Artificial Intelligence for Spatial Transcriptomics: A Scoping Review of Architectures, Datasets, and Models

Agampreet Saini
School of Computer Science
UPES Dehradun
Dehradun, India-248007
Agampreet.107828@stu.upes.ac.in

Supragya Gandotra
School of Computer Science
UPES Dehradun
Dehradun, India-248007
Supragya.107073@stu.upes.ac.in

Abhijit Kumar
School of Computer Science
UPES Dehradun
Dehradun, India-248007
abhijit.kumar@ddn.upes.ac.in

Pratik Dutta*
Department of Biomedical Informatics,
Stony Brook University
pratik.dutta@stonybrook.edu

Tirthankar Ghosal*
Oak Ridge National Laboratory
Oak Ridge, TN, USA
ghosalt@ornl.gov

Abstract

Learning meaningful representations from multimodal spatial transcriptomics data, which integrates histology images with gene expressions, is a fundamental challenge in biological vision. Spatial transcriptomics provides this data, enabling the mapping of the transcriptome onto tissue sections. This paper presents a thorough survey of representation learning in spatial omics, critically comparing over 40 deep learning frameworks. We group these models by the core tasks their learned representations are designed to solve: cell type deconvolution, spatial domain identification, gene expression imputation, 3D tissue reconstruction, and cell-cell interaction simulation. Special attention is given to the dominant architectures for representation learning in this domain, including graph neural networks, contrastive learning, and multimodal fusion methods. We evaluate representative models such as ADCL, CellMirror, and MuST for the scalability, interpretability, and biological impact of their learned embeddings. The survey also addresses common challenges that hinder representation learning, including spatial noise, modality imbalance, and low-resolution data. Finally, we outline future directions centered on building foundation models for spatial biology and improving 3D alignment. This review provides a critical guide for researchers developing foundational and task-specific representations from multimodal spatial data.

1 Introduction

Spatial transcriptomics (ST) has emerged as a transformative modality by integrating histology images with spatially resolved gene expression profiles, an achievement recognized when it was

*Jointly supervised this work.

The 3rd Workshop on Imageomics: Discovering Biological Knowledge from Images Using AI,39th Conference on Neural Information Processing Systems (NeurIPS 2025).

named *Nature Methods'* "Method of the Year" in 2020. This multimodal view enables researchers to link tissue morphology with molecular states, advancing our understanding of development, disease, and therapeutic response. For computational biology and computer vision, ST poses a unique set of challenges and opportunities. Unlike natural image tasks, spatial omics data are multimodal (images, gene counts, spatial coordinates), heterogeneous, and often limited in scale. Extracting meaningful representations that capture both morphological structure and molecular context is therefore a fundamental challenge.

The rapid evolution of this field has been driven by artificial intelligence. In just a few years, the computational landscape has matured from initial clustering algorithms to sophisticated architectures for representation learning, including graph neural networks (GNNs), contrastive learning, and multimodal transformers. Early landmark tools like SpaGCN began leveraging GNNs to integrate modalities, while others like Tangram used deep learning to register single-cell data onto spatial maps. The timeline in Figure S1 illustrates this accelerated development, highlighting the key architectural shifts that have defined the field.

In this paper, we survey recent advances in representation learning for spatial transcriptomics, with an emphasis on key modeling approaches across core tasks like cell type deconvolution, domain identification, and 3D reconstruction. We also examine the computational challenges that arise from multimodal data and discuss future opportunities to build foundation models that generalize across tissues and scales. Our goal is to provide a critical guide for researchers in the Imageomics community developing AI-driven methods to create interpretable and clinically relevant biological vision systems.

2 A Survey of Representation Learning Models

The field of AI-enhanced spatial transcriptomics can be organized into six core tasks, each addressing a key challenge in representation learning. These tasks range from resolving cellular composition within low-resolution spots to reconstructing entire 3D tissue architectures. Our full analysis is based on a comprehensive review of over 40 frameworks. Table 1 provides a comprehensive, side-by-side comparison of over 20 representative models, detailing the specific techniques, data modalities, and frameworks for each.

2.1 Cell Type Deconvolution

Deconvolution methodologies aim to computationally unmix the gene expression signals from heterogeneous cell types within a single ST spot.

- ADCL Zhang and Zhang [2023]: Integrates multi-head graph attention networks (MHGAT) and variational autoencoders (VAE) with dual-contrastive learning, outperforming previous models on resolution and accuracy.
- MHDGAT Chen et al. [2024b]: Deploys a multi-head dynamic GAT network fused with optimal transport, efficiently leveraging both spatial and transcriptional information and utilizing cell type labels.
- MACD Huang et al. [2024]: Combines masked autoencoders and adversarial learning to align latent representations from simulated and real data, achieving robust inference even under domain shift.
- CellMirror Xia et al. [2023] : Applies interpretable spatial graph-based contrastive learning with single-cell RNA-seq references, resolving finer cell-type subpopulations in mixed ST spots.
- TransST Liu et al. [2025]: Uses transfer learning from reference datasets with a Markov random field prior to boost segmentation and denoising in low-resolution or noisy spatial data.

2.2 Spatial Domain Identification

These models segment tissues into spatially coherent domains by learning representations that integrate gene expression, spatial coordinates, and sometimes histology.

Table 1: Survey of the rapidly evolving landscape of representation learning models for spatial transcriptomics, a key challenge in biological vision. The table compares over 40 frameworks from 2023-2025 across core tasks such as cell type deconvolution, 3D reconstruction, and multimodal fusion. We summarize the primary techniques, input data modalities, and computational frameworks to provide a benchmark for developing the next generation of foundational models. Abbreviations: **GNN** (Graph Neural Network), **VAE** (Variational Autoencoder), **GCN** (Graph Convolutional Network), **OT** (Optimal Transport), **LLM** (Large Language Model).

Method (Year)	Type	Technique Used	Input Modality	Resolution	Framework
SPACEL (2023)	Deep learning integrated framework	MLP + Probabilistic model (Spoint), GCN + adversarial training (Splane), differential evolution (Scube)	Expression (seq-based ST); multi-slice	Visium (~55–100 μm) to single-cell (~1–2 μm) (e.g. MERFISH/Stereo-seq)	MLP, GCN, adversarial network, evolutionary algorithm
SpaGCAC (2024)	Graph DL – spatial domain identification	GAT (graph attention) with adaptive feature balancing, contrastive loss	Expression + Coordinates	Visium (~55 μm)	GNN (Graph Attention Network)
TME-DeNoise (2023)	Semi-supervised learning – TME	Variational graph autoencoder with contrastive pretraining	Expression + Histology	Visium (~55 μm)	GNN (VGAE with contrastive loss)
MAFN (2023)	Deep multi-view clustering	View-specific encoders, attention-based fusion of expression + image features, joint optimization	Expression + H&E images	Visium (~55 μm)	Multi-branch CNN + fusion network
stBERT (2023)	Transformer-based transfer learning	BERT-like pretraining on ST slices, fine-tuning for domain classification	Expression (Visium slices)	Visium (~55 μm)	Transformer (pretrained on spatial data)
UPSST (2023)	Unsupervised GNN – tissue domain ID	Hierarchical graph autoencoder with adaptive clustering and reconstruction	Expression + Coordinates	Visium (~55 μm)	GNN (graph autoencoder + clustering)
ADCL (2023)	Deep deconvolution (contrastive)	Dual-contrastive framework (spot vs. pseudo-spot) using VAE architecture	Expression (spots)	Visium (~55 μm)	VAE + contrastive learning
STADS (2023)	Spatial drug repurposing (framework)	PREDICT algorithm on cell-type signatures; graph model on L1000 profiles + ST spots	Expression + drug targets + cell profiles	Visium (~55 μm)	GNN + matrix factorization
SpaCCC (2024)	LLM-based multi-modal communication	Large language model for cell-cell dialogue simulation, graph-based feature fusion	Expression + scRNA-seq + cell atlas	Various (e.g. Slide-seqV2 ~10 μm)	Transformer (LLM) + Graph Neural Network
MHDGATOT (2024)	GNN + optimal transport deconvolution	Multi-head dynamic GAT to build cell graph, fused with Gromov–Wasserstein OT	Expression + Coordinates + scRNA-seq	Visium (~55 μm)	GNN (multi-head) + optimal transport
HisToSGE (2024)	Gene expression prediction (ST→Histology)	CNN with spatial transformer on H&E to predict missing gene expression	H&E image + spot coordinates	Histology resolution (subcellular)	CNN + Spatial Transformer
MACD (2024)	Adversarial NN deconvolution	Two-stage CNN: cell type classifier followed by GAN to refine proportions	Simulated & real ST spots	Visium (~55 μm)	CNN + GAN
stMCDI (2024)	Imputation (conditional diffusion)	Masked gene modeling with spatial encoding, conditioned diffusion U-Net	Expression + Coordinates	Visium (~55 μm)	Conditional diffusion model + GNN
STG3Net (2024)	Spatial domain ID & batch correction	Encoder–decoder with clustering embedding, uses SCE + triplet loss for consistency	Expression (batch-integrated slices)	Visium (~55 μm)	MLP + clustering loss
SGAEC (2023)	GNN + VAE + contrastive learning	Spatial GNN autoencoder with contrastive and KL losses for clustering	Expression + Coordinates + Histology	Visium (~55 μm)	GNN (VGAE) + contrastive learning
MVCLST (2024)	Multi-view graph contrastive learning	Graph encoders on co-expression and spatial graphs, adaptive attention for modality weighting	Expression + multiple views	Visium (~55 μm)	GNNs with multi-loss (reconstruction, contrastive)
spGCLF (2024)	Deep graph contrastive learning	Two GNN views (expression vs. spatial graph), DGCNN, with denoising contrastive loss	Expression + Coordinates	Visium (~55 μm)	GNN (denoising contrastive)
SSGCN (2024)	GNN with adaptive reweighting	Multi-scale GCN with learnable reweighting of graph edges, clustering via MCL	Expression + Coordinates	Visium (~55 μm)	GNN (adaptive GCN) + spectral clustering
DGT (2024)	Domain-guided transfer clustering	Optimize label distributions with Wasserstein and self-training loss across domains	Expression + multiple ST datasets	Visium (~55 μm)	MLP + optimal transport
RepGraph (2024)	Representational graph learning	Graph clustering in latent space via Wasserstein distance; integrates histology graph	Expression + H&E image	Visium (~55 μm)	GNN with OT clustering
STCC (2024)	Consensus clustering (STCC)	Rotation-invariant smoothing + ensemble clustering to enhance domain detection	Expression + Coordinates (Visium/ST)	Visium (~55 μm)	Ensemble + graph smoothing
TransST (2025)	Transfer learning spatial factorization	Zero-shot learning: train on labelled ST to segment unlabeled images via MRF	Histology images + pretrained model	Histology (subcellular)	CNN (U-net) + transfer via MRF
FaST (2025)	Spatial Transcriptomics Pipeline (FaST)	Multi-step: fast alignment and counting (STAR), segmentation, integration (Scanpy)	Raw Visium sequencing data	Visium (~55 μm)	Software pipeline (Python tools)

- GCNCL Liang et al. [2024]: Adapts feature-spatial balances for each sample using GCNs, improving robustness in domain detection.
- stBERT Wang et al. [2024]: Brings transformer-based masked modeling and transfer learning to spatial contexts, achieving state-of-the-art tissue boundary delineation.
- vGraphST Li et al. [2023]: Blends graph variational autoencoders with contrastive learning for simultaneous clustering and denoising.
- spGCLF Chen et al. [2024a]: Applies contrastive learning with dual denoising for noisy data and cross-slice transfer.
- SSGCN Du et al. [2024]: A multi-scale GCN with adaptive reweighting for domain segmentation at various tissue resolutions.
- HexCNN Gao et al. [2022]: Employs hexagonal convolutional kernels to better model regular spatial spot grids, improving domain classification and noise resilience.

2.3 Gene Expression Imputation and Prediction

Methods in this category either fill in missing gene expression values (imputation) or predict expression patterns directly from histology images (prediction).

- CC + O2U Leng et al. [2023]: Use GNNs and semi-supervised error-resistant modules for improved tumor microenvironment analysis.
- **stMCDI** Li et al. [2024] is a state-of-the-art imputation framework that uses a masked conditional diffusion model, guided by a GNN encoder, to recover missing gene values while preserving the original data distribution.
- **HisToSGE** Shi et al. [2024] predicts high-resolution spatial gene expression directly from H&E histology images using a large-scale pretrained pathology model and a multi-head attention module to integrate spatial coordinates.
- BG-TRIPLEX Qu et al. [2024], EGDNN Yang et al. [2023], and STFormer Zhan et al. [2024] combine advanced attention and exemplar learning to improve gene prediction from imaging data.

2.4 3D Tissue Reconstruction

These frameworks focus on aligning and integrating multiple 2D ST slices to create cohesive 3D tissue atlases.

- **SPACEL** Xu et al. [2023] is a modular deep learning library featuring components for deconvolution (Spoint), 2D domain identification (Splane), and automated 3D reconstruction (Scube) that align and stack slices into a continuous 3D model.
- **OptiGraph3D** Zhang et al. [2024] is a hybrid framework that integrates pivot-driven registration to align slices, an ADMM-based solver for deconvolution, and a multi-head graph attention network to reconstruct the 3D tissue structure.
- **STG3Net** Fang et al. [2024]: Uses masked graph autoencoders and anchor-based correction for robust batch integration across multiple slices or samples, enhancing accuracy and consistency in composite tissue atlases.

2.5 Cell-Cell Interaction Inference

These models aim to decode the complex signaling networks and interactions between cells in their spatial context.

- **SpaCCC** Ji et al. [2024] is a novel framework that leverages Large Language Models (LLMs) to infer cell-cell communication by encoding cell types and their spatial context into natural language prompts, enabling zero-shot, knowledge-enhanced interaction prediction.
- **Hypergraph Wavelets** Sun et al. [2025] generalize standard graphs to hypergraphs to model higher-order cellular niches. It uses hypergraph diffusion wavelets to capture multi-scale spatial patterns, revealing distinct cellular communities associated with disease progression.

2.6 Drug Discovery and Target Identification

This emerging area uses spatially resolved transcriptomic maps to guide therapeutic strategies.

- **STADS** Karaaslanli et al. [2023] is a graph-based framework that integrates ST data with pharmacogenomic evidence from the L1000 Connectivity Map to rank and recommend drugs for repositioning.
- **Radiopharmaceutical Modeling** Hong et al. [2023] integrates ST data with pharmacokinetic and dosimetry modeling (solving PDEs) to enable *in silico* screening of radiopharmaceutical therapies and prioritize targets.

2.7 A unified evaluation and taxonomy framework

To move beyond a descriptive listing, we outline a compact task-centric framework that ties together all surveyed models. As illustrated in the AI-powered pipeline in supplemental material Figure S2, the taxonomy links each stage—from deconvolution to interaction modeling—to its evaluation axis: standard task metrics, dataset diversity, and robustness probes. This structure provides a simple, reproducible way to compare methods and forms the basis for the small meta-analysis summarized in the supplemental material.

To provide a compact comparative view, we conducted a brief meta-analysis of five representative cell-type deconvolution models (2023–2025) focusing on their architectural innovations, performance highlights, and interpretability trade-offs. Details are summarized in supplemental material Table S2.

3 Challenges and Open Problems

Despite rapid progress, the clinical and biological translation of AI-powered spatial transcriptomics remains hindered by several interlocking challenges. A central issue is the trade-off between **spatial resolution and molecular depth**: sequencing-based ST aggregates multiple cells per spot, while imaging-based approaches achieve subcellular resolution at the cost of limited gene panels and technical complexity. On the computational side, many multimodal fusion frameworks exhibit **modality bias**, where transcriptional signals dominate and underutilize morphological or spatial features. Progress is also slowed by **data scarcity**: Most public ST datasets are small, heterogeneous, and organ-specific, making overfitting and poor generalization common; a summary of commonly used datasets is provided in Table S1 in the supplementary material.

Tissue heterogeneity and batch effects further complicate the definition of consistent, biologically meaningful domains, while multi-slice registration and 3D reconstruction remain technically daunting. Beyond performance, issues of **interpretability and validation** are particularly pressing: state-of-the-art GNNs, transformers, and contrastive frameworks often act as "black boxes," offering limited biological explainability. At the same time, the absence of gold-standard annotated datasets forces reliance on proxies such as scRNA-seq overlap or histology concordance, introducing subjectivity and limiting reproducibility. Finally, as spatial data integrate genomics, histology and metadata, **privacy, ethics, and regulatory compliance** emerge as critical but unresolved obstacles, with no clear guidelines defined yet for AI-based spatial omics technologies.

Furthermore, bridging research models with clinical use requires validation beyond benchmarking. We suggest two practical evaluation tracks. First, *retrospective concordance studies*, using cohorts with matched ST and orthogonal assays (IHC, IF, or scRNA-seq), can measure agreement between model output, pathologist annotations, and molecular ground truth, identifying cases where AI predictions could refine diagnostic interpretation. Second, *operational reproducibility tests* should examine model stability across sample preparation methods and scanners, with clear calibration steps for deployment. Together, these tracks provide a feasible blueprint for future real-world validation and can guide laboratory collaborations seeking to assess translational readiness.

In addition, interpretability is vital for clinical trust. Common strategies include attention maps that show model focus on histological regions, saliency or gradient methods (Grad-CAM, Integrated Gradients) to reveal key image features, and GNN explainers that highlight influential cells or spatial edges. Concept based or counterfactual tests link human-readable traits to predictions. Combining these approaches provides a practical baseline for transparent and verifiable spatial models.

4 Future Directions

The field is advancing toward a new phase of scalable, generalizable, and clinically viable spatial intelligence, with several transformative trends on the horizon. First, the development of **foundation models for spatial biology**, large transformer-based encoders pretrained on diverse tissues and modalities, promises universal backbones for downstream tasks, enabling zero-shot and few-shot transfer learning. Second, the rise of **multi-omic integration** will allow ST to be co-analyzed with scRNA-seq, ATAC-seq, proteomics, and histology, using graph learning and contrastive objectives to capture regulatory circuits across scales. Third, **LLM-augmented discovery** is emerging, with systems like SpaCCC showing how large language models can perform zero-shot inference of cell–cell interactions and drug repurposing, accelerating translational insights directly from spatial data.

Beyond modeling, **privacy-preserving federated learning** pipelines will allow hospitals and labs to benefit from shared model updates without moving sensitive patient data. Finally, the paradigm of **human–AI co-analysis** is gaining traction: interactive visualization platforms and collaborative annotation systems will empower experts and machine intelligence to jointly interpret tissue architectures. Collectively, these directions point to a future where AI-driven ST moves beyond experimental research and becomes a cornerstone of precision medicine and Imageomics, turning multimodal biological images into interpretable, trustworthy, and clinically actionable knowledge.

Building on the unified pipeline outlined in Figure S2, these directions can be translated into concrete research roadmaps that guide community progress. In the near term (0–3 years), the focus should be on creating open, multi-platform pretraining corpora and standardized benchmarks that unify data formats and evaluation metrics. Mid-term goals (3–7 years) include federated infrastructures that enable secure cross-institutional learning and transparent leaderboards, while the longer horizon emphasizes clinically grounded validation studies where AI-driven predictions are prospectively reviewed by pathologists. These milestones convert the field's high-level ambitions into practical, measurable deliverables for the spatial transcriptomics community.

5 Conclusion

The combination of deep learning and spatial transcriptomics with modern AI has revolutionized our ability to decode tissue architecture with unprecedented molecular and spatial precision. Computational models now enable critical advances such as **cell-type deconvolution**, **spatial domain identification**, **3D tissue reconstruction**, and **cell–cell interaction modeling**. Yet major challenges remain, including **limited resolution**, **data scarcity**, **modality imbalance**, **lack of interpretability**, and unresolved issues of **privacy and regulation**. Addressing these barriers will require the development of **scalable, explainable, and multimodal frameworks**, supported by large annotated datasets and rigorous validation pipelines. The field is now shifting towards **foundation models** and translational use cases that bridge computational innovation with clinical practice. Success will depend on deep collaboration between computational, biological, and clinical communities, ensuring that AI-driven spatial transcriptomics evolves into a scalable, trustworthy, and clinically impactful system. This trajectory aligns with the broader Imageomics vision of discovering biological knowledge from images using AI. This is a short survey covering only the most essential and representative works in the area. Due to page limitations in the workshop, several relevant studies could not be included; we plan to extend this into a more exhaustive version in a forthcoming paper.

6 Acknowledgment

The UPES authors would like to thank the Center of Excellence-Artificial Intelligence, School of Computer Science, UPES Dehradun for supporting this research. Authors, Dutta and Ghosal, provided guidance to this line of research to the team.

References

Xiang Chen, Junnan Yu, and Min Li. spgclf: a versatile deep graph contrastive learning framework for spatial transcriptomics analysis. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 497–502. IEEE, 2024a.

Yuqi Chen, Feng Yang, Peng Jiang, Yuan Wang, Qiang Zhang, and Juan Liu. Deconvolution of spatial transcriptomics data based on multi-head dynamic gat and optimal transport. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 609–612. IEEE, 2024b.

Wenhui Du, Chunyang Chai, Yujie Gu, Shuang Hao, and Huiyan Sun. Ssgcn: A spatial domain identification framework based on graph convolutional neural network with adaptively reweighting edges. In *2024 12th International Conference on Bioinformatics and Computational Biology (ICBCB)*, pages 19–26. IEEE, 2024.

Donghai Fang, Fangfang Zhu, and Wenwen Min. Multi-slice spatial transcriptomics data integration analysis with stg3net. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 509–514. IEEE, 2024.

Jing Gao, Fa Zhang, Kai Hu, and Xuefeng Cui. Hexagonal convolutional neural network for spatial transcriptomics classification. In *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 200–205. IEEE, 2022.

J Hong, S Bae, L Cavinato, K Erlandsson, H Choi, and K Shi. Pharmacokinetic and dosimetry modeling of radiopharmaceutical therapy using spatial transcriptomics in microenvironment. In *2023 IEEE Nuclear Science Symposium, Medical Imaging Conference and International Symposium on Room-Temperature Semiconductor Detectors (NSS MIC RTSD)*, pages 1–1. IEEE, 2023.

Lin Huang, Xiaofei Liu, Shunfang Wang, and Wenwen Min. Masked adversarial neural network for cell type deconvolution in spatial transcriptomics. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 619–622. IEEE, 2024.

Boya Ji, Liwen Xu, and Shaoliang Peng. Spaccc: Large language model-based cell-cell communication inference for spatially resolved transcriptomic data. *bioRxiv*, pages 2024–02, 2024.

Abdullah Karaaslanli, Yijun Guo, Haodong Liang, Ziyin Huang, and Lana X Garmire. Stads: Spatial transcriptomics to aid drug-reposition recommendation. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2737–2740. IEEE, 2023.

Jiake Leng, Yiyan Zhang, Xiang Liu, Yiming Cui, Junhan Zhao, and Yongxin Ge. Error-robust and label-efficient deep learning for understanding tumor microenvironment from spatial transcriptomics. *IEEE Transactions on Circuits and Systems for Video Technology*, 2023.

Cuiyuan Li, Kai Hu, Fa Zhang, and Xuefeng Cui. Variational clustering and denoising of spatial transcriptomics. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 281–286. IEEE, 2023.

Xiaoyu Li, Wenwen Min, Shunfang Wang, Changmiao Wang, and Taosheng Xu. Masked conditional diffusion model with gnn for spatial transcriptomics data imputation. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2164–2169. IEEE, 2024.

Xuena Liang, Junliang Shang, Jin-Xing Liu, Chun-Hou Zheng, and Juan Wang. Enhancing spatial domain identification in spatially resolved transcriptomics using graph convolutional networks with adaptively feature-spatial balance and contrastive learning. *IEEE/ACM Transactions on Computational Biology and Bioinformatics*, 2024.

Shuo Shuo Liu, Shikun Wang, Yuxuan Chen, Anil K Rustgi, Ming Yuan, and Jianhua Hu. Transst: Transfer learning embedded spatial factor modeling of spatial transcriptomics data. *arXiv preprint arXiv:2504.12353*, 2025.

Mingcheng Qu, Yuncong Wu, Donglin Di, Anyang Su, Tonghua Su, Yang Song, and Lei Fan. Boundary-guided learning for gene expression prediction in spatial transcriptomics. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 445–450. IEEE, 2024.

Zhiceng Shi, Shuailin Xue, Fangfang Zhu, and Wenwen Min. High-resolution spatial transcriptomics from histology images using histosge. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 2402–2407. IEEE, 2024.

Xingzhi Sun, Charles Xu, João F Rocha, Chen Liu, Benjamin Hollander-Bodie, Laney Goldman, Marcello DiStasio, Michael Perlmutter, and Smita Krishnaswamy. Hyperedge representations with hypergraph wavelets: applications to spatial transcriptomics. In *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 1–5. IEEE, 2025.

Ziheng Wang, Xinhe Li, Hiroaki Takada, and Ryoichi Nagatomi. stbert: a pretrained model for spatial domain identification of spatial transcriptomics. *IEEE Access*, 2024.

Junjie Xia, Jiangnan Cui, Zhenyu Huang, Shuangquan Zhang, Fang Yao, Yijian Zhang, and Chunman Zuo. Cellmirror: deciphering cell populations from spatial transcriptomics data by interpretable contrastive learning. In *2023 IEEE International Conference on Medical Artificial Intelligence (MedAI)*, pages 165–176. IEEE, 2023.

Hao Xu, Shuyan Wang, Minghao Fang, Songwen Luo, Chunpeng Chen, Siyuan Wan, Rirui Wang, Meifang Tang, Tian Xue, Bin Li, et al. Spacel: deep learning-based characterization of spatial transcriptome architectures. *Nature Communications*, 14(1):7603, 2023.

Yan Yang, Md Zakir Hossain, Eric A Stone, and Shafin Rahman. Exemplar guided deep neural network for spatial transcriptomics analysis of gene expression prediction. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5039–5048, 2023.

Gan Zhan, Xiuju Du, Jing Liu, Yinhao Li, Lanfen Lin, Jingsong Li, and Yen-Wei Chen. Stformer: Learning to explore spot relationships for spatial transcriptomics prediction from histology of colorectal cancer. In *2024 46th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 1–4. IEEE, 2024.

Shilin Zhang and Qingchen Zhang. Adcl: an adaptive dual contrastive learning framework based on mhgat and vae for cell type deconvolution in spatial transcriptomics. In *2023 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 355–358. IEEE, 2023.

Zhengqian Zhang, Binghong Chen, Jialiang Wang, Haiyu Wang, Tianshuo Yu, Junjun Ren, Mingtao Zhang, Chongyao Liu, Wei Jiang, and Yongzhuang Liu. Robust 3d reconstruction and cell-type deconvolution approach for spatial transcriptomics by hybrid graph learning and iterative computation. In *2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, pages 1752–1757. IEEE, 2024.

A Supplemental material

A.1 Images

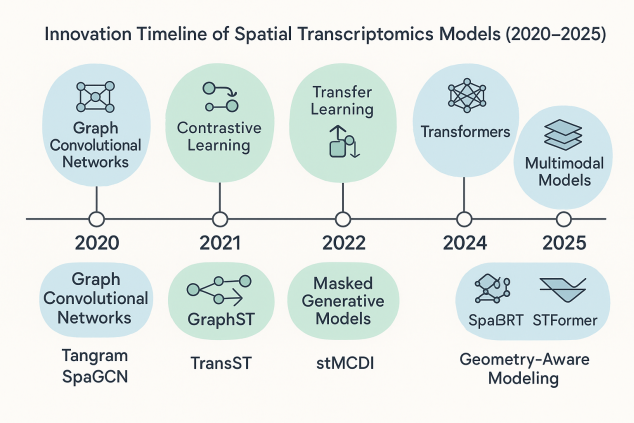

Figure S1: Timeline of key innovations in spatial transcriptomics models from 2020-2025, showing the progression from graph-based methods to transfer learning, transformers, and multimodal architectures.

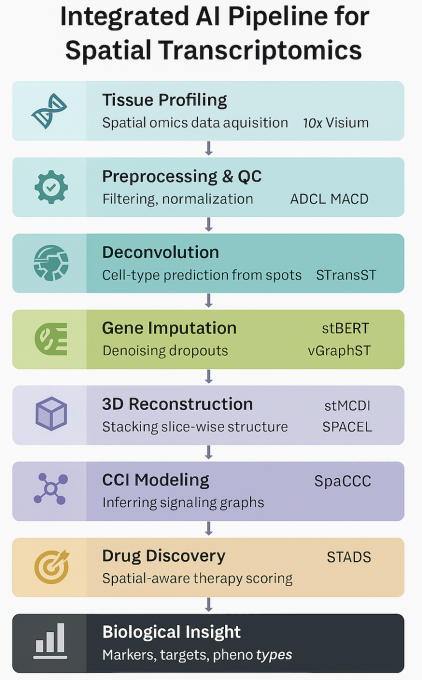

Figure S2: The AI-powered pipeline for spatial transcriptomics. This diagram illustrates the full lifecycle from tissue profiling to biological insight, highlighting where specific deep learning models intervene in preprocessing, deconvolution, domain identification, 3D reconstruction, interaction modeling, and drug targeting.

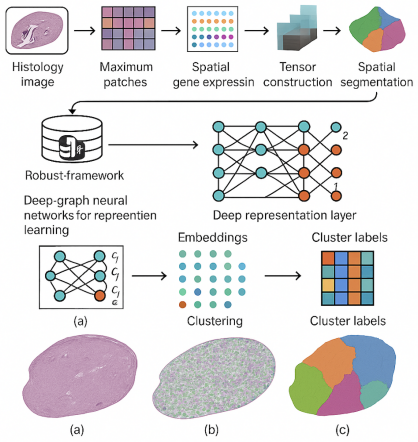

Figure S3: An overview of how histology images and gene expression data are combined in spatial transcriptomics. The process involves dividing tissue images into patches, linking them with spatial gene expression, constructing tensors, and segmenting regions. A deep graph neural network is then applied to learn representations, which are clustered to reveal spatial domains within the tissue.

A.2 Datasets

The table below summarizes the key datasets used by these models, including their sources, sample composition, and usage.

Table S1: Summary of spatial transcriptomics datasets used in benchmarking and modeling across recent studies.

Dataset Name	Data Source / Accession	Sample Size and Tissue	Benchmarking Use	Technical Details
Prostate Cancer (Visium) – PC1 (acinar) & PC2/PC3 (adenocarcinoma)	10x Genomics Visium (PC1); Mendeley Data	3 slides (2 patients; primary prostate tumors)	Used in Hong et al. (2024) for RPT dosimetry modeling	10x Visium (~55 µm spots; ~3–5K spots/slice)
Human DLPFC (Visium)	Public 10x Genomics dataset (Lieber Institute)	12 serial cortical slices (adult human prefrontal cortex)	Used by Fang et al. and Zhang et al. for multi-slice integration / 3D reconstruction	10x Visium (~55 µm hex grid; ~3K spots/slice)
Human embryonic heart (Stereo-seq)	Published dataset (9-week human embryo)	Multiple sections (embryonic heart)	Used by Zhang et al. for 3D tissue reconstruction	BGI Stereo-seq (high-res ~0.5 µm spots)
Adult mouse brain (Stereo-seq)	Published dataset	35 sections (adult mouse brain)	Used by Fang et al. for multi-slice integration	Stereo-seq (cellular resolution)
Mouse embryo brain (Stereo-seq)	Published dataset (E9.5–E11.5 mouse)	3 sections (embryonic brain)	Used by Fang et al. for multi-slice integration	Stereo-seq (cellular resolution)
Various (Visium; SlideseqV2; Stereo-seq)	Mixed public datasets	Human/mouse tissues (multiple studies)	Used by Zang et al. to evaluate MuST integration	Visium; Slide-seqV2; Stereo-seq platforms
CRC Charting cohort (Visium)	Internal (Zhan et al. 2024)	Colorectal cancer resections (histology + ST)	Used to train/test STFormer	10x Visium with H&E histology
CRC Intestine cohort (Visium)	External (Zhan et al. 2024)	Colorectal cancer resections	Used as external test set for STFormer	10x Visium with H&E histology

A.3 Meta-analysis of Representative Deconvolution Models

Table S2 provides a concise meta-analysis comparing five leading spatial transcriptomics deconvolution frameworks. Each entry synthesizes its core methodological idea, major strengths, improvements over prior work, and observed limitations as reported in recent literature (2023–2025). This table complements the main survey by highlighting how successive models balance interpretability, robustness to noise, and generalization across heterogeneous tissues.

Table S2: Meta-analysis of representative spatial transcriptomics deconvolution models (2023–2025). The table summarizes each model's core idea, reported strengths, comparative improvements, and practical limitations.

Model	Core Idea	Key Strength	Improvement Over Prior	Pros	Cons
Cell2location / GraphST (baseline)	Bayesian / GNN deconvolution	Simple mapping of scRNA to ST data	Limited spatial context, struggles with noise	Easy to use, widely adopted	Poor resolution, weak noise handling
ADCL (Zhang 2023)	Dual contrastive GNN + VAE framework	Captures both spatial topology and gene variance	More adaptive and interpretable than Cell2location or GraphST	Strong interpretability, adaptive learning	Requires careful training, moderate compute
MHDGAT + OT (Chen 2024)	Dynamic graph attention + optimal transport fusion	High accuracy in low-resolution data	Outperforms Tangram / DestVI in mapping tasks	Excels on noisy or low-res ST data	Computationally intensive (OT)
MACD (Huang 2024)	Masked autoencoder + adversarial alignment	Robust to noisy and heterogeneous datasets	Beats six top deconvolution models in PCC / SSIM metrics	Strong generalization, top quantitative metrics	Adversarial training harder to stabilize
CellMirror (Xia 2023)	Contrastive GNN with interpretable markers	Biologically explainable embeddings	Adds interpretability missing in GraphST and DestVI	Highlights marker genes, interpretable	Needs high-quality scRNA-seq reference data