PathSeqSAM: Sequential Modeling for Pathology Image Segmentation with SAM2

Mingyang Zhu¹ ¹ Shanghai Jiao Tong University, Shanghai, China Yinting Liu² ² University of Nebraska Medical Center, Omaha, USA Mingyu Li³ Jiacheng Wang³ ³ Vanderbilt University, Nashville, USA STJU18070285728@SJTU.EDU.CN

YILIU@UNMC.EDU

MINGYU.LI@VANDERBILT.EDU JIACHENG.WANG.1@VANDERBILT.EDU

Editors: Under Review for MIDL 2025

Abstract

Current methods for pathology image segmentation typically treat 2D slices independently, ignoring valuable cross-slice information. We present PathSeqSAM, a novel approach that treats 2D pathology slices as sequential video frames using SAM2's memory mechanisms. Our method introduces a distance-aware attention mechanism that accounts for variable physical distances between slices and employs LoRA for domain adaptation. Evaluated on the KPI Challenge 2024 dataset for glomeruli segmentation, PathSeqSAM demonstrates improved segmentation quality, particularly in challenging cases that benefit from cross-slice context. We have publicly released our code at https://github.com/JackyyyWang/PathSeqSAM. Keywords: Pathology image, SAM2, cross-slice attention, glomeruli segmentation

1. Introduction

Accurate segmentation of histopathological structures is fundamental for quantitative analysis of kidney pathology images, particularly in chronic kidney disease diagnosis (Deng et al., 2024, 2023). Traditional approaches often process each 2D pathology slice independently, overlooking potentially valuable contextual information from adjacent slices of the same specimen. This limitation becomes especially apparent in challenging cases with staining inconsistencies or complex pathological changes (Ginley et al., 2019; Altini et al., 2020).

Recent advances in foundation models, particularly SAM2 (Ravi et al., 2024), have shown promise in handling sequential data, but directly applying these methods to pathology remains challenging due to domain shift (Wu et al., 2023; Li et al., 2024; Ma et al., 2024). We propose PathSeqSAM, which interprets multiple 2D slices from the same subject as sequential video frames, enabling cross-slice contextual learning through:

- A sequential modeling strategy that treats pathology slices as video frames.
- A distance-aware attention mechanism to accommodate variable physical distances.
- Domain adaptation using Low-Rank Adaptation (LoRA) (Hu et al., 2021) for pathology-specific features.



Figure 1: Proposed SAM2 segmentation pipeline for pathology images. Features extracted via the LoRA-adapted SAM2 encoder are refined by distance-aware cross-slice attention and a memory bank, enabling coherent segmentation across slices.

2. Methods

Problem Formulation: Given a set of 2D pathology slices $\{S_1, S_2, ..., S_n\}$ from the same subject, we formulate a sequential segmentation task, where each slice functions as a frame in a video-like sequence. This viewpoint captures cross-slice relationships despite differences between typical video frames and pathology slices. As shown in Figure 1, our approach leverages the correspondence between pathology slices, where structures like glomeruli can be tracked across sequential slices, similar to objects moving through video frames.

Distance-Aware Cross-Slice Attention: SAM2 extends the original SAM (Kirillov et al., 2023) by incorporating a memory attention mechanism to handle sequential data. However, pathology slices often have variable physical distances, unlike uniformly spaced video frames. To address this, we introduce a distance-aware attention mechanism:

$$\alpha_{i,j} = \frac{\exp\left(sim(F_i, F_j) \cdot \phi(d_{i,j})\right)}{\sum_k \exp\left(sim(F_i, F_k) \cdot \phi(d_{i,k})\right)},\tag{1}$$

where F_i and F_j are feature embeddings for slices *i* and *j*, $sim(\cdot, \cdot)$ is a cosine similarity function, $d_{i,j}$ is the estimated physical distance, and $\phi(d) = \exp(-\lambda \cdot d^2)$ is a distance modulation function. The parameter λ is initialized to 0.1 and learned during training to adaptively weight the influence of physical distance on attention.

Adaptive Memory for Histopathology Context: We adopt an adaptive slice selection strategy instead of maintaining a fixed-size memory of recent frames, choosing the most informative slices based on feature similarity and confidence:

$$M_t = \{F_j \mid j \in \text{top-K}(sim(F_j, F_t) \cdot Conf(y_j)), j < t\}.$$
(2)

		0	
Me	thod	Mean±SD	
nn	UNet	$88.79 {\pm} 5.38$	
Sw	in-Unet	$89.65 {\pm} 6.41$	
\mathbf{SA}	M2	$92.48{\pm}6.13$	
Pa	thSeqSAM (Our	s) 94.71±5.89	

Table 1: Patch-level Diseased Glomeruli Segmentation Performance (Dice Score %)

Here, $Conf(y_j)$ represents the segmentation confidence derived from SAM2's cross-attention mechanism. This approach prioritizes slices with higher feature similarity and reliable segmentation confidence, regardless of strict sequential ordering. Such flexibility is critical for pathology slides, where the relationship between adjacent slices can be complex.

Domain Adaptation: We utilize Low-Rank Adaptation (LoRA) (Hu et al., 2021) to adapt SAM2's image encoder for the pathology domain. LoRA injects trainable low-rank matrices into attention layers, preserving most pre-trained weights. This method effectively handles domain adaptation with minimal computational overhead (Cheng et al., 2023).

3. Experimental and Results

We implemented PathSeqSAM on the SAM2 codebase (Ravi et al., 2024), applying LoRA with rank 8 to the image encoder. The model was trained on the KPI Challenge 2024 dataset (Deng et al., 2024) using a combined loss function:

$$\mathcal{L} = \mathcal{L}_{dice} + 0.5 \,\mathcal{L}_{BCE} + 0.2 \,\mathcal{L}_{consistency},\tag{3}$$

where $\mathcal{L}_{consistency}$ encourages consistent segmentation across similar slices by penalizing discrepancies in predictions between slices with high feature similarity (Ji et al., 2021). We set K = 5 to balance contextual information and computational efficiency. Physical distances were obtained from metadata or estimated via feature similarity.

Table 1 compares PathSeqSAM with state-of-the-art methods on patch-level glomeruli segmentation from the KPI Challenge 2024. PathSeqSAM achieved a mean Dice score of 94.71 ± 5.89 , outperforming nnUNet, Swin-Unet, and SAM2. The 2.23% improvement over SAM2 demonstrates the effectiveness of sequential modeling and distance-aware attention in pathology segmentation.

4. Discussion and Conclusion

PathSeqSAM introduces a sequential modeling paradigm for pathology image segmentation by leveraging SAM2's memory attention across multiple slices. The distance-aware attention mechanism and LoRA-based domain adaptation address the unique challenges of histopathological data, such as variable inter-slice spacing and staining inconsistencies. By prioritizing the most informative slices, our adaptive memory approach further enhances segmentation consistency and accuracy.

Acknowledgments

We thank the organizers of the KPI Challenge 2024 for providing the dataset.

References

- N Altini, G D Cascarano, A Brunetti, F Marino, M T Rocchetti, S Matino, U Venere, M Rossini, F Pesce, L Gesualdo, and V Bevilacqua. Semantic segmentation framework for glomeruli detection and classification in kidney histological sections. *Electronics*, 9 (3):503, 2020.
- Junlong Cheng, Jin Ye, Zhongying Deng, Jianpin Chen, Tianbin Li, Haoyu Wang, Yanzhou Su, Ziyan Huang, Jilong Chen, Lei Jiang, et al. Sam-med2d. arXiv preprint arXiv:2308.16184, 2023.
- Ruining Deng, Quan Liu, Can Cui, Tianyuan Yao, Jiancheng Long, Zaid Asad, Rebecca M Womick, Zhaoxiang Zhu, Agnes B Fogo, Shilin Zhao, et al. Omni-seg: A scale-aware dynamic network for renal pathological image segmentation. *IEEE Transactions on Biomedical Engineering*, 70(9):2636–2644, 2023.
- Ruining Deng, Tianyuan Yao, Yucheng Tang, Junlin Guo, Siqi Lu, Juming Xiong, Lining Yu, Quan Huu Cape, Pengzhou Cai, Libin Lan, et al. Kpis 2024 challenge: Advancing glomerular segmentation from patch- to slide-level. *Medical Image Analysis*, 2024.
- Brandon Ginley, Brendon Lutnick, Kuang-Yu Jen, Agnes B Fogo, Sanjay Jain, Avi Rosenberg, Vighnesh Walavalkar, Gregory Wilding, John E Tomaszewski, Rabi Yacoub, et al. Computational segmentation and classification of diabetic glomerulosclerosis. *Journal of* the American Society of Nephrology, 30(10):1953–1967, 2019.
- Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv* preprint arXiv:2106.09685, 2021.
- Wei Ji, Shuang Yu, Junde Wu, Kai Ma, Cheng Bian, Qi Bi, Jingjing Li, Hanruo Liu, Li Cheng, and Yefeng Zheng. Learning calibrated medical image segmentation via multirater agreement modeling. In *Proceedings of the IEEE/CVF Conference on Computer* Vision and Pattern Recognition, pages 12341–12351, 2021.
- Alexander Kirillov, Eric Mintun, Nikhila Ravi, Hanzi Mao, Chloe Rolland, Laura Gustafson, Tete Xiao, Spencer Whitehead, Alexander C. Berg, Wan-Yen Lo, Piotr Dollár, and Ross Girshick. Segment anything. arXiv preprint arXiv:2304.02643, 2023.
- Hao Li, Han Liu, Dewei Hu, Jiacheng Wang, and Ipek Oguz. Promise: Prompt-driven 3d medical image segmentation using pretrained image foundation models. In 2024 IEEE International Symposium on Biomedical Imaging (ISBI), pages 1–5. IEEE, 2024.
- Jun Ma, Yuting He, Feifei Li, Lin Han, Chenyu You, and Bo Wang. Segment anything in medical images. *Nature Communications*, 15(1):654, 2024.
- Nikhila Ravi, Valentin Gabeur, Yuan-Ting Hu, Ronghang Hu, Chaitanya Ryali, Tengyu Ma, Haitham Khedr, Roman R\u00e4dle, Chloe Rolland, Laura Gustafson, Eric Mintun, Junting Pan, Kalyan Vasudev Alwala, Nicolas Carion, Chao-Yuan Wu, Ross Girshick, Piotr Doll\u00e1r, and Christoph Feichtenhofer. Sam 2: Segment anything in images and videos. arXiv preprint arXiv:2402.19642, 2024.

PATHSEQSAM

Junde Wu, Rao Fu, Huihui Fang, Yuanpei Liu, Zhaowei Wang, Yanwu Xu, Yueming Jin, and Tal Arbel. Medical sam adapter: Adapting segment anything model for medical image segmentation. arXiv preprint arXiv:2304.12620, 2023.