Novelty Assessment Report

Paper: Fusing Pixels and Genes: Spatially-Aware Learning in Computational Pathology

PDF URL: https://openreview.net/pdf?id=uVXO6gzVzj

Venue: ICLR 2026 Conference Submission

Year: 2026

Report Generated: 2026-01-01

Abstract

Recent years have witnessed remarkable progress in multimodal learning within computational pathology. Existing models primarily rely on vision and language modalities; however, language alone lacks molecular specificity and offers limited pathological supervision, leading to representational bottlenecks. In this paper, we propose STAMP, a Spatial Transcriptomics-Augmented Multimodal Pathology representation learning framework that integrates spatially-resolved gene expression profiles to enable molecule-guided joint embedding of pathology images and transcriptomic data. Our study shows that self-supervised, gene-guided training provides a robust and task-agnostic signal for learning pathology image representations. Incorporating spatial context and multi-scale information further enhances model performance and generalizability. To support this, we constructed SpaVis-6M, the largest Visium-based spatial transcriptomics dataset to date, and trained a spatially-aware gene encoder on this resource. Leveraging hierarchical multi-scale contrastive alignment and cross-scale patch localization mechanisms, STAMP effectively aligns spatial transcriptomics with pathology images, capturing spatial structure and molecular variation. We validate STAMP across six datasets and four downstream tasks, where it consistently achieves strong performance. These results highlight the value and necessity of integrating spatially resolved molecular supervision for advancing multimodal learning in computational pathology. The code is included in the supplementary materials. The pretrained weights and SpaVis-6M will be released for community development after reviewing the manuscript.

Disclaimer

This report is **AI-GENERATED** using Large Language Models and WisPaper (a scholar search engine). It analyzes academic papers' tasks and contributions against retrieved prior work. While this system identifies **POTENTIAL** overlaps and novel directions, **ITS COVERAGE IS NOT EXHAUSTIVE AND JUDGMENTS ARE APPROXIMATE**. These results are intended to assist human reviewers and **SHOULD NOT** be relied upon as a definitive verdict on novelty.

Note that some papers exist in multiple, slightly different versions (e.g., with different titles or URLs). The system may retrieve several versions of the same underlying work. The current automated pipeline does not reliably align or distinguish these cases, so human reviewers will need to disambiguate them manually.

If you have any questions, please contact: mingzhang23@m.fudan.edu.cn

Core Task Landscape

This paper addresses: **multimodal representation learning integrating pathology images and spatial transcriptomics**

A total of **50 papers** were analyzed and organized into a taxonomy with **13 categories**.

Taxonomy Overview

The research landscape has been organized into the following main categories:

- **Foundation Models and Cross-Modal Pretraining**
- **Spatial Domain Identification and Tissue Segmentation**
- **Gene Expression Prediction from Histology**
- **Multi-Modal Disentanglement and Integration**
- **Datasets, Benchmarks, and Methodological Reviews**

Complete Taxonomy Tree

- multimodal representation learning integrating pathology images and spatial transcriptomics Survey Taxonomy
- Foundation Models and Cross-Modal Pretraining
  - Pan-Cancer and Multi-Organ Foundation Models ★ (7 papers)
  - [0] Fusing Pixels and Genes: Spatially-Aware Learning in Computational Pathology (Anon et al., 2026) View paper
  - [1] Past: A multimodal single-cell foundation model for histopathology and spatial transcriptomics in cancer (Yang Chang-chun, 2025) View paper
  - [9] A large-scale benchmark of cross-modal learning for histology and gene expression in spatial transcriptomics (Gindra, 2025) View paper
  - [10] STPath: a generative foundation model for integrating spatial transcriptomics and whole-slide images (Tinglin Huang, 2025) View paper
  - [14] Pan-cancer integrative histology-genomic analysis via multimodal deep learning (Richard J Chen, 2022) View paper
  - [15] spEMO: Leveraging Multi-Modal Foundation Models for Analyzing Spatial Multi-Omic and Histopathology Data (Hongyu Zhao, 2025) View paper
  - [36] Large-Scale Representation Learning and Generative Modeling for Multimodal Healthcare Data (Redekop, 2025) View paper
  - Contrastive Learning for Image-Gene Alignment (6 papers)
  - [20] Spatially resolved gene expression prediction from histology images via bi-modal contrastive learning (R Xie, 2023) View paper
  - [25] PathOmCLIP: Connecting tumor histology with spatial gene expression via locally enhanced contrastive learning of Pathology and Single-cell foundation model (Yongâ□□Ju Lee, 2024) View paper
  - [26] ST-Align: A Multimodal Foundation Model for Image-Gene Alignment in Spatial Transcriptomics (Lin Yuxiang, 2024) View paper
  - [27] Multimodal contrastive learning for spatial gene expression prediction using histology images (Shi, 2024) View paper
  - [33] Spatially Resolved Gene Expression Prediction from H&E Histology Images via Bi-modal Contrastive Learning (Xie, 2023) View paper
  - [45] Learning from Gene Names, Expression Values and Images: Contrastive Masked Text-Image Pretraining for Spatial Transcriptomics Representation Learning (Fang, 2025) View paper
  - Specialized Pretraining Paradigms (5 papers)
  - [28] Spatial omics driven crossmodal pretraining applied to graph-based deep learning for cancer pathology analysis (Zarif L. Azher, 2024) View paper

  - [37] RankByGene: Gene-Guided Histopathology Representation Learning Through Cross-Modal Ranking Consistency (Huang Wentao, 2024) View paper
  - [40] PEaRL: Pathway-Enhanced Representation Learning for Gene and Pathway Expression Prediction from Histology (Kapse, 2025) View paper
  - [44] Towards Unified Molecule-Enhanced Pathology Image Representation Learning via Integrating Spatial Transcriptomics (Han Minghao, 2024) View paper
  - [47] Multimodal Alignment Reveals Interpretable Geneâ□□Morphology Links in Perineuronal Net Pathology (W Liu, 2025) View paper
- Spatial Domain Identification and Tissue Segmentation
  - Graph-Based Spatial Domain Detection (3 papers)
  - [5] SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network (Jian Hu, 2021) View paper
  - [12] stGCL: A versatile cross-modality fusion method based on multi-modal graph contrastive learning for spatial transcriptomics (Yu Na, 2023) View paper
  - [48] A Multimodal Graph Learning Framework for Versatile Spatial Transcriptomics Analysis with SpatialModal (Xingyi Li, 2025) View paper
  - Contrastive Multi-Modal Feature Fusion (2 papers)
  - [4] TriCLFF: a multi-modal feature fusion framework using contrastive learning for spatial domain identification (Fenglan Pang, 2025) View paper
  - [22] MuCST: restoring and integrating heterogeneous morphology images and spatial transcriptomics data with contrastive learning. (Yu Wang, 2025) View paper
  - Anomalous Tissue Region Detection (2 papers)
  - [2] Meatrd: Multimodal anomalous tissue region detection enhanced with spatial transcriptomics (Kai-Chen Xu, 2025) View paper
  - [31] SPaSE: Spatially resolved pathology scores using optimal transport on spatial transcriptomics data. (Mohammad Nuwaisir Rahman, 2025) View paper
- Gene Expression Prediction from Histology
  - Generative Models for Expression Inference (3 papers)
  - [3] Diffusion generative modeling for spatially resolved gene expression inference from histology images (Zhu Sichen, 2025) View paper
  - [11] Cross-modal diffusion modelling for super-resolved spatial transcriptomics (Xiaofei Wang, 2024) View paper
  - [19] GenST: A generative cross-modal model for predicting spatial transcriptomics from histology images (R Wood, 2025) View paper
  - Contrastive and Dual-Scale Prediction (4 papers)
  - [13] Spatia: Multimodal model for prediction and generation of spatial cell phenotypes (Zhenglun Kong, 2025) View paper
  - [16] Deep Learning-Enabled Integration of Histology and Transcriptomics for Tissue Spatial Profile Analysis (Yongxin Ge, 2025) View paper
  - [21] Spatially gene expression prediction using dual-scale contrastive learning (Mingcheng Qu, 2025) View paper
  - [23] Deep learning-enabled integration of histology and transcriptomics for analyzing single-cell spatial profiles (Yongxin Ge, 2024) View paper
  - Graph and Topology-Aware Prediction (2 papers)
  - [7] Multi-modal Topology-embedded Graph Learning for Spatially Resolved Genes Prediction from Pathology Images with Prior Gene Similarity Information (Hang Shi, 2025) View paper
  - [35] MagNet: Multi-Level Attention Graph Network for Predicting High-Resolution Spatial Transcriptomics (Zhu Junchao, 2025) View paper
  - Relative Expression and Ranking-Based Prediction (2 papers)
  - [18] Learning Relative Gene Expression Trends from Pathology Images in Spatial Transcriptomics (Kazuya Nishimura, 2025) View paper
  - [49] Deep Association Multimodal Learning for Zero-Shot Spatial Transcriptomics Prediction (Yijing Zhou, 2025) View paper
  - Multi-Slice and High-Resolution Prediction (4 papers)
  - [32] Segmentation-free integration of nuclei morphology and spatial transcriptomics for retinal images (Chelebian, 2025) View paper
  - [34] ROICellTrack: a deep learning framework for integrating cellular imaging modalities in subcellular spatial transcriptomic profiling of tumor tissues. (Song Xiaofei, 2025) View paper
  - [43] Geometry-informed multimodal fusion network for enhancing high-density spatial transcriptomics from histology images (Zhiceng Shi, 2025) View paper
  - [46] Inferring multi-slice spatially resolved gene expression from H&E-stained histology images with STMCL. (Zhiceng Shi, 2025) View paper
- Multi-Modal Disentanglement and Integration (5 papers)
  - [8] Multi-modal disentanglement of spatial transcriptomics and histopathology imaging (Hassaan Maan, 2025) View paper
  - [39] Abstract B010: Spatially-resolved prediction of gene expression signatures in H&E whole slide images using additive multiple instance learning models (Miles Markey, 2023) View paper
  - [41] Multimodal Deep Learning for Subtype Classification in Breast Cancer Using Histopathological Images and Gene Expression Data (Shandiz, 2025) View paper
  - [42] Jointly leveraging spatial transcriptomics and deep learning models for pathology image annotation improves cell type identification over either approach alone (Asif Zubair, 2021) View paper
  - [50] Integrative deep learning of spatial multi-omics with SWITCH (Zhongzhan Li, 2025) View paper
- Datasets, Benchmarks, and Methodological Reviews (6 papers)
  - [6] Stimage-1k4m: A histopathology image-gene expression dataset for spatial transcriptomics (Chen Jiawen, 2024) View paper
  - [17] Statistical and machine learning methods for spatially resolved transcriptomics with histology (Jian Hu, 2021) View paper
  - [24] Deep learning methods for the integration of multi-omics and histopathology data for precision medicine in oncology (Benkirane, 2024) View paper
  - [29] Systematic benchmarking of high-throughput subcellular spatial transcriptomics platforms (Wang, 2024) View paper
  - [30] Integration of imaging-based and sequencing-based spatial omics mapping on the same tissue section via DBiTplus (Rong Fan, 2024) View paper

- [38] HAGE: Hierarchical Alignment Gene-Enhanced Pathology Representation Learning with Spatial Transcriptomics (Thao M. Dang, 2025) View paper

Narrative

Core task: multimodal representation learning integrating pathology images and spatial transcriptomics. This field seeks to bridge high-resolution histology with spatially resolved gene expression profiles, enabling richer characterizations of tissue architecture and molecular function. The taxonomy organizes research into several major branches. Foundation Models and Cross-Modal Pretraining encompasses large-scale efforts that learn joint embeddings across imaging and genomic modalities, often leveraging contrastive or generative objectives to capture pan-cancer or multi-organ patterns (e.g., Past Multimodal Foundation[1], STPath Foundation Model[10]). Spatial Domain Identification and Tissue Segmentation focuses on delineating biologically meaningful regions within tissue sections, combining graph-based and deep learning techniques (e.g., SpaGCN Spatial Domains[5]). Gene Expression Prediction from Histology aims to infer transcriptomic profiles directly from morphology, using regression, diffusion, or ranking-based models (e.g., Diffusion Gene Expression[3]). Multi-Modal Disentanglement and Integration addresses the challenge of separating and recombining modality-specific versus shared information (e.g., Multimodal Disentanglement[8]). Finally, Datasets, Benchmarks, and Methodological Reviews provide standardized resources and comparative analyses to guide method development (e.g., Stimage Dataset[6], Cross-Modal Benchmark[9]).

Several active lines of work highlight key trade-offs and open questions. Foundation model approaches emphasize scalability and transferability, training on diverse cohorts to produce general-purpose representations, yet they must balance pretraining complexity with downstream task performance. In contrast, spatial domain methods prioritize interpretability and biological fidelity, often incorporating graph structures or topological constraints, but may struggle with heterogeneity across tissue types. Gene expression prediction methods explore whether morphology alone suffices for transcriptomic inference or whether explicit spatial context is essential. Within this landscape, Fusing Pixels Genes[0] sits naturally among pan-cancer foundation models, sharing the ambition of Pan-Cancer Histology-Genomic[14] and spEMO Foundation Models[15] to learn cross-modal embeddings at scale. Compared to these neighbors, Fusing Pixels Genes[0] likely emphasizes tighter integration of pixel-level histology features with spatially indexed gene profiles, positioning itself as a bridge between large-scale pretraining and spatially aware representation learning.

Related Works in Same Category

The following **6 sibling papers** share the same taxonomy leaf node with the original paper:

1. Past: A multimodal single-cell foundation model for histopathology and spatial transcriptomics in cancer

Authors: Yang Chang-chun, Li, Haoyang, Changchun Yang, Wu Yushuai, et al. (23 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

While pathology foundation models have transformed cancer image analysis, they often lack integration with molecular data at single-cell resolution, limiting their utility for precision oncology. Here, we present PAST, a pan-cancer single-cell foundation model trained on 20 million paired histopathology images and single-cell transcriptomes spanning multiple tumor types and tissue contexts. By jointly encoding cellular morphology and gene expression, PAST learns unified cross-modal representatio...

Relationship Analysis

Both papers belong to the Pan-Cancer and Multi-Organ Foundation Models category, training large-scale multimodal models that integrate pathology images with molecular data across diverse tissue types and cancer contexts. They overlap in their core approach of using contrastive learning to align histopathology with transcriptomic data and both aim to learn generalizable cross-modal representations for downstream cancer analysis tasks. However, STAMP focuses on spot-level spatial transcriptomics (Visium technology) with 5.75M spots and emphasizes spatial-aware training with neighborhood context modeling, while PAST operates at single-cell resolution with 20M paired images and single-cell transcriptomes, targeting cellular-level molecular heterogeneity and virtual molecular staining applications.

2. A large-scale benchmark of cross-modal learning for histology and gene expression in spatial transcriptomics

Authors: Gindra, Rushin H., Palla, Giovanni, Wagner, et al. (15 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Spatial transcriptomics enables simultaneous measurement of gene expression and tissue morphology, offering unprecedented insights into cellular organization and disease mechanisms. However, the field lacks comprehensive benchmarks for evaluating multimodal learning methods that leverage both histology images and gene expression data. Here, we present HESCAPE, a large-scale benchmark for cross-modal contrastive pretraining in spatial transcriptomics, built on a curated pan-organ dataset spanning...

Relationship Analysis

Both papers belong to the Pan-Cancer and Multi-Organ Foundation Models category, focusing on large-scale multimodal pretraining that integrates pathology images with spatial transcriptomics across diverse tissue types. They overlap in their use of contrastive learning frameworks to align histology images with gene expression data and both construct large-scale datasets (SpaVis-6M vs. HESCAPE) for training generalizable cross-modal representations. The key difference is that STAMP emphasizes a two-stage pretraining strategy with spatially-aware gene encoders and hierarchical multi-scale alignment mechanisms, while HESCAPE focuses on systematic benchmarking of existing image and gene encoders across multiple pretraining strategies and downstream tasks, revealing that gene encoder selection is the primary determinant of performance.

3. STPath: a generative foundation model for integrating spatial transcriptomics and whole-slide images

Authors: Tinglin Huang, Tianyu Liu, Mehrtash Babadi, Rex Ying, Wengong Jin | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Abstract Spatial transcriptomics (ST) has shown remarkable promise in pathology applications, shedding light on the spatial organization of gene expression and its relationship to the tumor microenvironment. However, its clinical adoption remains constrained due to the limited scalability of current sequencing technologies. While recent methods attempt to infer ST from whole slide images (WSIs) using pretrained image encoders, they remain restricted by limited gene coverage, organ-specific train...

Relationship Analysis

Both papers belong to the Pan-Cancer and Multi-Organ Foundation Models category, training large-scale models on diverse tissue types to learn generalizable cross-modal representations between pathology images and spatial transcriptomics. They overlap in using spatial transcriptomics as molecular supervision for pathology image representation learning, both constructing large-scale datasets (SpaVis-6M vs. HEST-1K/STImage-1k4m) and employing multi-scale spatial-aware architectures. However, STAMP focuses on contrastive alignment with hierarchical multi-scale mechanisms and cross-scale patch localization, while STPath adopts a generative modeling paradigm with masked gene expression prediction and geometry-aware Transformers for direct gene expression inference across 38,984 genes.

4. Pan-cancer integrative histology-genomic analysis via multimodal deep learning

Authors: Richard J Chen, Ming Y. Lu, Richard J. Chen, Drew F. K. Williamson, Tiffany Y. Chen, et al. (14 authors total) | **Year/Venue**: 2022 | **URL**: View paper

Abstract

â¦ Though multimodal learning has been successful in technical domains such as the â¦ RNA-seq, mass cytometry, and spatial transcriptomics, these technologies continue to mature and â¦

Relationship Analysis

Both papers belong to the Pan-Cancer and Multi-Organ Foundation Models category, training on diverse tissue types to learn generalizable cross-modal representations. They overlap in integrating pathology images with molecular data (spatial transcriptomics in STAMP, bulk genomics in the candidate) for survival prediction across multiple cancer types using multimodal deep learning. The key difference is that STAMP focuses on spatially-resolved transcriptomics with spot-level alignment and spatial context modeling, while the candidate paper uses bulk RNA-seq and copy-number variation data for patient-level integration without spatial resolution.

5. spEMO: Leveraging Multi-Modal Foundation Models for Analyzing Spatial Multi-Omic and Histopathology Data

Authors: Hongyu Zhao, Tianyu Liu, Tinglin Huang, Tong Ding, Hao Wu, et al. (12 authors total) | **Year/Venue**: 2025 | **URL**: View paper

Abstract

Recent advances in pathology foundation models (PFMs), which are pretrained on large-scale histopathological images, have significantly accelerated progress in disease-centered applications. In parallel, spatial multi-omic technologies collect gene and protein expression levels at high spatial resolution, offering rich understanding of tissue context. However, current models fall short in effectively integrating these complementary data modalities. To fill in this gap, we...

Relationship Analysis

Both papers belong to the Pan-Cancer and Multi-Organ Foundation Models category, training large-scale models that integrate pathology images with spatial transcriptomics across diverse tissue types. They overlap in using contrastive learning to align histopathology images with spatial gene expression data and evaluating on tasks like spatial domain identification and disease prediction. However, STAMP focuses on a two-stage pretraining approach with a novel spatial-aware gene encoder trained on 5.75M spots and hierarchical multi-scale contrastive alignment, while spEMO emphasizes unifying embeddings from existing pathology foundation models with large language models (LLMs) to analyze spatial multi-omic data and includes automated medical report generation as a key application.

6. Large-Scale Representation Learning and Generative Modeling for Multimodal Healthcare Data

Authors: E Redekop | **Year/Venue**: 2025 | **URL**: View paper

Abstract

â¦ The next study integrates histology with paired spatial transcriptomics through a mixture-of-â¦ tools for processing 3D pathology images and predicting patient outcomes. MAMBA [SWWâ¦

Relationship Analysis

Both papers belong to the Pan-Cancer and Multi-Organ Foundation Models category, developing large-scale multimodal frameworks that integrate pathology images with molecular data across diverse tissue types. They overlap in using spatial transcriptomics to guide pathology representation learning through contrastive alignment mechanisms and demonstrate generalizability across multiple organs and cancer types. However, the original paper (STAMP) focuses specifically on spatially-aware pretraining with a novel two-stage approach using 5.75M ST entries and hierarchical multi-scale alignment, while the candidate paper presents a broader dissertation framework encompassing not only histology-transcriptomics integration but also MRI analysis, volumetric pathology reconstruction, and longitudinal EHR modeling through generative transformers.

Contributions Analysis

Overall novelty summary. STAMP proposes a foundation model that integrates spatial transcriptomics with histopathology images through self-supervised, gene-guided contrastive learning. The paper resides in the 'Pan-Cancer and Multi-Organ Foundation Models' leaf, which contains seven papers including the original work. This leaf represents a moderately populated research direction within the broader taxonomy of fifty papers, indicating active but not overcrowded exploration of large-scale cross-modal pretraining approaches that aim for generalizability across diverse tissue types and cancer contexts.

The taxonomy reveals that STAMP's immediate neighbors pursue similar pan-cancer foundation modeling goals, while adjacent leaves explore contrastive image-gene alignment and specialized pretraining paradigms. The 'Contrastive Learning for Image-Gene Alignment' leaf contains six papers focused on latent space alignment, and the 'Specialized Pretraining Paradigms' leaf includes five papers using alternative objectives like pathway-level alignment. STAMP appears to bridge these directions by combining contrastive alignment with spatial context modeling, distinguishing itself through explicit incorporation of spatially-resolved gene expression rather than bulk or pathway-level representations.

Among thirty candidates examined through semantic search, none clearly refuted any of STAMP's three core contributions. The STAMP framework itself was assessed against ten candidates with zero refutable overlaps; the SpaVis-6M dataset construction similarly showed no prior work among ten examined papers; and the unified alignment loss combining spatial and multi-scale objectives found no refuting evidence across ten candidates. These statistics suggest that within the limited search scope, STAMP's specific combination of spatial transcriptomics integration, large-scale dataset construction, and hierarchical multi-scale alignment appears relatively unexplored, though the search does not cover the entire literature landscape.

Based on the top-thirty semantic matches examined, STAMP's contributions appear to occupy a distinct position within the foundation model space. The absence of refuting candidates across all three contributions indicates potential novelty in the specific technical approach, though this assessment is constrained by the search methodology and does not preclude the existence of related work outside the examined set. The moderately populated taxonomy leaf suggests the paper enters an active research area with established precedents but room for methodological differentiation.

This paper presents **3 main contributions**, each analyzed against relevant prior work:

Contribution 1: STAMP framework for spatially-aware multimodal pathology learning

Description: The authors introduce STAMP, a novel framework that combines pathology images with spatial transcriptomics data through spatially-aware and multi-scale contrastive learning. The framework uses hierarchical multi-scale contrastive alignment and cross-scale patch localization to capture spatial structure and molecular variation.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Geometry-informed multimodal fusion network for enhancing high-density spatial transcriptomics from histology images

URL: View paper

Brief Assessment

Geometry-Informed Fusion[43] focuses on enhancing high-density spatial transcriptomics from histology images using geometry-guided multimodal fusion with spatial coordinates. This differs from STAMP's spatially-aware contrastive learning framework that combines hierarchical multi-scale alignment and cross-scale patch localization for joint representation learning of pathology images and transcriptomics data.

2. GenST: A generative cross-modal model for predicting spatial transcriptomics from histology images

URL: View paper

Brief Assessment

GenST Generative Model[19] focuses on generative cross-modal prediction of spatial transcriptomics from histology using VQ-VAEs with dictionary-based latent alignment, whereas STAMP employs hierarchical multi-scale contrastive learning with spatial transcriptomics-augmented pretraining. The technical approaches and objectives differ fundamentally.

3. Breast cancer histopathology image-based gene expression prediction using spatial transcriptomics data and deep learning

URL: View paper

Brief Assessment

Breast Cancer Prediction[54] focuses on predicting gene expression from histopathology images using spatial transcriptomics data for breast cancer specifically, employing deep learning architectures (ResNet, EfficientNet, Vision Transformer) with an auxiliary network. In contrast, the ORIGINAL paper proposes STAMP as a general framework for spatially-aware multimodal learning that integrates hierarchical multi-scale contrastive alignment and cross-scale patch localization mechanisms across multiple organs and tasks. The candidate does not demonstrate that similar spatially-aware multimodal contrastive learning frameworks with these specific architectural components existed prior to STAMP.

4. Past: A multimodal single-cell foundation model for histopathology and spatial transcriptomics in cancer

URL: View paper

Brief Assessment

Past Multimodal Foundation[1] focuses on single-cell resolution transcriptomics paired with cellular-level histopathology images, while STAMP operates at the tissue spot level using spatial transcriptomics (Visium platform with 55-micrometer spots). These represent fundamentally different scales and technical approaches to multimodal pathology learning.

5. Histopathologic analysis of human kidney spatial transcriptomics data: toward precision pathology

URL: View paper

Brief Assessment

Kidney Precision Pathology[53] focuses on morphology-based histopathologic analysis of kidney spatial transcriptomics data for precision pathology, not on developing a general multimodal learning framework that integrates spatial transcriptomics with histology images through contrastive learning mechanisms.

6. Benchmarking the translational potential of spatial gene expression prediction from histology

URL: View paper

Brief Assessment

Translational Potential Benchmarking[55] focuses on benchmarking existing methods for predicting spatial gene expression from histology images, evaluating their translational potential and clinical applicability. It does not propose a novel multimodal learning framework like STAMP, which introduces hierarchical multi-scale contrastive alignment and cross-scale patch localization mechanisms for joint representation learning.

7. Combining spatial transcriptomics with tissue morphology

URL: View paper

Brief Assessment

Combining Spatial Morphology[51] is a review paper that surveys methods combining spatial transcriptomics with tissue morphology, categorizing them into translation and integration approaches. It does not present a novel framework comparable to STAMP's hierarchical multi-scale contrastive alignment and cross-scale patch localization mechanisms.

8. Multi-modal disentanglement of spatial transcriptomics and histopathology imaging

URL: View paper

Brief Assessment

Multimodal Disentanglement[8] focuses on disentangling overlapping versus unique sources of variation between spatial transcriptomics and histopathology through a disentanglement framework (SpatialDIVA). In contrast, STAMP addresses hierarchical multi-scale contrastive alignment and cross-scale patch localization for joint embedding. The technical objectives and methodologies differ fundamentally.

9. STPath: a generative foundation model for integrating spatial transcriptomics and whole-slide images

URL: View paper

Brief Assessment

STPath Foundation Model[10] focuses on generative modeling for predicting gene expression from whole-slide images across diverse organs and sequencing technologies, whereas STAMP emphasizes contrastive learning with hierarchical multi-scale alignment and cross-scale patch localization for joint representation learning of pathology images and spatial transcriptomics.

10. Predicting breast cancer molecular subtypes from h &e-stained histopathological images using a spatial-transcriptomics-based patch filter

URL: View paper

Brief Assessment

Spatial-Transcriptomics Patch Filter[52] focuses on identifying clinically significant patches for breast cancer subtype prediction, not on developing a general spatially-aware multimodal pretraining framework with hierarchical multi-scale contrastive alignment.

Contribution 2: SpaVis-6M dataset construction

Description: The authors constructed SpaVis-6M, the largest Visium-based spatial transcriptomics dataset containing 5.75 million spatial transcriptomics entries from 35 organs, 1,982 slices, and 262 datasets or publications. This resource supports training of a robust spatial-aware gene encoder.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Integrating single-cell and spatially resolved transcriptomic strategies to survey the astrocyte response to stroke in male mice

URL: View paper

Brief Assessment

Astrocyte Stroke Response[71] focuses on astrocyte responses to stroke using Visium technology but does not construct a large-scale multi-organ spatial transcriptomics dataset comparable to SpaVis-6M.

2. High-definition spatial transcriptomic profiling of immune cell populations in colorectal cancer

URL: View paper

Brief Assessment

Colorectal Immune Profiling[69] focuses on high-resolution spatial transcriptomics of colorectal cancer using Visium HD technology, not on constructing a large-scale multi-organ Visium-based dataset like SpaVis-6M.

3. Single-cell, single-nucleus, and spatial transcriptomics characterization of the immunological landscape in the healthy and PSC human liver

URL: View paper

Brief Assessment

PSC Liver Landscape[70] focuses on immunological characterization of healthy and PSC human liver tissue using single-cell, single-nucleus, and spatial transcriptomics. It does not describe construction of a large-scale, multi-organ Visium-based spatial transcriptomics dataset comparable to SpaVis-6M.

4. Systematic benchmarking of high-throughput subcellular spatial transcriptomics platforms across human tumors

URL: View paper

Brief Assessment

Tumor Platform Benchmarking[72] focuses on systematic evaluation of four commercial spatial transcriptomics platforms (Stereo-seq v1.3, Visium HD FFPE, CosMx 6K, Xenium 5K) across tumor samples, not on constructing large-scale Visium-based datasets for training foundation models.

5. A practical guide to spatial transcriptomics: lessons from over 1000 samples

URL: View paper

Brief Assessment

Practical Guide[68] discusses experience with spatial transcriptomics across multiple platforms (Visium, Visium HD) but does not describe constructing a large-scale dataset comparable to SpaVis-6M. The candidate focuses on practical guidance from analyzing samples rather than dataset curation for model training.

6. A spatially resolved transcriptome landscape during thyroid cancer progression

URL: View paper

Brief Assessment

Thyroid Cancer Landscape[66] focuses on thyroid cancer progression using spatial transcriptomics but does not describe constructing a large-scale, multi-organ Visium dataset comparable to SpaVis-6M's scope (5.75M entries, 35 organs, 262 datasets).

7. Museum of spatial transcriptomics

URL: View paper

Brief Assessment

Museum Spatial Transcriptomics[67] appears to be a general repository or collection of spatial transcriptomics data. The provided candidate context is too fragmentary to determine whether it contains a comparable large-scale Visium-based dataset or challenges the novelty of SpaVis-6M's scale and construction.

8. Hest-1k: A dataset for spatial transcriptomics and histology image analysis

URL: View paper

Brief Assessment

Hest Dataset[65] focuses on paired spatial transcriptomics and histology images (1,229 samples) for multimodal analysis, while SpaVis-6M emphasizes large-scale gene expression data (5.75 million entries) for training spatial-aware gene encoders. The datasets serve different primary purposes and scales.

9. Spatial transcriptomics in health and disease

URL: View paper

Brief Assessment

Health Disease Review[73] is a review paper discussing spatial transcriptomics applications in health and disease contexts. It does not describe the construction of any large-scale Visium-based dataset comparable to SpaVis-6M.

10. Spatial transcriptomics at subspot resolution with BayesSpace

URL: View paper

Brief Assessment

BayesSpace Subspot[74] focuses on computational methods for enhancing spatial resolution of existing Visium data through Bayesian statistical modeling, not on constructing large-scale spatial transcriptomics datasets for training foundation models.

Contribution 3: Unified alignment loss combining spatial and multi-scale objectives

Description: The authors develop a unified alignment loss function that integrates multiple objectives including cross-scale patch positioning, inter-modal contrastive alignment between images and genes, and intra-modal alignment between patches and regions. This design enables the model to learn spatial relationships and multi-scale features effectively.

This contribution was assessed against **10 related papers** from the literature. Papers with potential prior art are analyzed in detail with textual evidence; others receive brief assessments.

1. Counterfactual contrastive learning for weakly-supervised vision-language grounding

URL: View paper

Brief Assessment

Counterfactual Contrastive[63] focuses on weakly-supervised vision-language grounding using counterfactual transformations and contrastive learning between positive/negative samples. It does not address spatial transcriptomics, multi-scale patch positioning, or cross-modal alignment between images and genes as in the original paper's unified loss design.

2. Spatial-aware Multi-modal Contrastive Learning for RGB-D salient object detection and beyond

URL: View paper

Brief Assessment

RGB-D Contrastive Learning[56] focuses on RGB-D salient object detection with multi-scale cross-modal contrastive learning between RGB and depth modalities, not on spatial transcriptomics alignment with pathology images or cross-scale patch positioning mechanisms.

3. Spatialâ□□temporal video grounding with cross-modal understanding and enhancement

URL: View paper

Brief Assessment

Spatial-Temporal Grounding[62] focuses on video grounding with temporal-spatial alignment between video frames and text queries, not on pathology image-gene expression alignment with cross-scale patch positioning and multi-modal contrastive objectives as in the original paper.

4. 3d coca: Contrastive learners are 3d captioners

URL: View paper

Brief Assessment

3D Coca[60] focuses on 3D scene captioning with contrastive vision-language learning for point clouds, not on spatial transcriptomics or multi-scale pathology image alignment. The technical domains and objectives are fundamentally different.

5. Hierarchical set-to-set representation for 3-d cross-modal retrieval

URL: View paper

Brief Assessment

Hierarchical Set-to-Set[61] addresses 3-D cross-modal retrieval between 3D objects and other modalities, not spatial transcriptomics with pathology images. The hierarchical similarity combines global-to-global and local-to-local metrics for 3D object retrieval, which is fundamentally different from the original paper's spatial localization and multi-scale patch positioning in computational pathology.

6. Multi-modal multi-scale representation learning via cross-attention between chest radiology images and free-text reports

URL: View paper

Brief Assessment

Cross-Attention Radiology[64] focuses on chest radiology image-text alignment using cross-attention mechanisms. The candidate does not address spatial transcriptomics, gene expression alignment, or the specific combination of cross-scale patch positioning with inter/intra-modal contrastive objectives described in the original paper.

7. Transcending fusion: A multi-scale alignment method for remote sensing image-text retrieval

URL: View paper

Brief Assessment

Multi-Scale Alignment[57] focuses on remote sensing image-text retrieval with multi-scale visual-textual alignment, not computational pathology with spatial transcriptomics. The domains, modalities, and technical objectives are fundamentally different.

8. Multi-scale multi-instance visual sound localization and segmentation

URL: View paper

Brief Assessment

Multi-Scale Sound Localization[59] focuses on audio-visual localization in videos using multi-scale visual features for sound source localization, not on spatial transcriptomics alignment with pathology images or gene expression data.

9. Complementary and Contrastive Learning for Audio-Visual Segmentation

URL: View paper

Brief Assessment

Audio-Visual Segmentation[58] focuses on audio-visual segmentation tasks using contrastive learning for cross-modal alignment, not on spatial transcriptomics or multi-scale patch positioning in computational pathology.

10. Multimodal contrastive learning for spatial gene expression prediction using histology images

URL: View paper

Brief Assessment

Multimodal Contrastive Prediction[27] focuses on contrastive learning for spatial gene expression prediction from histology images. The provided candidate context is extremely limited and does not contain sufficient technical detail about alignment loss functions, spatial objectives, or multi-scale learning mechanisms to assess novelty claims.

Appendix: Text Similarity Detection

No high-similarity text segments were detected across any compared papers.

References

- [0] Fusing Pixels and Genes: Spatially-Aware Learning in Computational Pathology View paper
- [1] Past: A multimodal single-cell foundation model for histopathology and spatial transcriptomics in cancer View paper
- [2] Meatrd: Multimodal anomalous tissue region detection enhanced with spatial transcriptomics View paper
- [3] Diffusion generative modeling for spatially resolved gene expression inference from histology images View paper
- [4] TriCLFF: a multi-modal feature fusion framework using contrastive learning for spatial domain identification View paper
- [5] SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network View paper
- [6] Stimage-1k4m: A histopathology image-gene expression dataset for spatial transcriptomics View paper
- [7] Multi-modal Topology-embedded Graph Learning for Spatially Resolved Genes Prediction from Pathology Images with Prior Gene Similarity Information View paper
- [8] Multi-modal disentanglement of spatial transcriptomics and histopathology imaging View paper
- [9] A large-scale benchmark of cross-modal learning for histology and gene expression in spatial transcriptomics View paper
- [10] STPath: a generative foundation model for integrating spatial transcriptomics and whole-slide images View paper
- [11] Cross-modal diffusion modelling for super-resolved spatial transcriptomics View paper
- [12] stGCL: A versatile cross-modality fusion method based on multi-modal graph contrastive learning for spatial transcriptomics View paper
- [13] Spatia: Multimodal model for prediction and generation of spatial cell phenotypes View paper
- [14] Pan-cancer integrative histology-genomic analysis via multimodal deep learning View paper
- [15] spEMO: Leveraging Multi-Modal Foundation Models for Analyzing Spatial Multi-Omic and Histopathology Data View paper
- [16] Deep Learning-Enabled Integration of Histology and Transcriptomics for Tissue Spatial Profile Analysis View paper
- [17] Statistical and machine learning methods for spatially resolved transcriptomics with histology View paper
- [18] Learning Relative Gene Expression Trends from Pathology Images in Spatial Transcriptomics View paper
- [19] GenST: A generative cross-modal model for predicting spatial transcriptomics from histology images View paper
- [20] Spatially resolved gene expression prediction from histology images via bi-modal contrastive learning View paper
- [21] Spatially gene expression prediction using dual-scale contrastive learning View paper
- [22] MuCST: restoring and integrating heterogeneous morphology images and spatial transcriptomics data with contrastive learning. View paper
- [23] Deep learning-enabled integration of histology and transcriptomics for analyzing single-cell spatial profiles View paper
- [24] Deep learning methods for the integration of multi-omics and histopathology data for precision medicine in oncology View paper
- [25] PathOmCLIP: Connecting tumor histology with spatial gene expression via locally enhanced contrastive learning of Pathology and Single-cell foundation model View paper
- [26] ST-Align: A Multimodal Foundation Model for Image-Gene Alignment in Spatial Transcriptomics View paper
- [27] Multimodal contrastive learning for spatial gene expression prediction using histology images View paper
- [28] Spatial omics driven crossmodal pretraining applied to graph-based deep learning for cancer pathology analysis View paper
- [29] Systematic benchmarking of high-throughput subcellular spatial transcriptomics platforms View paper
- [30] Integration of imaging-based and sequencing-based spatial omics mapping on the same tissue section via DBiTplus View paper
- [31] SPaSE: Spatially resolved pathology scores using optimal transport on spatial transcriptomics data. View paper
- [32] Segmentation-free integration of nuclei morphology and spatial transcriptomics for retinal images View paper
- [33] Spatially Resolved Gene Expression Prediction from H&E Histology Images via Bi-modal Contrastive Learning View paper
- [34] ROICellTrack: a deep learning framework for integrating cellular imaging modalities in subcellular spatial transcriptomic profiling of tumor tissues. View paper
- [35] MagNet: Multi-Level Attention Graph Network for Predicting High-Resolution Spatial Transcriptomics View paper
- [36] Large-Scale Representation Learning and Generative Modeling for Multimodal Healthcare Data View paper
- [37] RankByGene: Gene-Guided Histopathology Representation Learning Through Cross-Modal Ranking Consistency View paper
- [38] HAGE: Hierarchical Alignment Gene-Enhanced Pathology Representation Learning with Spatial Transcriptomics View paper
- [39] Abstract B010: Spatially-resolved prediction of gene expression signatures in H&E whole slide images using additive multiple instance learning models View paper
- [40] PEaRL: Pathway-Enhanced Representation Learning for Gene and Pathway Expression Prediction from Histology View paper
- [41] Multimodal Deep Learning for Subtype Classification in Breast Cancer Using Histopathological Images and Gene Expression Data View paper
- [42] Jointly leveraging spatial transcriptomics and deep learning models for pathology image annotation improves cell type identification over either approach alone View paper
- [43] Geometry-informed multimodal fusion network for enhancing high-density spatial transcriptomics from histology images View paper
- [44] Towards Unified Molecule-Enhanced Pathology Image Representation Learning via Integrating Spatial Transcriptomics View paper
- [45] Learning from Gene Names, Expression Values and Images: Contrastive Masked Text-Image Pretraining for Spatial Transcriptomics Representation Learning View paper
- [46] Inferring multi-slice spatially resolved gene expression from H&E-stained histology images with STMCL. View paper
- [47] Multimodal Alignment Reveals Interpretable Geneâ□□Morphology Links in Perineuronal Net Pathology View paper
- [48] A Multimodal Graph Learning Framework for Versatile Spatial Transcriptomics Analysis with SpatialModal View paper
- [49] Deep Association Multimodal Learning for Zero-Shot Spatial Transcriptomics Prediction View paper
- [50] Integrative deep learning of spatial multi-omics with SWITCH View paper
- [51] Combining spatial transcriptomics with tissue morphology View paper
- [52] Predicting breast cancer molecular subtypes from h &e-stained histopathological images using a spatial-transcriptomics-based patch filter View paper
- [53] Histopathologic analysis of human kidney spatial transcriptomics data: toward precision pathology View paper
- [54] Breast cancer histopathology image-based gene expression prediction using spatial transcriptomics data and deep learning View paper
- [55] Benchmarking the translational potential of spatial gene expression prediction from histology View paper
- [56] Spatial-aware Multi-modal Contrastive Learning for RGB-D salient object detection and beyond View paper
- [57] Transcending fusion: A multi-scale alignment method for remote sensing image-text retrieval View paper

- [58] Complementary and Contrastive Learning for Audio-Visual Segmentation View paper
- [59] Multi-scale multi-instance visual sound localization and segmentation View paper
- [60] 3d coca: Contrastive learners are 3d captioners View paper
- [61] Hierarchical set-to-set representation for 3-d cross-modal retrieval View paper
- [62] Spatialâ□□temporal video grounding with cross-modal understanding and enhancement View paper
- [63] Counterfactual contrastive learning for weakly-supervised vision-language grounding View paper
- [64] Multi-modal multi-scale representation learning via cross-attention between chest radiology images and free-text reports View paper
- [65] Hest-1k: A dataset for spatial transcriptomics and histology image analysis View paper
- [66] A spatially resolved transcriptome landscape during thyroid cancer progression View paper
- [67] Museum of spatial transcriptomics View paper
- [68] A practical guide to spatial transcriptomics: lessons from over 1000 samples View paper
- [69] High-definition spatial transcriptomic profiling of immune cell populations in colorectal cancer View paper
- [70] Single-cell, single-nucleus, and spatial transcriptomics characterization of the immunological landscape in the healthy and PSC human liver View paper
- [71] Integrating single-cell and spatially resolved transcriptomic strategies to survey the astrocyte response to stroke in male mice View paper
- [72] Systematic benchmarking of high-throughput subcellular spatial transcriptomics platforms across human tumors View paper
- [73] Spatial transcriptomics in health and disease View paper
- [74] Spatial transcriptomics at subspot resolution with BayesSpace View paper