Topo-GraT: Learning to Grow with Causal Graph Transformers (Student Abstract)

Ashim Dhor[1]*, **Smily Bharadwaj**[1]

[1]Indian Institute of Science Education and Research Bhopal, Madhya Pradesh 462066, India
ashimdhor2003@gmail.com, smily23@iiserb.ac.in

Abstract

Automated cancer segmentation in Whole Slide Images (WSIs) has been dominated by a paradigm of static pattern recognition, where even advanced methods leveraging Transformers, Multiple Instance Learning, or topology-aware losses remain fundamentally descriptive and correlational. To address this limitation, we reframe WSI segmentation from a descriptive task to one of causal process modeling. We introduce Topo-GraT, a novel framework featuring a Causal Growth Field (CGF) to model tumor invasion dynamics and a Causal Flow Attention (CFA) mechanism that embeds this field as an architectural prior. This causal engine is integrated within an iterative graph refinement loop that uses segmentation uncertainty to dynamically focus computational resources on the most ambiguous tissue regions. Our comprehensive experiments on multiple WSI datasets demonstrate that Topo-GraT establishes a new state-of-the-art, significantly outperforming existing methods and reducing the 95% Hausdorff Distance, a key boundary metric, by over 15%. Crucially, our framework yields the CGF as a rich, interpretable output whose structure correlates with tumor aggressiveness, positioning it as a novel biomarker for downstream prognostic tasks. By shifting the paradigm from static recognition to causal reasoning, Topo-GraT offers a more robust, efficient, and clinically insightful approach, setting a new direction for the causally-aware medical image analysis.

Code — https://github.com/AshimDhor/Topo-GraT

Introduction

Histopathological analysis remains the gold standard for cancer diagnosis, with Whole Slide Imaging (WSI) opening the door for computational pathology (Van der Laak, Litjens, and Ciompi 2021). However, the gigapixel scale of WSIs and the morphological heterogeneity of tumors pose profound challenges. Current methods, including U-Net-based architectures (Ronneberger, Fischer, and Brox 2015), Transformer models like UNETR (Hatamizadeh et al. 2022), and MIL-based approaches like TransMIL (Shao et al. 2021), treat segmentation as a static pattern recognition task. Even topology-aware methods like TA-Net (Wang, Xian, and Vakanski 2022) are descriptive, using geometric abstractions of the final tumor shape rather than modeling the underlying biological process. This descriptive approach limits robustness in ambiguous cases and restricts the output to inert masks that lack prognostic insight. To bridge this gap, we introduce Topo-GraT, a framework that reframes WSI segmentation as a task of causal process modeling. We introduce the Causal Growth Field (CGF) to model the dynamics of tumor progression and a novel Causal Flow Attention (CFA) that embeds this model as a strong inductive bias. This causal engine is integrated within an iterative graph refinement loop, allowing the model to dynamically allocate resources to the most challenging tissue regions.

Methodology

Topo-GraT operates as a two-stage, iterative framework.

Stage 1: Iterative Graph Construction. The framework first transforms the WSI into a sparse graph. A Swin Transformer (Liu et al. 2021) extracts patch features. A salience score, which combines a patch's cancer probability with its contextual importance via a gating mechanism (Ilse, Tomczak, and Welling 2018), is used to select an initial set of K patch nodes. In subsequent iterations ($t > 0$), a pixel-wise uncertainty map from Stage 2 is used to guide the selection of K' new patches from the most ambiguous regions, a form of self-directed active learning (Settles 2009). This allows the graph to dynamically expand and focus on challenging boundaries. Stage 2: Causal Graph Transformer Segmentation. The graph nodes are processed by a hierarchical encoder-decoder network built on our novel Causal Topology-Guided EPA (CTG-EPA) blocks. The core of this block is the CGF prediction branch, which learns a vector field modeling the direction of tumor growth. The CGF branch is supervised using a pseudo-ground-truth vector field generated by inverting the gradient of a distance transform from the boundary of the ground truth mask, with the origin defined by the medial axis. This CGF is then used to create a causal attention mask for our CFA mechanism, which constrains the flow of information in the self-attention module to follow causally-plausible pathways. Specifically, the attention score from a token i to a token j is up-weighted if j lies "downstream" along the growth vector from i, and suppressed otherwise. This embeds the causal model directly into the architecture, providing a strong inductive bias.

The model is trained end-to-end with a composite loss

*Corresponding author.

function: $L_{\text{total}} = L_{\text{seg}} + \alpha L_{\text{topo}} + \beta L_{\text{instance}}$. The segmentation loss (L_{seg}) is a combination of Dice and Cross-Entropy losses for pixel-level accuracy. The topology preservation loss (L_{topo}) is a Cosine Similarity loss that directly supervises the CGF branch, compelling it to learn the correct *direction* of tumor growth. The instance selection loss (L_{instance}) is a Binary Cross-Entropy loss that trains the Stage 1 patch selector. This multi-task approach ensures all components of the framework are jointly optimized.

Experiments and Results

We evaluated Topo-GraT on four WSI datasets, including a held-out dataset HNSCC (Grossberg et al. 2020) to test for out-of-domain generalization. The results, summarized in Table 1 and Table 2, demonstrate that Topo-GraT consistently establishes a new state-of-the-art and that each of our novel components provides a significant performance gain. Table 1 benchmarks Topo-GraT against a suite of leading models. On our primary benchmarks - CAMELYON16 (C16) (Bejnordi et al. 2017), GlaS challenge (Sirinukunwattana et al. 2017), RINGS (Salvi et al. 2021), Topo-Grat outperforms all baselines. The efficacy of our causal approach is evident in the boundary-based metrics: Topo-GraT reduces the 95% Hausdorff Distance by an average of 15.7% compared to the strongest baseline, UNETR++. The model shows robust out-of-domain (OOD) generalization on the unseen HNSCC dataset, maintaining a high Dice score of 87.2% where baselines experience a more pronounced performance degradation.

Model	C16		GlaS		RINGS		HNSCC	
	DSC ↑	HD95 ↓	DSC ↑	HD95 ↓	DSC ↑	HD95 ↓	DSC ↑	HD95 ↓
U-Net	81.5	16.4	85.2	15.1	83.1	17.2	80.1	18.9
UNETR	83.2	12.8	87.1	13.5	84.9	14.1	81.5	16.5
Swin-UNET	84.0	11.5	88.5	12.4	86.2	12.9	82.1	15.2
UNETR++	84.3	10.2	89.1	12.1	86.8	12.5	82.5	14.1
DTFD-MIL	78.1	21.5	82.3	18.9	80.5	22.4	75.5	25.8
MedNeXt	84.5	10.8	89.3	11.9	87.0	12.2	82.8	14.5
TA-Net	85.1	9.9	89.9	11.8	87.5	11.9	83.1	13.5
Topo-GraT	**89.3**	**8.1**	**91.2**	**9.1**	**89.8**	**9.5**	**87.2**	**10.3**

Table 1: Unified quantitative comparison with state-of-the-art models across all datasets. Topo-GraT consistently outperforms baselines on primary benchmarks and demonstrates superior out-of-domain (OOD) generalization.

A systematic ablation study on CAMELYON16 (Table 2) validates the contribution of each novel component. While the MIL selection improves efficiency, the most significant performance gain is derived from the architectural integration of the CGF via Causal Flow Attention (CFA). The final addition of the iterative loop provides a crucial boost to boundary metrics, confirming the value of our full, synergistic system. To validate our choice of T=3 for the number of iterative refinement steps, we conducted an ablation study on the CAMELYON16 dataset. As shown in Table 3, the model's performance, particularly on the boundary-sensitive HD95 metric, improves with each iteration. The most significant gains are observed when moving from a non-iterative model (T=0, which is configuration (4) from our main ablation study) to one iteration (T=1). The improvements continue up to T=3, after which the performance gains become marginal while the inference time continues to increase. This confirms that T=3 represents the optimal balance between segmentation accuracy and computational efficiency for our framework.

Configuration	DSC(%)↑	IoU(%)↑	HD95↓	FLOPs(G)↓*
(1) UNETR++	84.3	73.8	10.2	289.5
(2) + MIL Selection	85.9	75.8	9.9	**168.2**
(3) + CGF Loss	86.7	77.0	9.5	172.8
(4) + CFA	88.2	79.1	8.8	174.1
(5) Full Topo-GraT	**89.3**	**80.7**	**8.1**	181.5

*FLOPs are measured for a single forward pass.

Table 2: Ablation study on CAMELYON16

The robustness of Topo-GraT on non-radial tumors stems from the flexibility of the CGF prediction branch. As a convolutional network, it is not restricted to generating purely radial fields. Instead, it learns the underlying principle of predicting vectors that point from the tumor's medial axis outwards to its boundary, regardless of the overall shape. Even for elongated morphologies, the resulting non-radial CGF provides a valid directional prior for the Causal Flow Attention mechanism. Consequently, the model maintains high segmentation accuracy in these challenging cases.

Iterations (T)	DSC(%) ↑	HD95 ↓	Inference Time (s)↓
0 (No Iteration)	88.2	8.8	4.7
1	88.9	8.3	4.9
2	89.2	8.2	5.0
3 (Ours)	**89.3**	**8.1**	**5.1**
4	89.3	8.1	5.3

Table 3: Ablation study on the number of iterative refinement steps (T) on CAMELYON16

Conclusion

In this work, we reframed WSI segmentation from a static pattern recognition task to one of causal process modeling. Our novel framework, Topo-GraT, integrates a Causal Growth Field and Causal Flow Attention to create a system that learns not just what a tumor looks like, but how it grows. Our results show that this causal approach yields state-of-the-art segmentation accuracy and produces a rich, interpretable CGF that has potential as a new prognostic biomarker, setting a new direction for the field of computational pathology. Looking forward, the implications of this work extend beyond segmentation: Topo-GraT shifts from mere pattern recognition to understanding tumor growth, enabling prediction of progression and automated grading, and laying a foundation for a more dynamic future in computational pathology.

Acknowledgments

We gratefully acknowledge the Indian Institute of Science Education and Research (IISER) Bhopal for providing the computational resources and research environment that made this work possible. We would like to express our sincere gratitude to Dr. Tanmay Basu and Biomedical Data Science Research Lab, IISER Bhopal for their constant support, valuable guidance, and encouragement throughout the course of this work.

References

Bejnordi, B. E.; Veta, M.; Van Diest, P. J.; Van Ginneken, B.; Karssemeijer, N.; Litjens, G.; Van Der Laak, J. A.; Hermsen, M.; Manson, Q. F.; Balkenhol, M.; et al. 2017. Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *Jama*, 318(22): 2199–2210.

Grossberg, A.; Elhalawani, H.; Mohamed, A.; Mulder, S.; Williams, B.; White, A. L.; Zafereo, J.; Wong, A. J.; Berends, J. E.; AboHashem, S.; et al. 2020. Hnscc. *(No Title)*.

Hatamizadeh, A.; Nath, V.; Tang, Y.; Yang, D.; Myronenko, A.; Landman, B.; Roth, H. R.; and Xu, D. 2022. Unetr: Transformers for 3d medical image segmentation. In *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, 574–584.

Ilse, M.; Tomczak, J.; and Welling, M. 2018. Attention-based deep multiple instance learning. In *International conference on machine learning*, 2127–2136. PMLR.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision*, 10012–10022.

Ronneberger, O.; Fischer, P.; and Brox, T. 2015. U-net: Convolutional networks for biomedical image segmentation. In *International Conference on Medical image computing and computer-assisted intervention*, 234–241. Springer.

Salvi, M.; Bosco, M.; Molinaro, L.; Gambella, A.; Papotti, M.; Acharya, U. R.; and Molinari, F. 2021. A hybrid deep learning approach for gland segmentation in prostate histopathological images. *Artificial Intelligence in Medicine*, 115: 102076.

Settles, B. 2009. Active learning literature survey. *University of Wisconsin-Madison Department of Computer Sciences*.

Shao, Z.; Bian, H.; Chen, Y.; Wang, Y.; Zhang, J.; and Zhang, H. 2021. Transmil: Transformer based correlated multiple instance learning for whole slide image classification. In *Advances in Neural Information Processing Systems*, volume 34, 2136–2147.

Sirinukunwattana, K.; Pluim, J. P.; Chen, H.; Qi, X.; Heng, P.-A.; Guo, Y. B.; Wang, L. Y.; Matuszewski, B. J.; Bruni, E.; Sanchez, U.; et al. 2017. Gland segmentation in colon histology images: The glas challenge contest. *Medical image analysis*, 35: 489–502.

Van der Laak, J.; Litjens, G.; and Ciompi, F. 2021. Deep learning in histopathology: the path to the clinic. *Nature medicine*, 27(5): 775–784.

Wang, H.; Xian, M.; and Vakanski, A. 2022. TA-Net: Topology-aware network for gland segmentation. *IEEE transactions on medical imaging*, 41(12): 3637–3648.