

There's No Place Like Genome: Mapping the Architecture of Spatial Transcriptomic AI

Jessica Xu^{1,3}, Dr. Guangyu Wang²

¹William P. Clements High School, Sugar Land, TX

²Houston Methodist Research Institute, Houston, TX

³Gifted and Talented Mentorship Program, Fort Bend ISD, TX

Abstract

The spatial organization of cells within tumor tissue is a critical determinant of disease progression and treatment response. Spatial transcriptomics captures gene expression across thousands of tissue locations while preserving spatial coordinates, but extracting insight from this high-dimensional data requires sophisticated computational tools.

This project developed a publication-quality architectural diagram for a self-supervised Vision Transformer (ViT) segmentation model in Dr. Guangyu Wang's lab at Houston Methodist, intended for the lab's forthcoming research paper.

Introduction and Summary

Two patients with nearly identical tumor profiles can respond to the same treatment completely differently, pointing to something beyond genetics. Research shows that where a cell is located in tissue can be just as consequential as what it is doing genetically. Spatial transcriptomics addresses this gap by mapping RNA activity across tissue simultaneously, revealing tumor heterogeneity and microenvironment interactions. High-resolution platforms like Visium HD now enable whole-transcriptome discovery at single-cell-scale resolution, making sophisticated AI tools essential for interpretation.

The ViT model uses a three-component teacher-student architecture (Image Teacher (fixed weights), ST Teacher, and ST Student) to identify spatially distinct tissue regions sharing similar transcriptomic profiles, at a scale between single-cell and whole-tissue. It is trained through contrastive loss and masked patch reconstruction objectives. This project produced a comprehensive architectural diagram of the full pipeline for the lab's forthcoming publication

Methodology

Phase 1: A literature review established foundational knowledge of spatial transcriptomics, ViT architecture, and self-supervised learning, including seminal works like "Attention Is All You Need."

Phase 2: The diagram was developed iteratively, visualizing data flow through the three-component structure, crop resolution distinctions, and both training objectives.

Phase 3: Weekly review sessions with postdoc Hannah Zhao refined each draft for technical accuracy and publication-quality visual clarity, with final approval from Dr. Wang's lab

Results and Findings

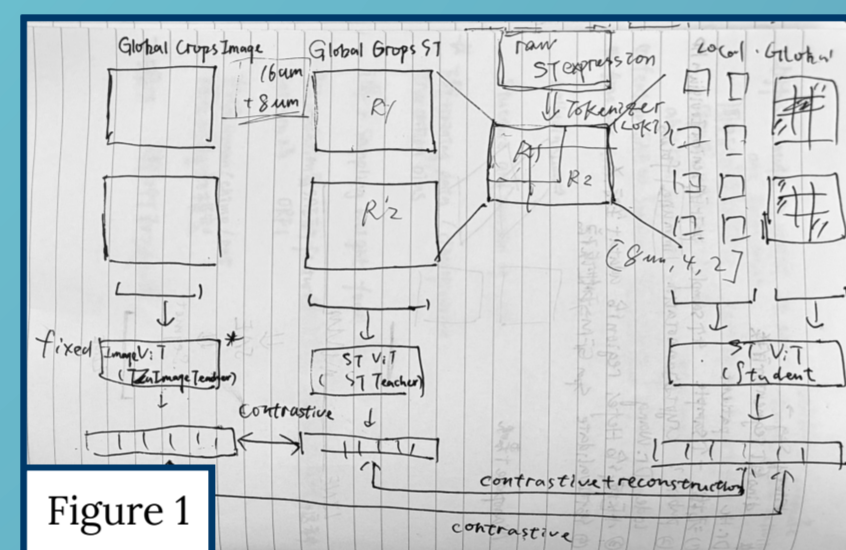


Figure 1

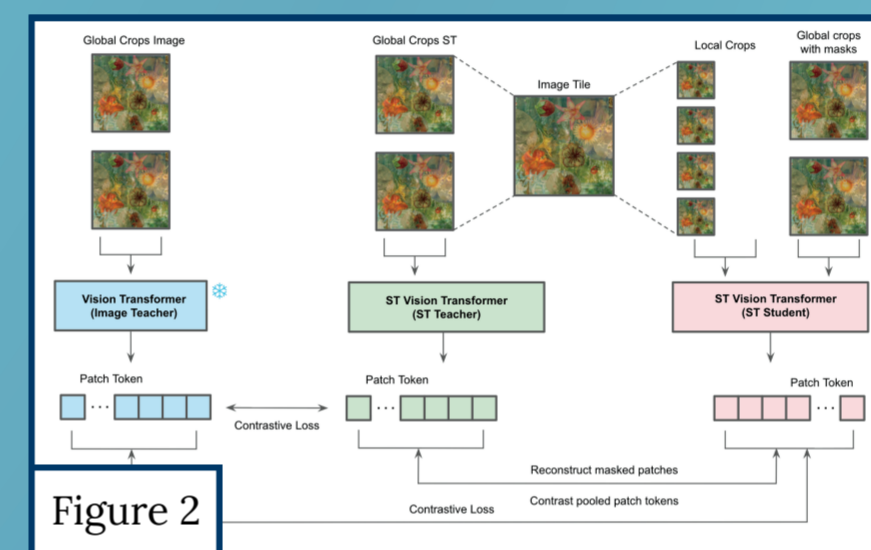


Figure 2

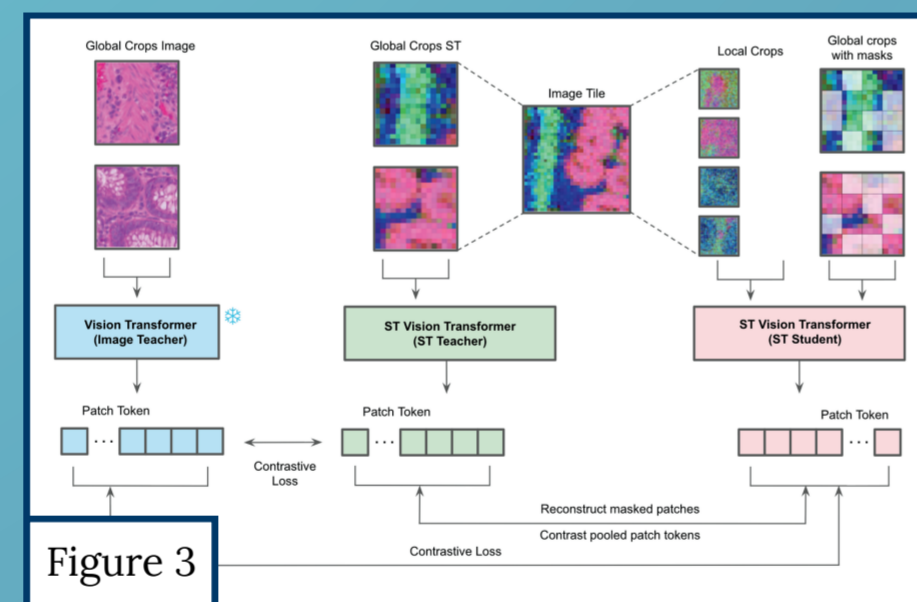


Figure 3



Figure 4

Figure 1: The hand-drawn drafts identified core components but left key relationships unclear. Data flow between crops, the fixed-weight designation, and dual training objectives were missing or insufficiently labeled. Structural organization improved after the first round of feedback, though crop resolution distinctions remained ambiguous.

Figure 2: This model marked a strong intermediate milestone, where all three component data flows were clearly defined, crop distinctions were explicit, and both training objectives were fully integrated. The remaining gap was representational authenticity, as non-biological placeholder imagery was still in use.

Figure 3: Real H&E staining slides and false-color transcriptomic overlays replaced all placeholder imagery, substantially increasing the diagram's scientific credibility. This substitution also revealed that crop thumbnails needed to be enlarged for tissue features to remain visible at publication resolution.

Figure 4: The final version added bounding boxes grouping the teacher and student modules, repositioned the Image Tile at the top center with dashed crop lines, and refined the fixed-weight snowflake icon placement. This model was ultimately submitted as a potential architecture representation for the lab's forthcoming manuscript.

Discussion

The hypothesis was confirmed: a rigorous, iterative approach to diagram development produced a figure suitable for peer-reviewed publication. The process validates expert-guided scientific visualization as an effective model for communicating complex multimodal AI architectures in computational biology. Limitations included an initial knowledge gap in deep learning that slowed early cycles, a constrained weekly review cadence, and the absence of systematic evaluation of alternative layouts. One unexpected finding emerged when replacing placeholders with real H&E data: crop thumbnails needed to be larger than originally sized for tissue features to remain visible at publication resolution, underscoring that diagrams should always be finalized with actual experimental imagery. Future work could explore interactive or animated diagrams for dynamic training loops, and the iterative methodology developed here could serve as a replicable framework for other student researchers in computational biology labs.

References

1. Arnason, Kristjan, and Heike Krebber. "Spatial Heterogeneity in the Tumor Microenvironment." PMC, National Institutes of Health, 2016, <https://pmc.ncbi.nlm.nih.gov/articles/PMC4968167/>.
2. Luo, Zhen, et al. "Spatial Transcriptomics: A New Frontier in Cancer Research." *Clinical Cancer Bulletin*, vol. 3, no. 18, 2024, <https://link.springer.com/article/10.1007/s44272-024-00018-8>.
3. "10x Genomics Commercially Launches Visium HD Spatial Gene Expression Assay." PR Newswire, 10x Genomics, 26 Mar. 2025, <https://www.prnewswire.com/news-releases/10x-genomics-commercially-launches-visium-hd-spatial-gene-expression-assay-302099533.html>.
4. Jimenez-Santos, Maria Jose, et al. "Spatial Transcriptomics in Breast Cancer Reveals Tumour Microenvironment-Driven Drug Responses and Clonal Therapeutic Heterogeneity." *bioRxiv*, 21 Feb. 2024, doi.org/10.1101/2024.02.18.580660.