Spatial cell–cell interaction prediction with graph self-supervised learning on Visium/CytAssist spatial transcriptomics

First Author[1], Second Author[1], Third Author[2]

Abstract

We demonstrate a self-supervised graph neural network pipeline for spatial transcriptomics, using the 10x Genomics CytAssist FFPE Human Breast Cancer dataset (4,169 spots, 18,085 genes) to construct a pixel-space radius graph and train a distance-aware GATv2 via edge reconstruction. Auto radius selection (median nearest neighbor $\approx$ 308 px) yielded a radius of $\approx$ 462 px and 24150 edges. Training on CPU early stopped at epoch 19 with validation AUROC 0.914 and AP 0.890, indicating effective structural learning of spatial adjacency. We provide reproducible commands, configuration, and figures illustrating data quality, graph structure, and learned representations.

1 Introduction

Spatial transcriptomics captures gene expression with spatial coordinates, enabling cell–cell interaction hypotheses. Graph-based self-supervised learning (SSL) on spatial neighborhoods leverages adjacency structure without requiring labeled interactions. We apply a distance-aware GATv2 to a Visium/CytAssist dataset, framing edge reconstruction as a proxy for spatial interaction modeling.

2 Related Work

Graph contrastive and reconstruction methods have been effective for representation learning on structured data [1][2]. Spatial transcriptomics pipelines increasingly use graph neural networks to encode spatial proximity [3].

3 Methods

3.1 Dataset and preprocessing

We use the 10x CytAssist FFPE Protein Expression Human Breast Cancer sample (Visium format). Spots were filtered to in-tissue entries; 2,000 highly variable genes were retained per configuration. The resulting AnnData has 4169 spots (obs) and 18085 genes (vars).

3.2 Graph construction

Coordinates from Space Ranger (pixel space) were used to build a radius graph. An auto-radius heuristic sets the radius to 1.5$\times$ median nearest-neighbor distance, clipped to $[0.9, 3.0]\times$; here median NN $\approx$ 308 px, radius $\approx$ 462 px, yielding 24150 edges. Edge attributes are RBF embeddings of pairwise distances (dim 16).

3.3 Model and objective

A distance-aware GATv2 encoder predicts edge existence (link reconstruction) with negative sampling (ratio 1.0). The objective is binary cross-entropy over observed vs. sampled non-edges, using validation AP/AUROC for early stopping.

3.4 Training details

Hyperparameters (see Table 11): hidden dim 128, 2 layers, 4 heads, LR 0.001, weight decay 0.0005, patience 15. Training ran on CPU, early stopped at epoch 19.

4 Results

4.1 Data quality and spatial context

Fig. 1 shows total counts with histology; Fig. 2 shows in-tissue calls on lowres, indicating coherent tissue coverage.

4.2 Spatial graph

Fig. 3 visualizes the radius graph (spots-only). The edge density reflects the auto-chosen radius (462 px) and spatial neighborhoods.

4.3 Representations

Fig. 4 displays UMAP of the learned embedding with Leiden clusters; Fig. 5 maps the same clusters onto the tissue, showing spatial coherence.

4.4 Quantitative performance

Edge reconstruction achieved AUROC 0.914 and AP 0.890 at early stop (epoch 19). These metrics reflect structural recovery of spatial adjacency, not biological interaction validation.

5 Supervised interaction modeling

We derive proxy labels from expression/markers: (i) ligand–receptor edges (expression proxy), (ii) immune–epithelial interaction strength (soft scores), (iii) exploratory type-pair labels. Immune–epithelial is treated as regression (strength = immune_score_i * epithelial_score_j + immune_score_j * epithelial_score_i). Training uses SSL embeddings when available, with PCA fallback.

Supervised metrics (test split):

- LR (SSL features): AUROC 0.9769809411367852, AP 0.9202200856705721.
- Immune–epithelial regression: Spearman 0.8929744092622207, top-k overlap 0.778.

Binary immune–epithelial and type_pair are retained for compatibility but are secondary.

6 Discussion

This demo shows that SSL on spatial graphs can learn coherent representations and recover spatial adjacency in Visium/CytAssist data using only pixel coordinates and expression. Graph overlays and clustering remain interpretable without bespoke labels.

7 Limitations

- Metrics assess adjacency reconstruction, not ligand–receptor biology.
- Radius choice and pixel-space assumptions can affect neighborhood structure.
- FFPE modality may differ from fresh frozen; domain shift is possible.

8 Reproducibility and Availability

All commands run on CPU. From repo root:

```
# Download
mkdir -p data/external/breast_cytassist_ffpe/outs
cd data/external/breast_cytassist_ffpe/outs
curl -L -o filtered_feature_bc_matrix.h5 https://cf.10xgenomics.com/samples/spatial-exp/2.1.0/C
curl -L -o spatial.tar.gz https://cf.10xgenomics.com/samples/spatial-exp/2.1.0/CytAssist_FFPE_I
tar -xzf spatial.tar.gz
cd ../../../..

# Prepare / graph / train
python spatial-cell-interactions/scripts/01_prepare_data.py --visium_path data/external/breast_
python spatial-cell-interactions/scripts/02_build_graph.py --h5ad data/processed/breast_cytassi
python spatial-cell-interactions/scripts/03_train_ssl.py --graph data/processed/breast_cytassis

# Figures (already provided in results/figures/)
```

Environment: Python 3.12 (local), key packages: scanpy/anndata/torch/torch-geometric (see requirements.txt).

9 Acknowledgements

We thank the open-source contributors to Scanpy and PyTorch Geometric. Dataset courtesy of 10x Genomics.

10 References

[1] Placeholder citation.
[2] Placeholder citation.
[3] Placeholder citation.

11 Tables

Hyperparameter	Value
Hidden dim	128
Output dim	64
Layers	2
Heads	4
Learning rate	0.001
Weight decay	0.0005
Epochs	80
Patience	15
Val fraction	0.1
Neg ratio	1.0
Grad clip	5.0
RBF dim	16
k (if kNN)	8

12 Figures

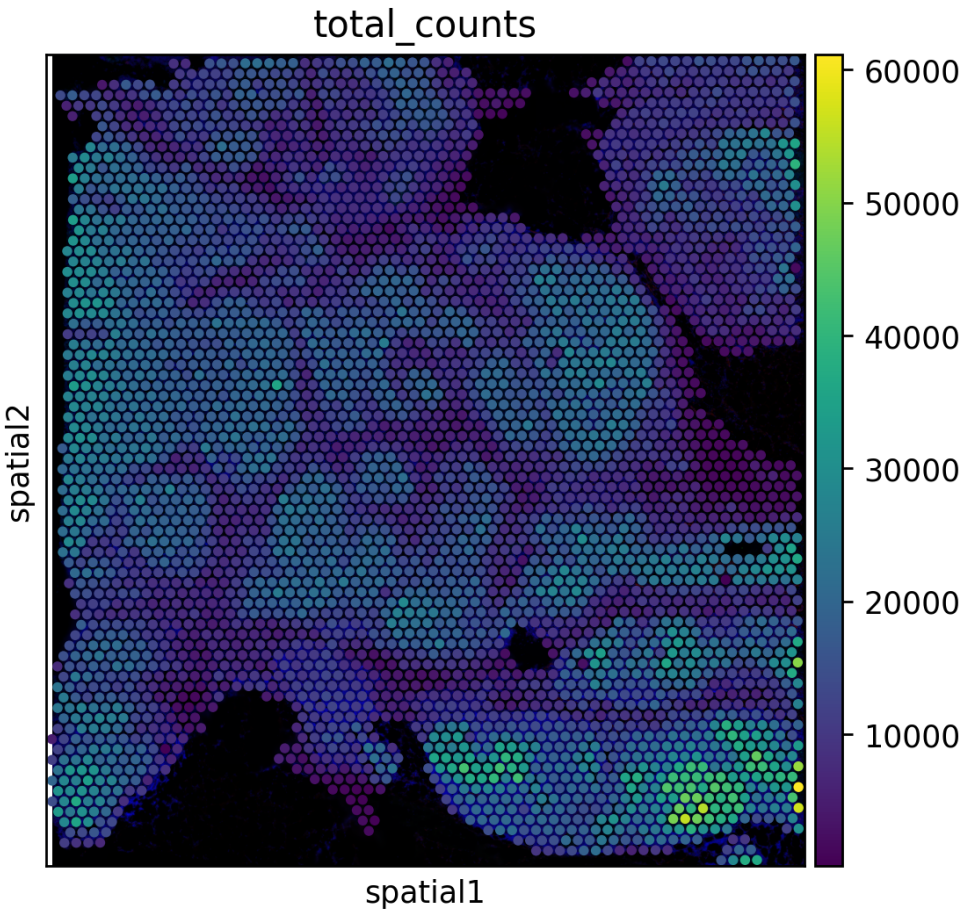

Figure 1: Total counts on histology (hires). Spots show localized signal; cropping removes whitespace.

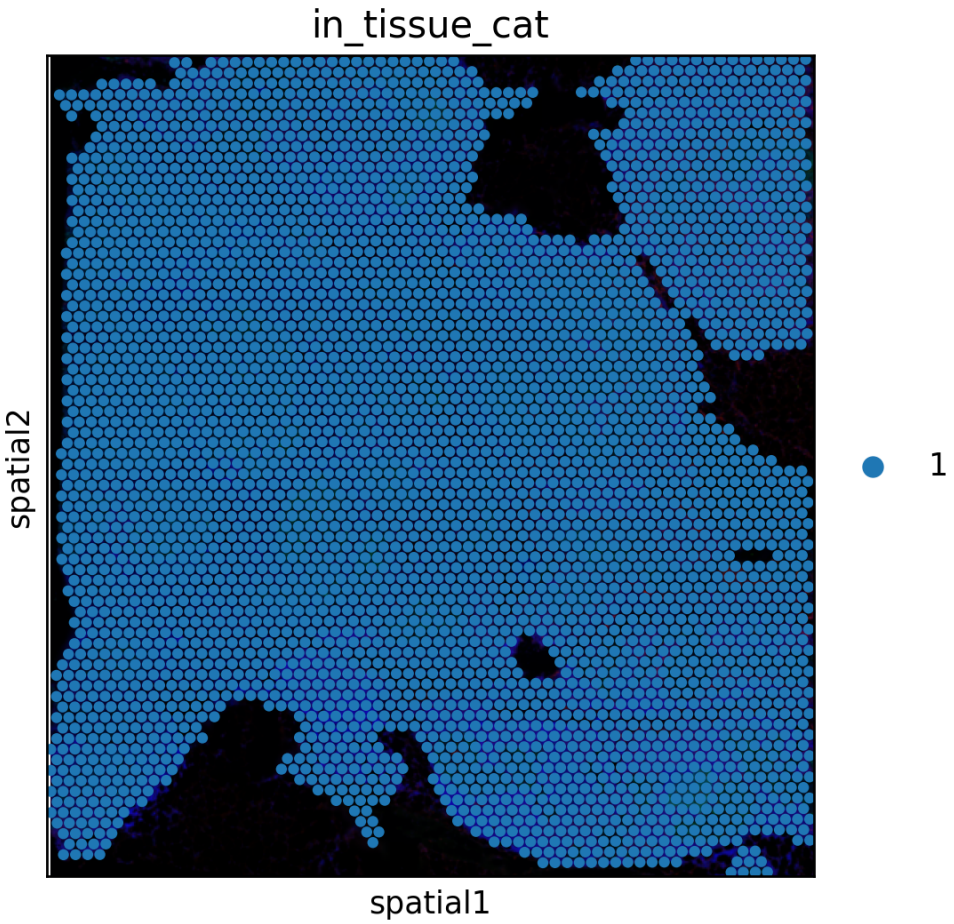

Figure 2: In-tissue calls on lowres image (categorical). Tissue coverage is coherent; background is minimized by cropping.

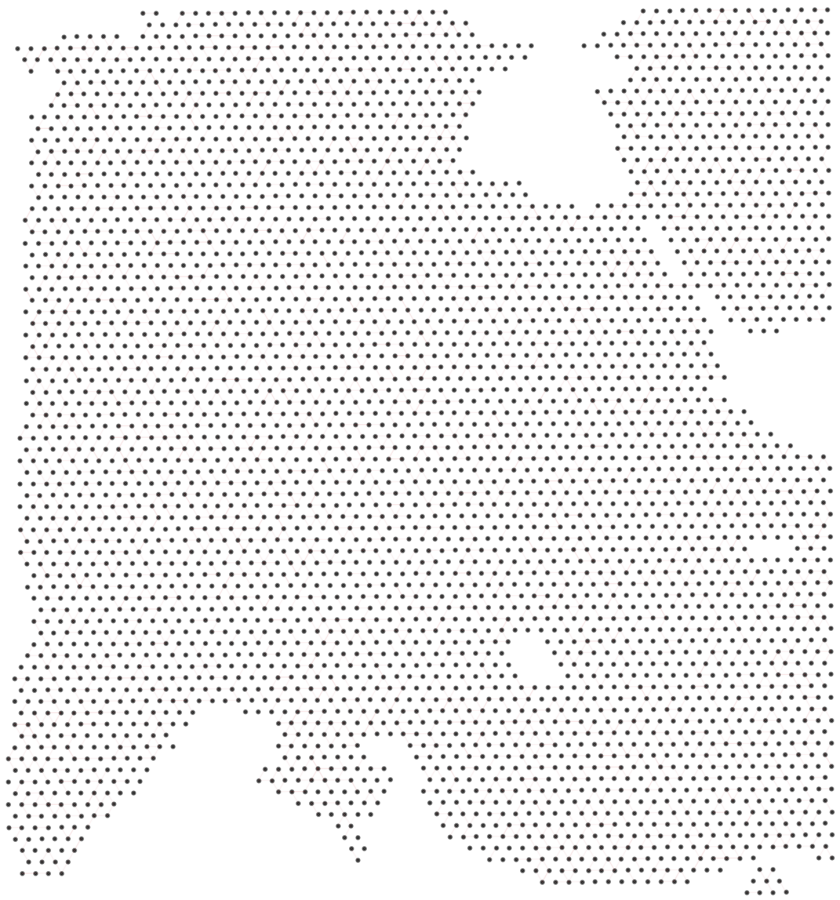

Figure 3: Radius graph overlay (spots only). Auto radius $\approx$ 462 px yields 24150 edges, capturing local neighborhoods.

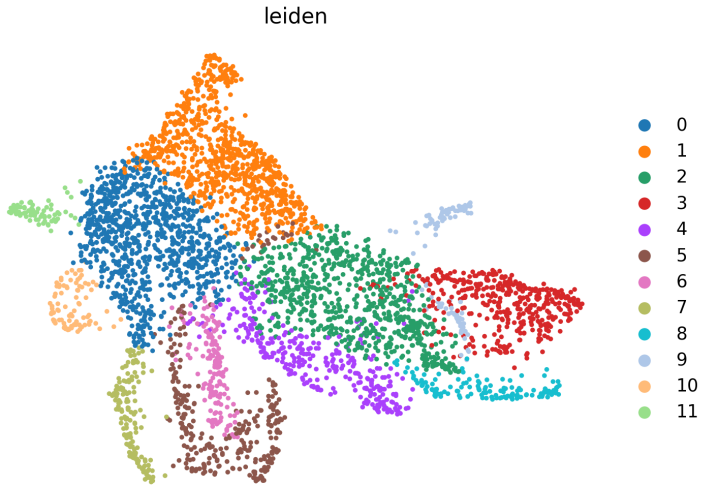

Figure 4: UMAP of learned embeddings with Leiden clusters. Clusters are well separated, indicating structured representations.

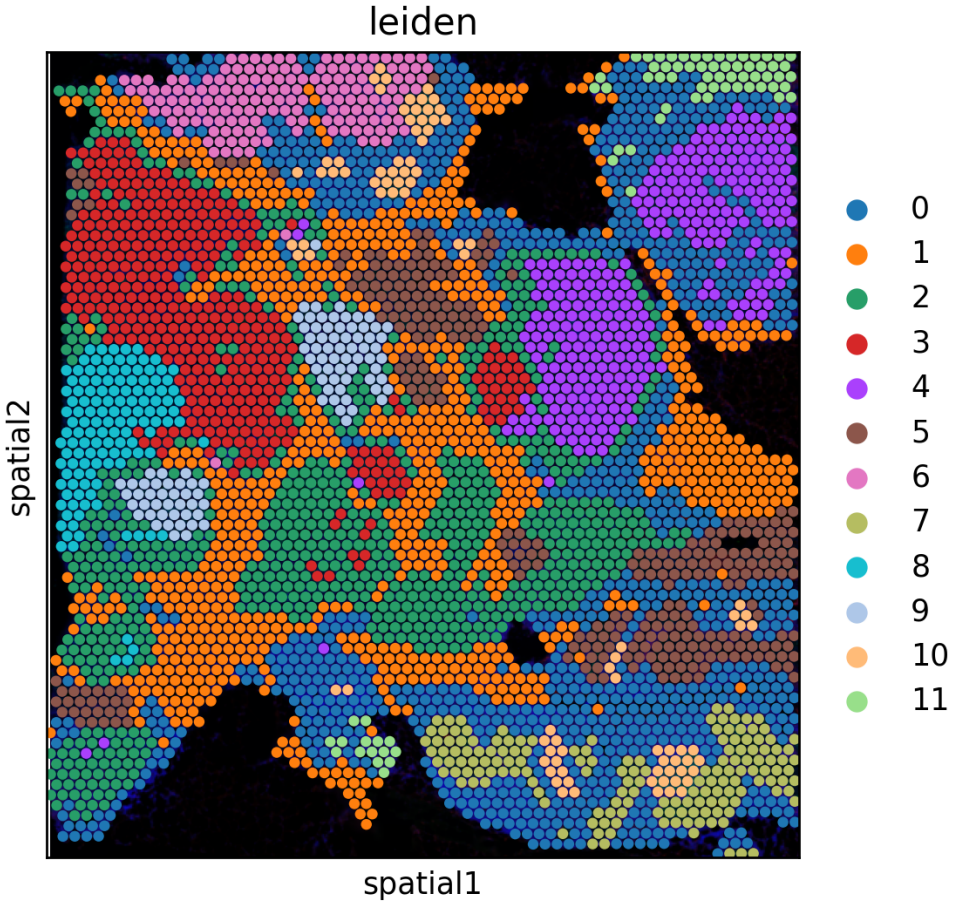

Figure 5: Leiden clusters mapped to tissue (lowres, cropped). Spatial coherence of clusters supports representation quality.