Self Supervised Vision Transformers for Learning 3D Colorectal Cancer Tissue Architecture from Open-Top Light Sheet Microscopy

Comparing 2D, 2.5D, and 3D Self-Supervised Foundation Models for Colorectal Cancer Tissue

Animesh Page, Aniruddh Maheshwari, Prahlad Uchila, Gautham Agilan

Edit AI Internship, Dartmouth College | Mentors: Dr. Levy, Dr. Khang

Introduction

Hematoxylin and eosin (H&E) staining is the foundation of tumor pathology and the substrate for nearly all archived clinical tissue, yet it is fundamentally limited: it reveals tissue structure but carries no direct molecular readout, and it is two-dimensional, discarding the volumetric architecture of tissue. Spatial transcriptomics (ST) recovers molecular state in situ, but it is sparse, expensive, and typically limited to a few 2D sections, so it cannot scale to whole tumors or the millions of archived H&E slides - leaving a persistent gap between the modality that is abundant (H&E) and the information that is most informative (ST). Open-Top Light-Sheet (OTLS) microscopy closes the imaging side of that gap by capturing intact tissue blocks in 3D, non-destructively and at near-histologic resolution, with fluorescence rendered as familiar virtual H&E. This enables a self-supervised 3D foundation model that learns tumor structure directly from volumes and offers a scaffold for registering sparse ST into 3D as future supervision. Colorectal cancer - the second-deadliest cancer overall, rising sharply in adults under 50 - is a setting where scalable, morphology-driven biomarker discovery is urgently needed. We build this pipeline end to end and directly test whether the third dimension helps.

Research Questions

- Does 3D volumetric modeling yield stronger representations than a 2D single-slice model, given its higher cost?
- Is 2.5D channel-stacking an effective middle ground, or must depth be a true spatial axis?
- Where does 3D help most - accuracy, depth-consistency, or boundary detection?

Methods & Materials

OTLS volumes from five colorectal blocks (paired nuclear + cytoplasmic channels, ~1,800-1,980 z-planes at ~7,000 x 8,000 px) were converted to RGB virtual H&E on the Dartmouth Discovery HPC cluster, and tissue-filtered 3D patches trained three masked-autoencoder (MAE) models with a ViT encoder (75% masking, 100 epochs), differing only in depth handling.

1. **OTLS imaging - image intact tissue blocks non-destructively in 3D; acquire paired nuclear + cytoplasmic fluorescence channels.**
2. **Virtual H&E - apply a Beer-Lambert absorption model per z-slice to render RGB pseudo-H&E (nuclei purple, cytoplasm pink).**
3. **3D volumes - generate whole-volume pseudo-H&E slice-by-slice into resumable, memory-mapped TIFFs on the HPC cluster.**
4. **3D patches - stream tissue-filtered [3, D, H, W] samples from the finished volumes, rejecting near-white background.**
5. **Self-supervised MAE - train 2D, 2.5D, and 3D masked-autoencoder variants, then evaluate via linear probes and clustering.**

Figures & Results

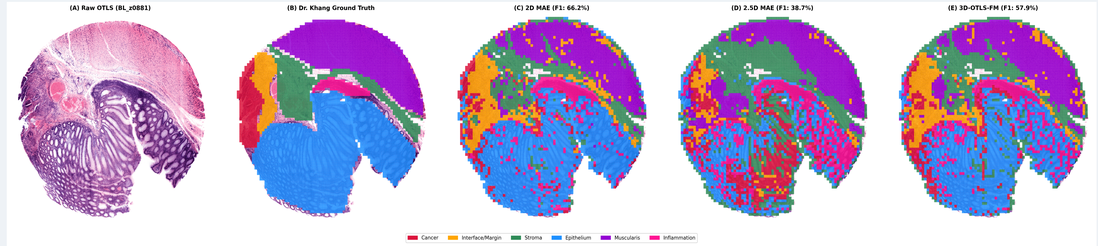
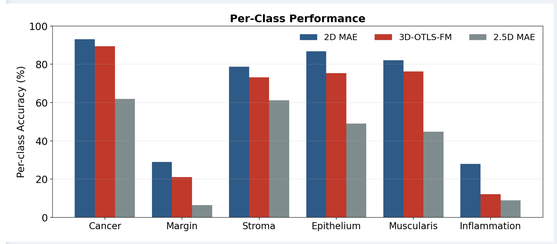
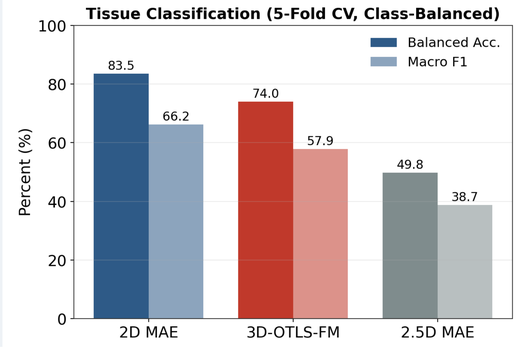
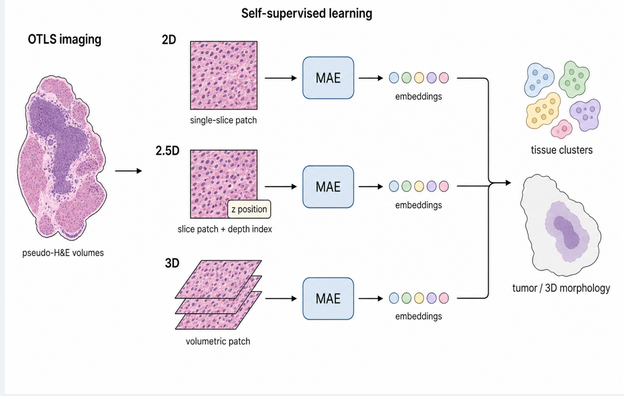


Figure 1. *Dense tissue segmentation on a held-out section: (A) raw OTLS virtual H&E; (B) Dr. Khang pathologist ground truth; (C-E) 2D, 2.5D, and 3D predictions (macro F1 shown). The 3D model recovers boundary structure the 2D model smooths over.*







Results

- 2D Architecture Leads Diagnostic Accuracy: 2D ViT-MAE achieved the highest 5-fold balanced accuracy (83.5%, 66.2% macro F1), followed by 3D (74.0%, 57.9% F1), while 2.5D channel-stacking collapsed (49.8%, 38.7% F1).
- Robust Malignancy Detection: Both 2D and 3D models achieved high cancer recall (93.0% in 2D, 89.4% in 3D) and robust recovery of benign stroma and epithelium (73–87%), while boundary-defined invasive margins remained challenging for all models (21–29%).
- Channel-Stacking (2.5D) Breaks Optical Physics: Fusing depth slices into color channels provides depth without geometric reasoning, proving that volumetric context requires true 3D tubelet tokenization.
- 3D Volumetric Continuity Advantage: The 3D model maintained superior axial stability along the
- Z
- Z-axis with lower prediction flicker (1.36% vs 1.79% across 5,903 coordinates), capturing continuous infiltrating margins that 2D missed.
- Autonomous Unsupervised Clustering: Pretrained with 75% masked autoencoding without human labels, latent embeddings autonomously grouped by true tissue histology rather than specimen batch artifacts.

Figure 4. *Embedding clusters at z=1040 for the 2D, 2.5D, and 3D encoders vs. the volumetric OTLS section; the 3D clusters are the most spatially coherent.*

Discussion

These results give a nuanced answer. On raw class-balanced accuracy the simple 2D model is strongest, and its far lower computational cost makes it attractive. But accuracy alone understates what 3D captures: the 3D model's better depth-consistency and its recovery of boundaries invisible to 2D show it encodes real volumetric structure a 2D model cannot access in principle. The failure of the 2.5D variant is instructive - stacking slices as channels supplies depth without a mechanism to reason over it, and can even hinder learning; depth helps only when modeled as a true spatial axis. A likely factor is that the 2D model benefits from a larger pool of independent training views and an easier reconstruction task, while our probe - trained on 2D annotations drawn on flat sections - under-credits the depth-wise capability that motivates 3D modeling. 3D-aware evaluation, or molecular targets registered in 3D, could reveal a larger advantage.

Conclusion

We built an end-to-end pipeline converting raw OTLS fluorescence volumes into virtual-H&E training data, and used it to train and compare three self-supervised foundation models that differ only in how they treat depth. Under a common class-balanced evaluation, the 2D model attained the best aggregate accuracy (83.5%), the 3D model followed closely (74.0%) while producing more depth-consistent predictions and recovering boundaries hidden to 2D, and 2.5D underperformed markedly (49.8%). The evidence supports 3D modeling as capturing genuine volumetric structure even where it does not dominate aggregate metrics, and motivates 3D-aware evaluation. The next step is to register spatial transcriptomics into the OTLS volumes and fine-tune the 3D encoder toward dense, whole-volume molecular reconstruction on archived H&E cohorts.

Limitations

The study uses five blocks from a single specimen, so absolute numbers may not generalize across specimens, sites, or scanners. Annotations are defined and evaluated on 2D sections, which structurally favors 2D-friendly features. The 3D model was trained on relatively few volumetric samples at modest patch sizes; scaling data and volume size could shift the balance. Molecular prediction - the ultimate goal - is not yet evaluated, so tissue classification serves as a proxy for representation quality.

Acknowledgments

We thank our mentors Dr. Levy and Dr. Khang for guidance and pathologist annotations, and the Edit AI program at Dartmouth College for compute access on the Discovery HPC cluster.

References

1. Gurcan MN, Boucheron LE, Can A, Madabhushi A, Rajpoot NM, Yener B. Histopathological Image Analysis: A Review. IEEE Reviews in Biomedical Engineering [Internet]. 2009 [cited 2026 Aug 26];2:147–71. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2910932/
2. McInnes, Elizabeth F. *Artifacts in Histopathology*. *Background Lesions in Laboratory Animals*, 2012, pp. 93–99, https://doi.org/10.1016/B978-0-7020-3519-7.00008-5.
3. Laurino A, Franceschini A, Pesce L, Cinci L, Montalbano A, Mazzamuto G, et al. A Guide to Perform 3D Histology of Biological Tissues with Fluorescence Microscopy. International Journal of Molecular Sciences [Internet]. 2023 Apr 4 [cited 2026 Aug 27];24(7):6747. Available from: https://pmc.ncbi.nlm.nih.gov/articles/PMC10094801/
4. Keklikoglou K, Arvanitidis C, Chatzigeorgiou G, Chatzinikolaou E, Karagiannidis E, Koletsa T, et al. Micro-CT for Biological and Biomedical Studies: A Comparison of Imaging Techniques. Journal of Imaging. 2021 Sept 1;7(9):172.
5. OTLS Microscopy [Internet]. The Jon Liu Lab. 2025 [cited 2026 Aug 27]. Available from: https://med.stanford.edu/jonliulab/research/otls.html
6. Shia J, Schultz N, Kuk D, Vakiani E, Middha S, Segal NH, et al. Morphological characterization of colorectal cancers in The Cancer Genome Atlas reveals distinct morphology–molecular associations: clinical and biological implications. Modern Pathology [Internet]. 2017 Apr 1 [cited 2026 Aug 26];30(4):599–609. Available from: https://www.nature.com/articles/modpathol2016198
7. *Chen RJ, Ding T, Lu MY, Drew, Jaume G, Chen B, et al. A General-Purpose Self-Supervised Model for Computational Pathology [Internet]. arXiv.org. 2023 [cited 2026 Aug 26]. Available from:* https://arxiv.org/abs/2308.15474
8. Awan R, Sirinukunwattana K, Epstein D, Jefferyes S, Qidwai U, Aftab Z, et al. Glandular Morphometrics for Objective Grading of Colorectal Adenocarcinoma Histology Images. Scientific Reports [Internet]. 2017 Dec 4 [cited 2026 Aug 26];7(1):16852. Available from: https://www.nature.com/articles/s41598-017-16516-w
9. Voigt B, Fischer O, Schilling B, Krumnow C, Herta C. Investigation of semi- and self-supervised learning methods in the histopathological domain. Journal of pathology informatics [Internet]. 2023 Dec [cited 2026 Aug 26];14:100305. Available from: https://pubmed.ncbi.nlm.nih.gov/37025325/