

Spatial Transcriptomics-Conditioned Latent Diffusion Models for Synthetic Histopathology Tissue Patch Generation

Arnav Chaphalkar

Emerging Diagnostic and Investigative Technologies, Department of Pathology, Dartmouth Hitchcock Medical Center

ABSTRACT

- **Problem:** WSIs are essential for cancer diagnosis but are huge, hard to annotate, and current synthetic methods fail to capture tumor microenvironment (TME) features.
- **Approach:** Built a spatial transcriptomics-conditioned latent diffusion model with a new encoder (GeneCondNet) to directly map gene expression into Stable Diffusion.
- **Results/Impact:** Generated realistic colorectal cancer tissue patches (FID ~55), first to directly integrate ST into histology generation, with potential for dataset augmentation and precision oncology.

INTRODUCTION

KEY CONTEXT:

Whole Slide Images (WSI):

- Gigapixel pathology slides that preserve both cellular and tissue-level architecture
- WSIs are powerful for training machine learning and deep learning models, which transform cancer diagnosis and treatment



Fig 1: Example of a Whole Slide Image

Spatial Transcriptomics (ST):

Provides genome-wide expression data at defined tissue locations, linking molecular profiles directly to histological features.

Latent Diffusion Models (LDMs): Modern generative models that synthesize realistic images in compressed latent space, offering stability and scalability.

Problem: WSIs are gigapixel in size, costly to digitize and annotate, and require significant storage and computational resources. Publicly available datasets are limited in quantity, often biased toward certain cancer types or institutions, making large-scale use challenging.

Impact: These constraints hinder the training of machine learning and deep learning models, limiting accuracy, slowing progress in development, and reducing generalizability across diverse patient populations

Current Methods:

- Researchers have **begun** using AI to create **synthetic histology images** as a way to expand datasets and **reduce the need for manual annotations**. Most of these approaches use bulk RNA-seq or single-cell data as input.
- **Limitation:** While these inputs provide useful molecular information, they **do not preserve the spatial organization of cells in tissue**. As a result, current models often miss **critical** tumor microenvironment (TME) features such as immune infiltration, stromal remodeling, or hypoxic regions.

Hypothesis: Conditioning latent diffusion models (LDMs) on spatial transcriptomics can generate histology that is both realistic and biologically faithful.

There are currently **NO** existing studies that incorporate spatial transcriptomics into generating histopathology images

METHODS

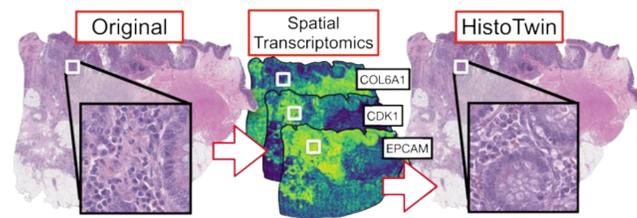


Fig 2: Conceptual Idea of ST-Conditioned WSI Generation - Image: Gokul Srinivasan

Project Goals: Advance previous work by developing an **improved** method for generating synthetic histology images that are both **visually realistic and biologically accurate**, reflecting underlying ST data and demonstrating the feasibility of ST-conditioned latent diffusion models for histopathology image generation.

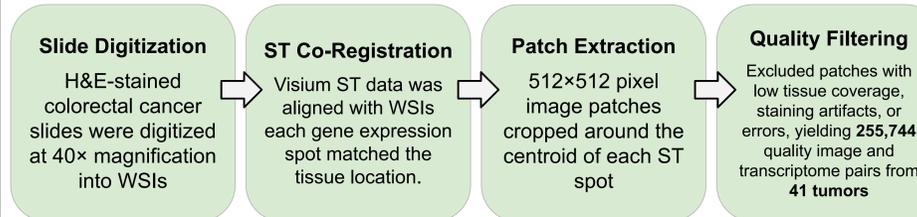
DATA

41 CRC

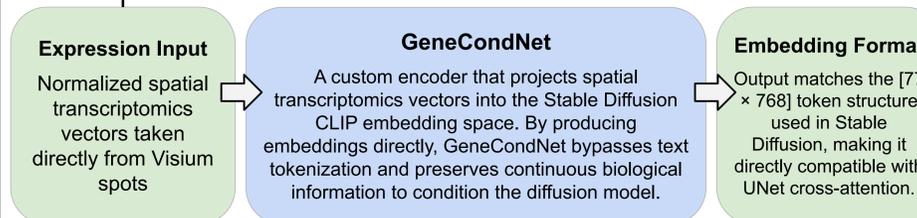
40X H&E WSI

Visium ST

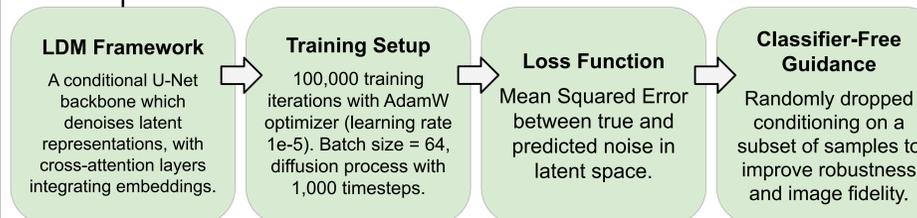
Preprocessing



Encoding Pipeline



Latent Diffusion Model



Synthetic Images Generated:

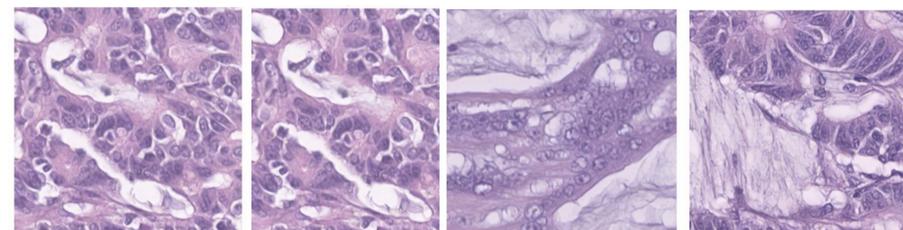


Fig 3: Synthetic Tissue Patch Images Generated by the Improved Model Workflow

RESULTS

Model Evaluation:

Previous Study: Earlier work evaluated the synthetic patches with blinded pathologists, who compared real and generated images and assessed similarity.

Current Evaluation: Due to logistical limitations, we were unable to conduct the pathologist review. Instead, we evaluated our model using a **Frechet Inception Distance (FID) score alongside qualitative visual inspection of outputs**.

Limitations of FID:

- Relies on Inception-V3, trained on natural images, not H&E slides - absolute scores may be inflated.
- Captures overall distributional similarity but ignores whether TME features (hypoxia, immune infiltration, stromal remodeling) are reproduced.

FID offers a convenient benchmark but cannot replace blinded pathologist review for assessing biological and clinical fidelity.

RESULTS:

~55

Average Frechet Inception Distance(FID)

Qualitative Assessment: Manual inspection of outputs showed that most generated patches captured tumor, stromal, immune, and adipose compartments, though occasional artifacts and unrealistic textures remained.

FUTURE WORK

- **Whole-Slide Expansion:** Extend from patch-level generation to full gigapixel WSIs, enabling direct integration into digital pathology workflows.
- **Pathologist Review:** Reintroduce blinded pathologist evaluation to validate biological fidelity and clinical relevance of generated images.
- **Broader Datasets:** Apply the framework to other cancer types and larger, more diverse cohorts to test generalizability.
- **Larger Cohorts:** Train and evaluate on a larger cohort of patient WSIs to improve robustness and statistical power.

CONCLUSION

Developed the **first** framework directly integrating spatial transcriptomics with latent diffusion models to **generate synthetic tissue patches**

- Generating realistic synthetic slides helps **overcome the scarcity of annotated pathology** data, one of the **biggest barriers** holding back AI in medicine
- Synthetic datasets **reduce bias** and bring **representation** to patient groups often missing from existing repositories, making the field more **equitable**.
- The work lays a **foundation** for AI systems that improve cancer **diagnosis, prognosis, and treatment decisions**, bringing us closer to precision oncology for **all** patients.
- **References:** <https://tinyurl.com/syntheticwsi>
- **Acknowledgements:** Dr. Khang Le, Dr. Joshua Levy, Dr. Louis Vaickus and Dartmouth College