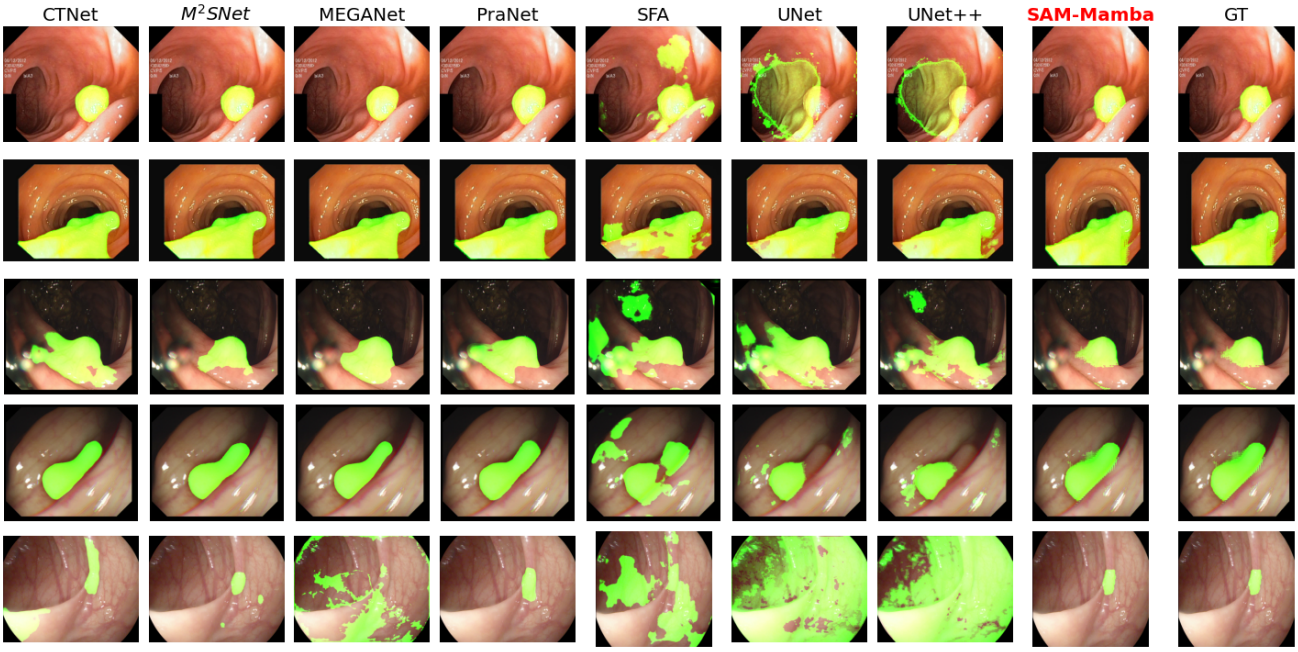

Figure 3. **Visualization and comparison** of segmentation results of state-of-the-art methods with SAM-Mamba (ours). Here, `row-1 and 2` corresponds to the samples from seen datasets *i.e.* Kvasir-SEG and Clinic-DB. `Row-3,4 and 5` corresponds to the samples from unseen datasets *i.e.* CVC-300, CVC-ColonDB, and ETIS.

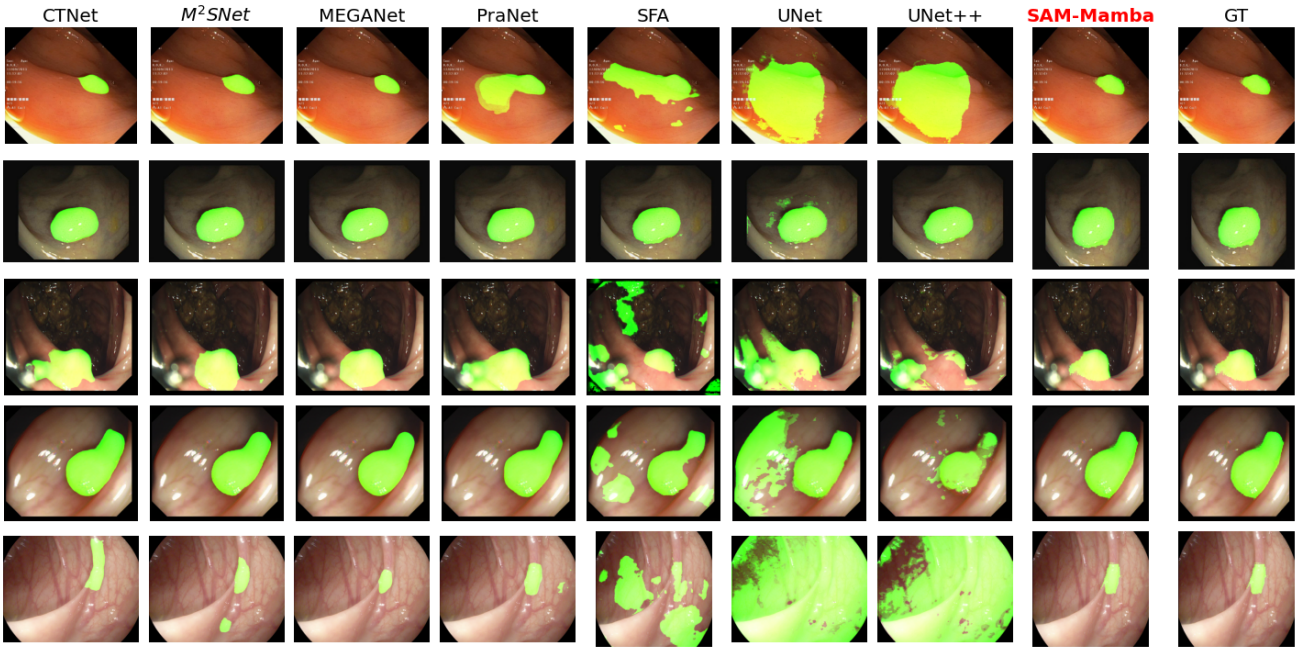

Figure 4. **Visualization and comparison** of segmentation results of state-of-the-art methods with SAM-Mamba (ours). Here, `row-1 and 2` corresponds to the samples from seen datasets *i.e.* Kvasir-SEG and Clinic-DB. `Row-3,4 and 5` corresponds to the samples from unseen datasets *i.e.* CVC-300, CVC-ColonDB, and ETIS.

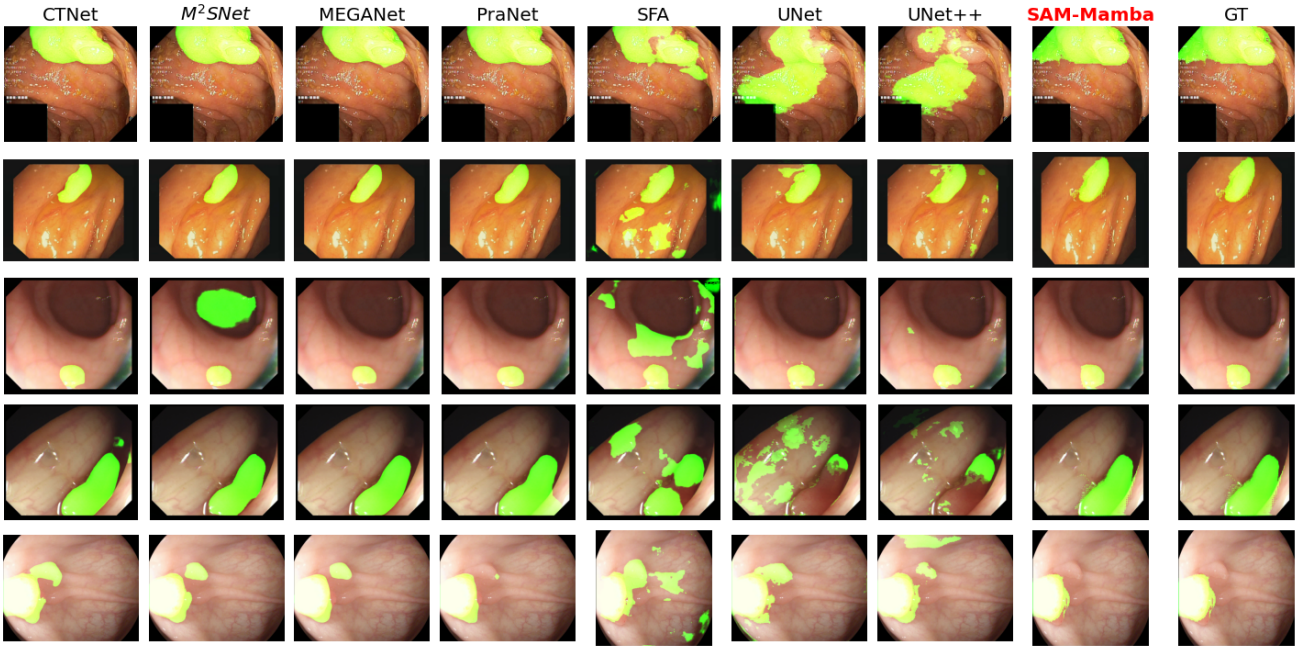

Figure 5. **Visualization and comparison** of segmentation results of state-of-the-art methods with SAM-Mamba (ours). Here, `row-1 and 2` corresponds to the samples from seen datasets *i.e.* Kvasir-SEG and Clinic-DB. `Row-3,4 and 5` corresponds to the samples from unseen datasets *i.e.* CVC-300, CVC-ColonDB, and ETIS.

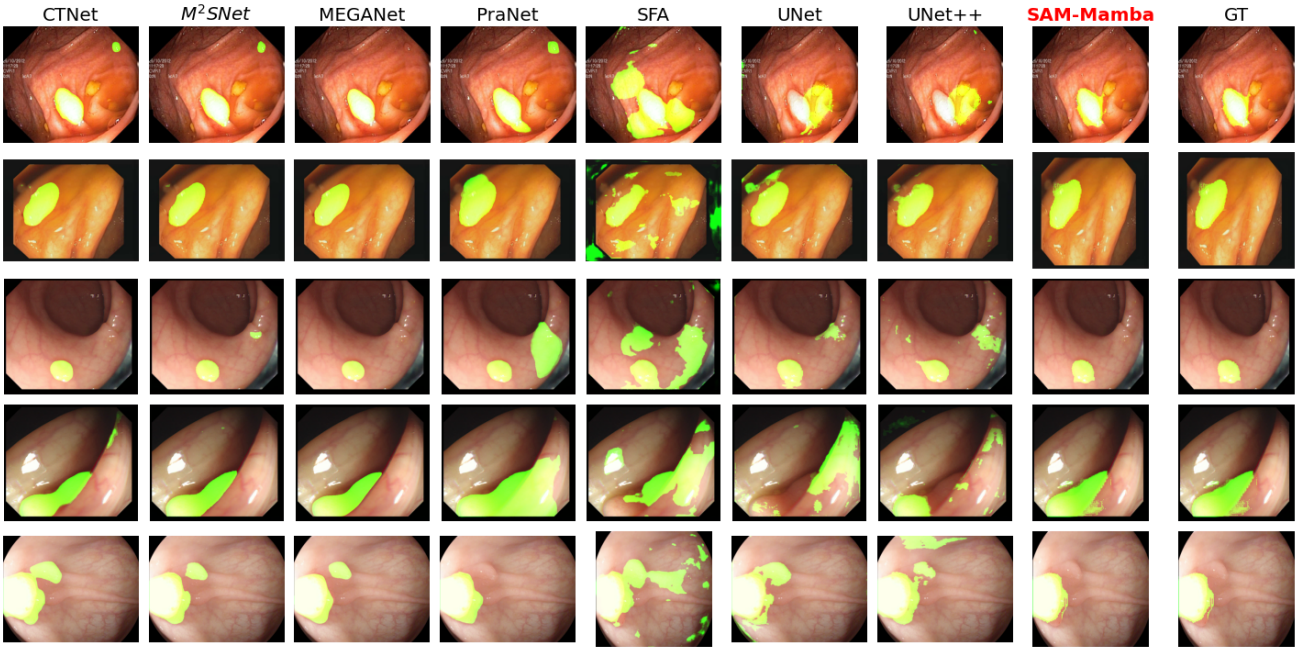

Figure 6. **Visualization and comparison** of segmentation results of state-of-the-art methods with SAM-Mamba (ours). Here, `row-1 and 2` corresponds to the samples from seen datasets *i.e.* Kvasir-SEG and Clinic-DB. `Row-3,4 and 5` corresponds to the samples from unseen datasets *i.e.* CVC-300, CVC-ColonDB, and ETIS.

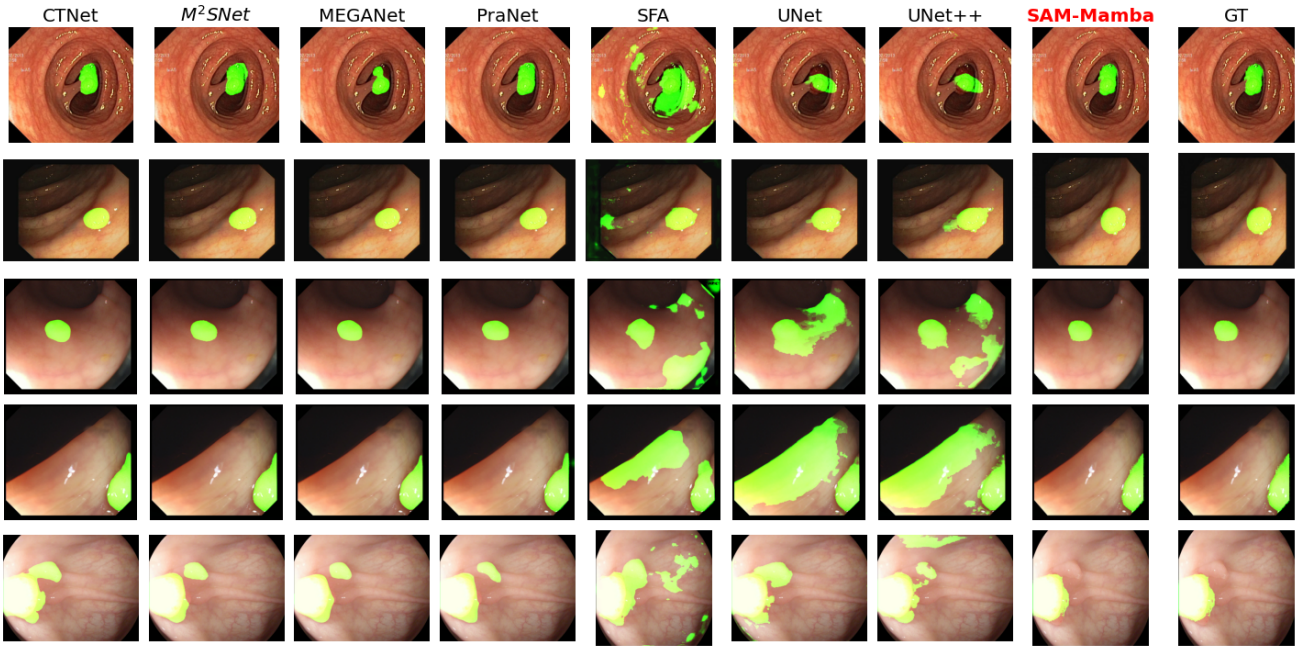

Figure 7. **Visualization and comparison** of segmentation results of state-of-the-art methods with SAM-Mamba (ours). Here, `row-1 and 2` corresponds to the samples from seen datasets *i.e.* Kvasir-SEG and Clinic-DB. `Row-3,4 and 5` corresponds to the samples from unseen datasets *i.e.* CVC-300, CVC-ColonDB, and ETIS.

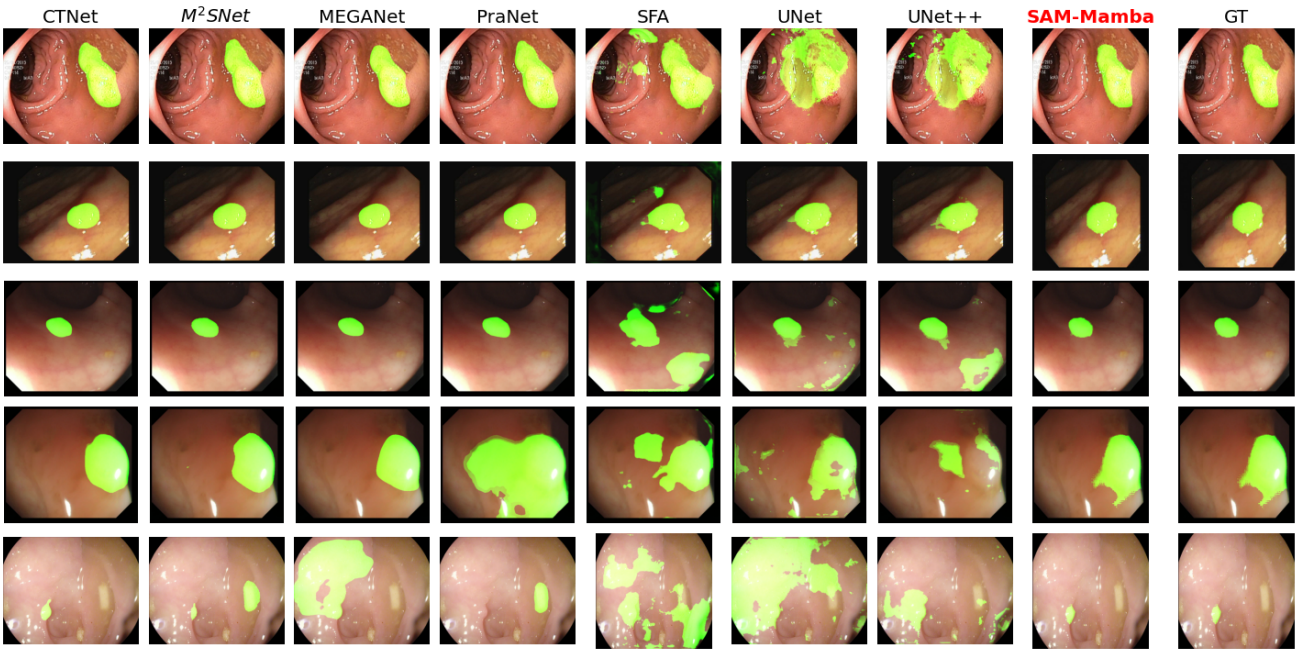

Figure 8. **Visualization and comparison** of segmentation results of state-of-the-art methods with SAM-Mamba (ours). Here, `row-1 and 2` corresponds to the samples from seen datasets *i.e.* Kvasir-SEG and Clinic-DB. `Row-3,4 and 5` corresponds to the samples from unseen datasets *i.e.* CVC-300, CVC-ColonDB, and ETIS.

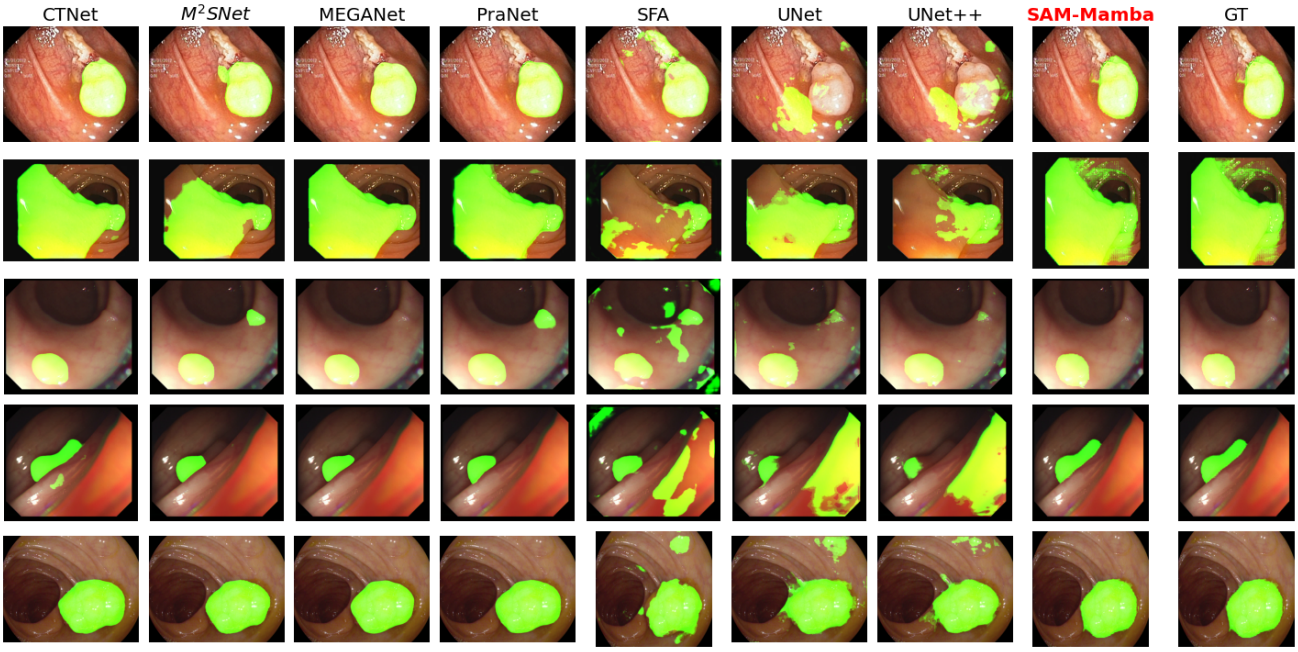

Figure 9. **Visualization and comparison** of segmentation results of state-of-the-art methods with SAM-Mamba (ours). Here, `row-1 and 2` corresponds to the samples from seen datasets *i.e.* Kvasir-SEG and Clinic-DB. `Row-3,4 and 5` corresponds to the samples from unseen datasets *i.e.* CVC-300, CVC-ColonDB, and ETIS.

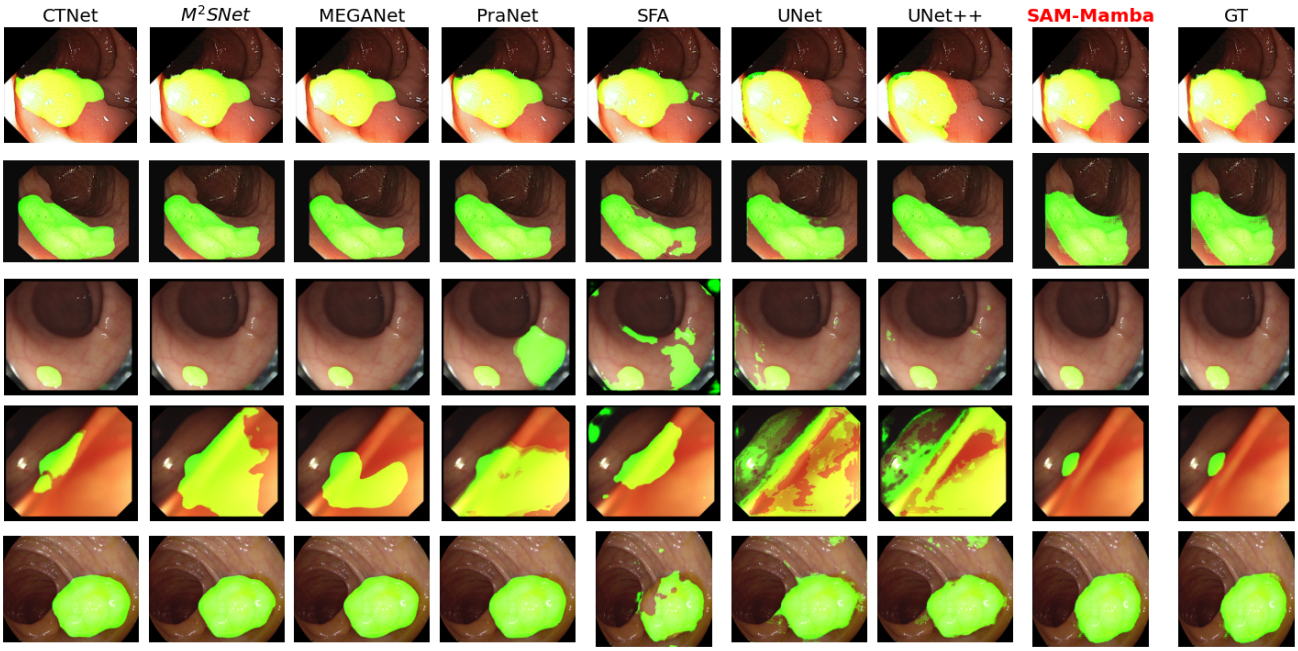

Figure 10. **Visualization and comparison** of segmentation results of state-of-the-art methods with SAM-Mamba (ours). Here, `row-1 and 2` corresponds to the samples from seen datasets *i.e.* Kvasir-SEG and Clinic-DB. `Row-3,4 and 5` corresponds to the samples from unseen datasets *i.e.* CVC-300, CVC-ColonDB, and ETIS.