HaDM-ST: Histology-Assisted Differential Modeling for Spatial Transcriptomics Generation

Xuepeng Liu[2,*], Zheng Jiang[2,*], Pinan Zhu[2], Hanyu Liu[2], Chao Li (Corresponding)[1]
[1]University of Cambridge, UK
[2]Northeastern University, China

I. Motivation

Background

Spatial transcriptomics (ST) is a technique that measures gene expression while preserving spatial context within tissue sections. It enables the study of how genes are expressed across different regions, but mainstream platforms suffer from coarse resolution and high costs. Although high-resolution technologies have emerged, they remain inefficient and hard to scale. Leveraging H&E histology provides a practical path to enhance ST resolution.

Challenge&Motivation

Histology-guided ST holds great promise but faces persistent obstacles. Complex tissue morphology makes it difficult to isolate expression-relevant features. Most methods lack mechanisms to capture gene-specific variability across multiple expression channels. These limitations reduce both accuracy and interpretability, motivating the development of new frameworks that can integrate histology and transcriptomics more effectively for reliable high-resolution ST.

Contributions

Main Contributions:

- **H&E-driven Semantic Distillation (HSD)** Utilize HE to combine the cell segmentation map and BERT encoding and distill predictive cues.
- **Cross-Modal Spatial Alignment (CMSA)** for pixel-level alignment via contrastive learning.
- **Gene-wise Differential Adversarial Learning (GDAL)** to capture per-gene patterns with a graph-aware discriminator.

II. Methodology

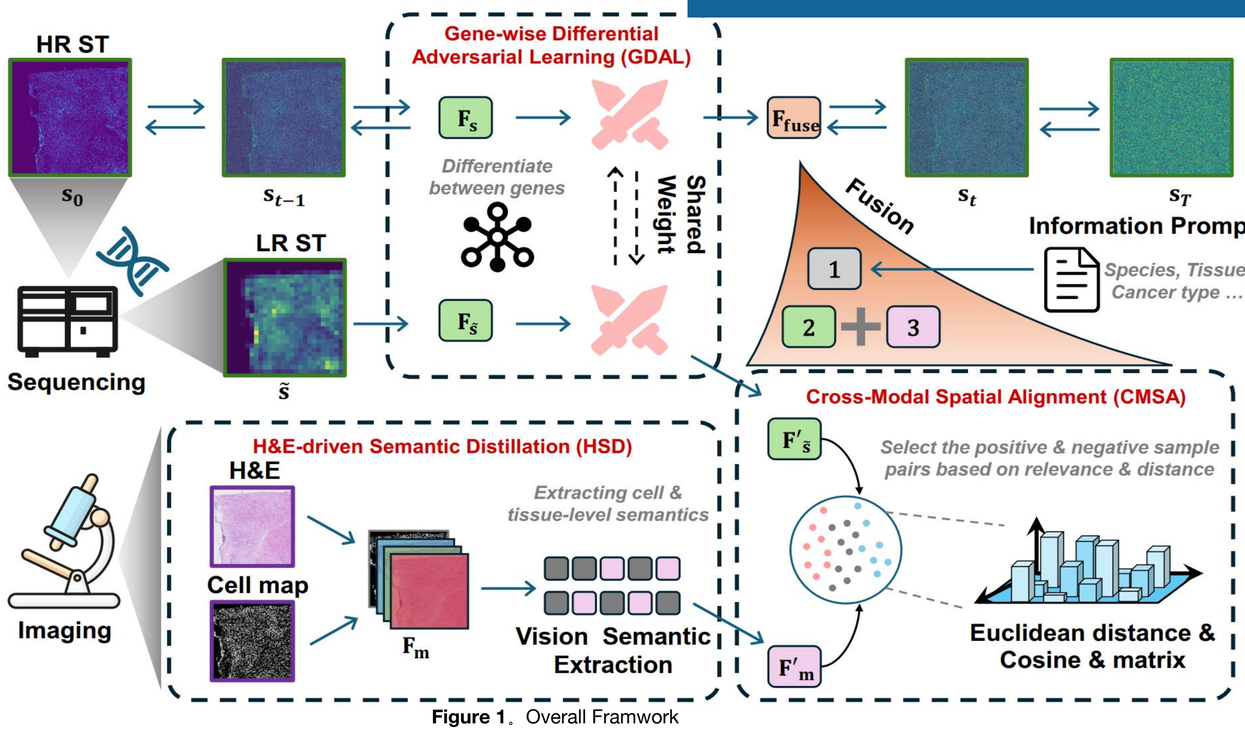

Figure 1。Overall Framwork

Overall Framework

Diffusion Framework: Denoising process conditioned on histology + LR ST.

Training: add noise to HR ST; model learns to remove noise stepwise with multimodal guidance.

Inference: supports histology-only generation if LR ST is unavailable .

HSD

HSD:H&E-Driven Semantic Distillation (HSD) fuses H&E images and ST data to bridge the semantic gap. Transformer generates features, BERT processes cancer prompts, and fusion optimizes high-resolution ST map reconstruction. H&E-ST modeling enhances generation.

CMSA

CMSA: We align H&E and LR ST features via contrastive learning. Cosine similarity and Euclidean distance are used to select positive (top 30%) and negative (bottom 30%) region pairs. A joint loss combines cosine, Euclidean, and InfoNCE terms to achieve precise spatial and semantic alignment.

GDAL & Objectives

GDAL: Channel-aware adversarial learning with a gene graph; each gene channel is a node; edges encode co-expression.

Objective:Use gcn learn per-gene distinctions while leveraging inter-gene structure; sharpen fine-grained patterns.

Losses: reconstruction + adversarial + alignment (contrastive) → stable training and biologically faithful outputs.

$$\mathcal{L}_{\text{contrast}} = \mathcal{L}_{\text{cosine}} + \lambda_1 \mathcal{L}_{\text{euclidean}} + \lambda_2 \mathcal{L}_{\text{InfoNCE}}.$$

III. Results

Experimental Setup

- **Datasets**: Xenium Mouse Brain & Human Breast; 120 unique genes per dataset; paired H&E–ST tiles.
- **Metrics**: RMSE, SSIM; higher SSIM and lower RMSE indicate better fidelity and structure.
- **HaDM-ST**: achieves SOTA on both datasets with large SSIM gains and reduced RMSE compared to TESLA, HSG, iStar.As shown in **Table 1.**

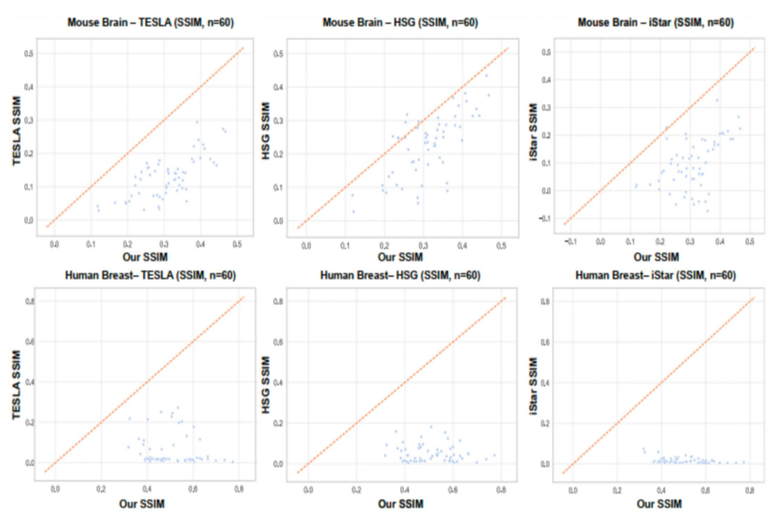

Figure 3. Gene-level SSIM scatter plots comparing HaDM-ST with TESLA, HSG, and iStar.

Key Fingdings

Spatial Alignment: we generated local SSIM-based alignment maps. As shown in **Figure 2**, HaDM-ST exhibits predominantly green regions , indicating strong spatial concordance with H&E histology.

Gene-Level Structural Recovery:We further analyzed gene-level performance using SSIM scatter plots. As illustrated in **Figure 3**

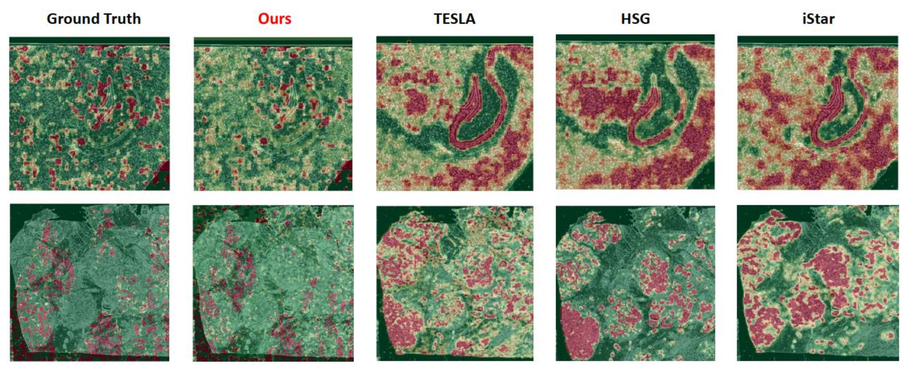

Figure 2. Local SSIM-based alignment maps, where green denotes high concordance.

Qualitative Insights

- **Qualitative**: As shown in **Figure 4**.HaDM-ST restores sharper edges and finer cellular structures; hotspots match ground truth.
- **Generalization**: robust across tissues and genes; maintains high fidelity in diverse morphology.
- **Utility**: enables high-confidence in silico HR ST for downstream spatial analyses and precision medicine.

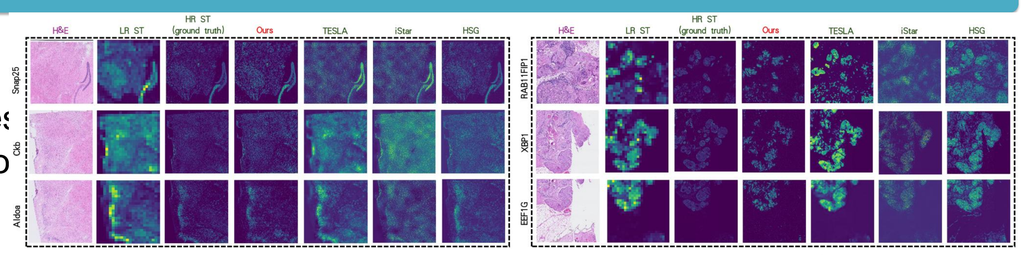

Figure 4. Visual comparison of restored gene expression maps, where HaDM-ST achieves superior fidelity.

Approach	RMSE	SSIM
TESLA	0.2489	0.1373
iStar	0.3088	0.0995
HSG	0.2000	0.2648
Ours	**0.1630**	**0.3184**

(a) Mouse brain

Approach	RMSE	SSIM
TESLA	0.3302	0.0655
iStar	0.3071	0.0486
HSG	0.2832	0.0533
Ours	**0.2304**	**0.4663**

(b) Human breast

Table1. Quantitative comparison of RMSE and SSIM across two datasets.

IV. Conclusions & Acknowledgments

Conclusions

- We introduced HaDM-ST, a diffusion-based cross-modal framework for generating HR ST maps from H&E + LR ST.HSD, CMSA, and GDAL jointly address semantic extraction, precise alignment, and gene-wise modeling.
- HaDM-ST delivers superior accuracy and structural fidelity, providing reliable high-resolution data for ST research.

Contact

- **Contact: cl647@cam.ac.uk**