DeepSpot2Cell: Predicting Virtual Single-Cell Spatial Transcriptomics from H&E images using Spot-Level Supervision

Kalin Nonchev [*] [†]
kalin.nonchev@inf.ethz.ch

Glib Manaiev[*] [†]
gmanaiev@ethz.ch

Viktor H Koelzer[‡] [§] [¶]
viktor.koelzer@usb.ch

Gunnar Rätsch[†] [¶]
raetsch@inf.ethz.ch

Abstract

Spot-based spatial transcriptomics (ST) technologies like 10x Visium combine RNA sequencing with spatial imaging to quantify genome-wide gene expression while retaining tissue organization. However, their coarse spot-level resolution aggregates signals from multiple cells, preventing accurate single-cell analysis and detailed cellular characterization. Here, we present DeepSpot2Cell, a novel multi-modal DeepSet neural network that fuses sub-spot image detail and neighborhood context with pathology foundation-model features to effectively predict virtual single-cell gene expression from histopathological images using spot-level supervision. DeepSpot2Cell substantially improves gene expression correlations on a newly curated benchmark we specifically designed for single-cell ST deconvolution and prediction from H&E images. The benchmark includes 20 lung, 7 breast, and 2 pancreatic cancer samples, across which DeepSpot2Cell outperformed previous super-resolution methods, achieving respective improvements of 46%, 65%, and 38% in cell expression correlation for the top 100 genes. We hope that DeepSpot2Cell and this benchmark will stimulate further advancements in virtual single-cell ST, enabling more precise delineation of cell-type-specific expression patterns and facilitating enhanced downstream analyses.
Code availability: https://github.com/ratschlab/DeepSpot

1 Introduction

Spatial transcriptomics (ST) reveals spatial heterogeneity of tissue microenvironments and disease mechanisms [1, 2] by combining RNA sequencing with spatial imaging. However, current methods face a trade-off between resolution and transcriptome coverage [3]. For example, 10x Visium profiles the whole transcriptome but at coarse spot-level resolution (1–10 cells per spot) [4], while 10x Xenium achieves single-cell resolution but is limited to targeted gene panels. Emerging approaches, such as 10x Visium HD, provide subcellular, full-transcriptome coverage, yet still exhibit low sensitivity,

[*]Equal contribution.
[†]Institute for Machine Learning, Department of Computer Science, ETH Zurich, Switzerland
[‡]Institute of Medical Genetics and Pathology Group, University Hospital of Basel, Basel, Switzerland
[§]Computational and Translational Pathology Group, Department of Biomedical Engineering, University of Basel, Basel, Switzerland
[¶]Equal supervision.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: 2nd Workshop on Multi-modal Foundation Models and Large Language Models for Life Sciences.

high error rates, and substantial costs [5]. Achieving true single-cell, whole-transcriptome ST would enable precise cell annotations and deeper insight into biological mechanisms and interactions.

Meanwhile, spot-level Visium data are rapidly accumulating in public repositories [6, 7, 8] and supporting large-scale cohorts (e.g., 7,000 patients in MOSAIC [9]), highlighting the need for robust deconvolution methods to reconstruct single-cell transcriptomes.

Recently, advances in deep learning demonstrated that hematoxylin and eosin (H&E)-stained histological images can be used to effectively predict ST profiles [10, 11, 12, 13, 14]. These methods represent a promising, cost-effective, and scalable alternative to conventional sequencing techniques. Building on this, early studies explored the prediction of super-resolution transcriptomic data [15, 16], which produce superpixel-level expression maps rather than precise cell-level profiles. Despite these advances, achieving true single-cell transcriptomic resolution remains a major challenge.

To this end, we present DeepSpot2Cell, a novel multi-modal deep learning model that leverages recent pathology foundation models alongside spatial multi-level context to accurately predict virtual single-cell gene expression from H&E images using spot-level supervision. Built on a permutation-invariant DeepSet architecture, DeepSpot2Cell represents each spot as a bag of cells, learns cell-level contributions during training, and performs single-cell prediction at inference. This enables robust mappings from histology to single-cell transcriptomes using abundant multi-modal spot-level ST.

We evaluate the model on a newly curated benchmark designed to test two key tasks: deconvolving retrospective ST datasets and predicting single-cell expression from unseen H&E images. The benchmark spans 29 cancer samples (20 lung, 7 breast, 2 pancreatic) profiled with 10x Xenium and structured to mimic Visium spot-level data. DeepSpot2Cell substantially outperforms prior super-resolution models in reconstructing single-cell transcriptomes, with consistent performance across in-sample, out-of-sample, and out-of-distribution settings.

To our knowledge, DeepSpot2Cell is the first model to leverage pathology foundation models for virtual single-cell ST prediction from H&E images using spot-level transcriptomic supervision.

2 Related Works

2.1 Pathology foundation models

Pathology foundation models (PFM) are trained on large-scale histopathology datasets using self-supervised techniques such as contrastive learning or masked image modeling. Notable examples include UNI [17], Phikon-v2 [18], and H-Optimus-0 [19], which mostly rely on vision transformers (ViT) to learn high-dimensional morphological representations. These models achieve state-of-the-art performance across a range of computational pathology tasks [20, 21].

2.2 Spatial transcriptomics prediction from H&E images

ST sequencing methods generate spatially resolved transcriptomic profiles aligned with H&E images. For example, on the 10x Genomics Visium platform, each spot covers 55µm of tissue area and captures transcripts from 1-10 cells depending on the cell size [4].

With the growth of such molecular datasets [6, 7, 8], specialized machine-learning models have been developed to predict ST from H&E images using CNNs [10], vision transformers [12, 13], or contrastive learning [11]. The recent DeepSpot model [14] advances this by employing a PFM for spot representations and integrating spatial multi-level context, enabling prediction of 5,000 genes with markedly higher accuracy and up to six-fold greater coverage than prior models.

2.3 Super-resolution-based deconvolution models from H&E images

Super-resolution methods aim to improve the spatial resolution of ST by integrating H&E images. For example, iStar [22] uses hierarchical vision transformers to extract histology features at a 16×16-pixel scale, capturing fine-grained tissue characteristics. scstGCN [23] combines graph convolutional networks with PFM and spatial information to capture the relationships among adjacent superpixels. However, these methods output high-dimensional superpixel expression maps rather than precise cell transcriptomic profiles, requiring custom post-processing to approximate individual cell expression.

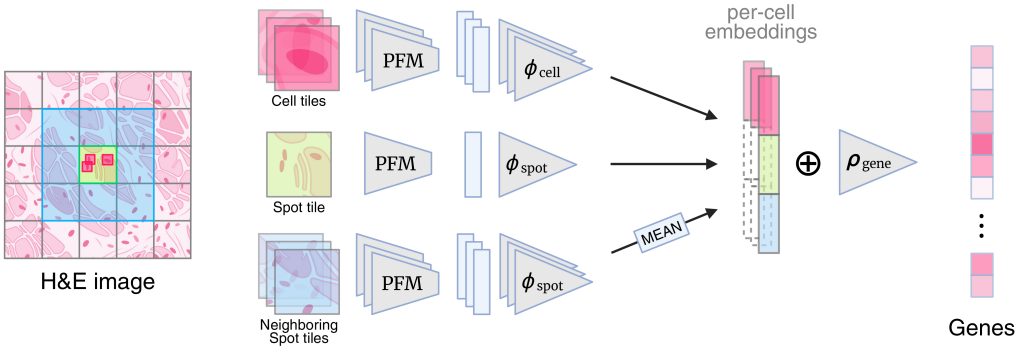

Figure 1: DeepSpot2Cell predicts virtual single-cell spatial transcriptomics as follows: (1) During training, the model takes as input (i) the cropped cell tile defined by the segmentation mask, (ii) the full spot tile containing the cell, and (iii) the neighboring spot tile(s). All tiles are first processed through a pathology foundation model (PFM) before being used to train the model to regress spot-level gene expression; (2) During inference, the model takes as input only the cell tile of interest along with (ii) and (iii), again after PFM processing, and predicts the virtual transcriptomic profile at the cell level.

3 Methods

3.1 Model architecture

DeepSpot2Cell extends the DeepSets architecture [24, 14] to integrate spatial multi-tissue context from histopathology images for accurate cell expression prediction using spot-level transcriptomic supervision. As illustrated in Figure 1, for each cell j within spot i, the model extracts features from three H&E image inputs using frozen PFM: (i) the cropped cell tile defined by the segmentation mask $\mathbf{x}_j^{\text{cell}}$, (ii) the full spot tile containing the cell $\mathbf{x}_i^{\text{spot}}$, and (iii) the neighboring spot tile(s) $\mathbf{x}_i^{\text{neighbor}}$.

For each cell, PFM embeddings are processed via dedicated two-layer multilayer perceptrons (MLP): ϕ_{cell} for cell tiles, and ϕ_{spot} for spot and neighboring contexts. These embeddings are concatenated to form the integrated cell representation:

$$\mathbf{h}_j = \text{Concat}\big(\phi_{\text{cell}}(\mathbf{x}_j^{\text{cell}}),\ \phi_{\text{spot}}(\mathbf{x}_i^{\text{spot}}),\ \phi_{\text{spot}}(\mathbf{x}_i^{\text{neighbor}})\big).$$

Cell embeddings within each spot are aggregated by summation, ensuring permutation invariance and accommodating variable cell counts within a spot:

$$\hat{\mathbf{s}}_i = \rho_{\text{gene}}\big(\textstyle\sum_{j\in\mathcal{C}_i} \mathbf{h}_j\big),$$

where ρ_{gene} is a two-layer MLP gene prediction head generating the predicted gene expression vector $\hat{\mathbf{s}}_i \in \mathbb{R}^G$ for spot i, and $\mathcal{C}_i$ denotes the set of cells within that spot. Notably, the summation aggregation naturally models transcript count additivity and ensures robustness to cell order permutations.

3.2 Benchmarking of virtual cell transcriptomic profiles inferred from H&E images

10x Visium measures gene expression at the spot level, but single-cell resolution is needed for accurate evaluation of prediction methods. To address this, we gathered 10x Xenium datasets across multiple cancer and tissue types (Table 2, Figure 3) with true single-cell resolution to establish a newly dedicated benchmark. Building upon the HEST-1k benchmark [20], we derived pseudo spot-level gene expression profiles by aggregating single-cell transcript counts within each 55µm spatial spot, consistent with 10x Visium. To account for cell size variability [25], a cell was considered fully contained within a spot if its nucleus was located at least 10µm inside the spot boundary.
Evaluation: Models are trained exclusively on spot-level data, and their performance is assessed by comparing predicted gene expression to single-cell ground truth using per-gene Pearson correlation. Two key evaluation tasks were considered: deconvolution of in-sample (IS) cells in spots seen during

training, and prediction of single-cell expression for out-of-sample (OOS) cells from unseen samples within the cohort, and out-of-distribution (OOD) cells from samples belonging to a different cohort.

4 Experiments

4.1 DeepSpot2Cell enables cell expression deconvolution from spatial transcriptomics spots

Table 1 summarizes the performance of a two-layer MLP baseline, previous super-resolution-based models (iStart and scstGCN), and DeepSpot2Cell in deconvolving spot-based ST from H&E images. The in-sample (IS) scenario benchmarks deconvolution performance. IS is further subdivided into IS_{in}, assessing gene correlation among cells within spots, and IS_{out}, assessing cells outside spots.

DeepSpot2Cell substantially improved the single-cell gene expression deconvolution across the three cancer datasets. For example, in the lung cancer dataset, DeepSpot2Cell increased the IS_{in} Pearson correlation across the top 100 genes by 22%, from 0.32 (best competitor, scstGCN) to 0.39, with some genes exceeding a correlation of 0.50 (Figure 5). Notably, DeepSpot2Cell's transcriptomic predictions for cells located within spots (IS_{in}) and outside of spots (IS_{out}) are similarly accurate in lung and breast cancer, indicating that the model does not overfit to the spot-level signals (Table 3, 4).

Table 1: Benchmark of single-cell expression prediction across lung, breast, and pancreatic cancer datasets. Average Pearson correlation between predicted and ground-truth single-cell gene expression is reported for the top 100 most predictive genes.

	Lung cancer (n=20)			Breast cancer (n=7)				Pancreatic cancer (n=2)		
Model	IS_{in}	IS_{out}	OOS	IS_{in}	IS_{out}	OOS	OOD	IS_{in}	IS_{out}	OOS
MLP	0.20 (0.01)	0.24 (0.00)	0.19 (0.00)	0.30 (0.01)	<u>0.34</u> (0.01)	0.14 (0.00)	<u>0.25</u> (0.01)	0.18 (0.00)	0.09 (0.00)	<u>0.10</u> (0.00)
iStar	0.28 (0.01)	0.28 (0.00)	0.15 (0.00)	0.34 (0.01)	0.25 (0.01)	0.10 (0.00)	0.17 (0.00)	0.28 (0.01)	0.13 (0.00)	<u>0.10</u> (0.00)
scstGCN	<u>0.32</u> (0.01)	<u>0.35</u> (0.01)	<u>0.24</u> (0.01)	<u>0.37</u> (0.01)	0.33 (0.01)	<u>0.17</u> (0.00)	0.24 (0.01)	<u>0.29</u> (0.01)	<u>0.14</u> (0.00)	0.08 (0.00)
DeepSpot2Cell (ours)	**0.39** (0.01)	**0.41** (0.01)	**0.35** (0.01)	**0.43** (0.01)	**0.41** (0.01)	**0.28** (0.01)	**0.37** (0.01)	**0.32** (0.01)	**0.16** (0.00)	**0.11** (0.00)

IS_{in}: gene correlation among cells within spots. IS_{out}: gene correlation among cells outside spots. OOS: gene correlation among cells on hold-out patients, same cohort. OOD: gene correlations on cells from different cohort.

Figure 2 qualitatively illustrates the deconvolution of *MSLN*, a known non-small cell lung cancer marker gene, on slide NCBI867 from the lung cancer dataset. Predictions from iStar ($r = 0.21$) and scstGCN ($r = 0.30$) are noisy and spatially diffuse, whereas DeepSpot2Cell ($r = 0.45$) produces coherent, spatially structured patterns that align better with the ground truth.

4.2 DeepSpot2Cell predicts virtual single-cell spatial transcriptomics from H&E images

Furthermore, the out-of-sample (OOS) and out-of-distribution (OOD) scenarios assess a model's ability to infer virtual single-cell gene expression for samples unseen during training, simulating its application to novel data from the same or a different cohort, respectively (Table 1).

For example, in the lung cancer OOS scenario, DeepSpot2Cell consistently increased the Pearson correlation by 46%, from 0.24 (best competitor, scstGCN) to 0.35 (DeepSpot2Cell). In the more challenging breast cancer OOD scenario, DeepSpot2Cell improved the gene correlations by more than 50%, demonstrating that it has learned robust single-cell transcriptomic mappings that generalize to unseen histopathological images from other cohorts.

4.3 DeepSpot2Cell ablation experiments

Next, we evaluated how specific modeling choices in DeepSpot2Cell contribute to its accuracy in single-cell prediction (Figure 4). We make two important observations: **1)** Leveraging spatial multi-level tissue context through both spot representations and their neighbors improves DeepSpot2Cell's

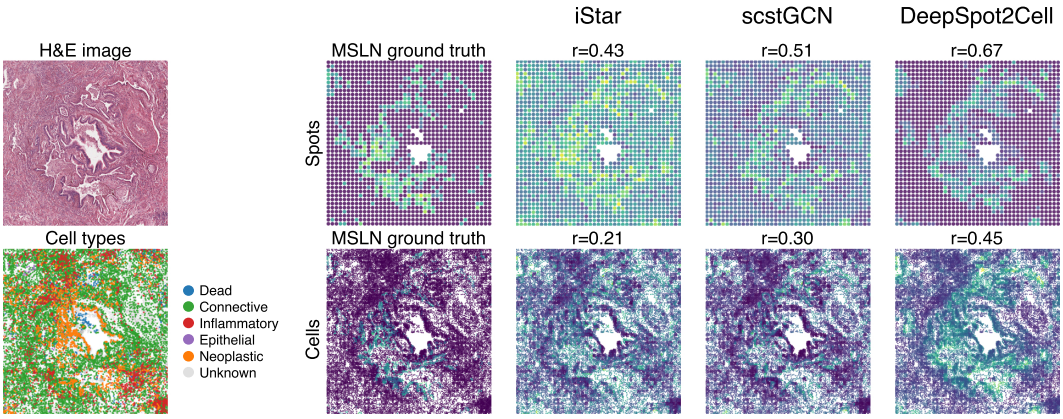

Figure 2: H&E image and CellViT [26] cell-type annotations for slide NCBI867 (lung dataset). Annotations and model predictions of *MSLN* expression across spots and cells.

gene correlations compared with using only the spot and cell or only the cell itself, consistent with previous findings [27]. **2)** The choice of PFM is important, with Phikon-v2 outperforming UNI and H-Optimus-0, potentially due to Phikon-v2's multi-resolution training design. **3)** A more advanced GRU network for learning the set convolution operation performs worse than simple summation, likely due to cell order sensitivity.

5 Discussion & Conclusion

In this work, we propose DeepSpot2Cell, a novel DeepSet neural network that leverages PFM and spatial multi-level tissue context to accurately infer virtual single-cell ST from routine histology images using spot-level supervision. The method's key innovation is modeling spots as bags of cells: DeepSpot2Cell learns how individual cells contribute to spot-level gene expression and uses these mappings to predict single-cell expression. Further, we curated a newly dedicated benchmark designed for single-cell ST deconvolution and prediction, enabling systematic comparison of models.

Our results demonstrate that DeepSpot2Cell outperforms previous super-resolution models in single-cell deconvolution and prediction across multiple cancer types, even in out-of-distribution settings. The variable performance of other super-resolution models across different scenarios indicates overfitting to the specific images and transcriptomic spots. In contrast, DeepSpot2Cell gene correlations were consistent, indicating that it has learned general single-cell mappings that could be transferred to unseen images. Notably, both iStar and scstGCN underperformed on OOD breast cancer compared to our MLP baseline, highlighting their limited ability to generalize beyond the training distribution.

In summary, DeepSpot2Cell uses the abundance of spot-based ST both to augment existing multimodal ST cohorts with cell-level resolution and to learn generalizable single-cell transcriptomic mappings, enabling the prediction of single-cell expression profiles from H&E images.

6 Limitations & Future work

Several limitations merit acknowledgment. First, the experiments used pseudo-Visium spots derived from Xenium data, which may not fully reflect real Visium measurements. We also relied on available old Xenium datasets with $\sim$300 genes rather than the newer 5k panels. Second, accuracy depends on the quality of cell segmentation, as errors in segmentation propagate to expression assignments. In our benchmark, we relied on Xenium ground-truth nuclei locations, but in practice, these must be inferred computationally, requiring accurate nucleus detection. Improving and integrating these methods is essential for single-cell resolution in real-world settings. Finally, this work motivates further research on the utility of the virtual cells in downstream biological applications, including identifying genes that correlate with tissue architecture and those that cannot be reliably predicted.

References

[1] Qichao Yu, Miaomiao Jiang, and Liang Wu. Spatial transcriptomics technology in cancer research. *Frontiers in Oncology*, 12:1019111, 2022.

[2] Marco De Zuani, Haoliang Xue, Jun Sung Park, Stefan C Dentro, Zaira Seferbekova, Julien Tessier, Sandra Curras-Alonso, Angela Hadjipanayis, Emmanouil I Athanasiadis, Moritz Gerstung, et al. Single-cell and spatial transcriptomics analysis of non-small cell lung cancer. *Nature communications*, 15(1):4388, 2024.

[3] Lambda Moses and Lior Pachter. Museum of spatial transcriptomics. *Nature methods*, 19(5): 534–546, 2022.

[4] 10X Genomics. How many cells are captured in a single spot?, n.d. URL `https://kb.10xgenomics.com/hc/en-us/articles/360035487952-How-many-cells-are-captured-in-a-single-spot`. Accessed: 2025-08-26.

[5] Yixing Dong, Chiara Saglietti, Quentin Bayard, Almudena Espin Perez, Sabrina Carpentier, Daria Buszta, Stephanie Tissot, Rémy Dubois, Atanas Kamburov, Senbai Kang, et al. Transcriptome analysis of archived tumors by visium, geomx dsp, and chromium reveals patient heterogeneity. *Nature communications*, 16(1):4400, 2025.

[6] Zhicheng Xu, Weiwen Wang, Tao Yang, Ling Li, Xizheng Ma, Jing Chen, Jieyu Wang, Yan Huang, Joshua Gould, Huifang Lu, et al. Stomicsdb: a comprehensive database for spatial transcriptomics data sharing, analysis and visualization. *Nucleic acids research*, 52(D1):D1053–D1061, 2024.

[7] Zhen Fan, Runsheng Chen, and Xiaowei Chen. Spatialdb: a database for spatially resolved transcriptomes. *Nucleic acids research*, 48(D1):D233–D237, 2020.

[8] Guoliang Wang, Song Wu, Zhuang Xiong, Hongzhu Qu, Xiangdong Fang, and Yiming Bao. Crost: a comprehensive repository of spatial transcriptomics. *Nucleic Acids Research*, 52(D1): D882–D890, 2024.

[9] MOSAIC Consortium and Caroline Hoffmann. Mosaic: Intra-tumoral heterogeneity characterization through large-scale spatial and cell-resolved multi-omics profiling. *bioRxiv*, pages 2025–05, 2025.

[10] Bryan He, Ludvig Bergenstråhle, Linnea Stenbeck, Abubakar Abid, Alma Andersson, Åke Borg, Jonas Maaskola, Joakim Lundeberg, and James Zou. Integrating spatial gene expression and breast tumour morphology via deep learning. *Nature biomedical engineering*, 4(8):827–834, 2020.

[11] Ronald Xie, Kuan Pang, Sai Chung, Catia Perciani, Sonya MacParland, Bo Wang, and Gary Bader. Spatially resolved gene expression prediction from histology images via bi-modal contrastive learning. *Advances in Neural Information Processing Systems*, 36:70626–70637, 2023.

[12] Yuansong Zeng, Zhuoyi Wei, Weijiang Yu, Rui Yin, Yuchen Yuan, Bingling Li, Zhonghui Tang, Yutong Lu, and Yuedong Yang. Spatial transcriptomics prediction from histology jointly through transformer and graph neural networks. *Briefings in Bioinformatics*, 23(5), 2022.

[13] Yuran Jia, Junliang Liu, Li Chen, Tianyi Zhao, and Yadong Wang. Thitogene: a deep learning method for predicting spatial transcriptomics from histological images. *Briefings in Bioinformatics*, 25(1), 2023.

[14] Kalin Nonchev, Sebastian Dawo, Karina Silina, Holger Moch, Sonali Andani, Tumor Profiler Consortium, Viktor H Koelzer, and Gunnar Rätsch. Deepspot: Leveraging spatial context for enhanced spatial transcriptomics prediction from h&e images. *medRxiv*, pages 2025–02, 2025.

[15] Daiwei Zhang, Amelia Schroeder, Hanying Yan, Haochen Yang, Jian Hu, Michelle YY Lee, Kyung S Cho, Katalin Susztak, George X Xu, Michael D Feldman, et al. Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology. *Nature biotechnology*, 42(9):1372–1377, 2024.

[16] Shuailin Xue, Fangfang Zhu, Jinyu Chen, and Wenwen Min. Inferring single-cell resolution spatial gene expression via fusing spot-based spatial transcriptomics, location, and histology using gcn. *Briefings in Bioinformatics*, 26(1), 2024.

[17] Richard J Chen, Tong Ding, Ming Y Lu, Drew FK Williamson, Guillaume Jaume, Andrew H Song, Bowen Chen, Andrew Zhang, Daniel Shao, Muhammad Shaban, et al. Towards a general-purpose foundation model for computational pathology. *Nature medicine*, 30(3):850–862,

2024.

[18] Alexandre Filiot, Paul Jacob, Alice Mac Kain, and Charlie Saillard. Phikon-v2, a large and public feature extractor for biomarker prediction. *arXiv preprint arXiv:2409.09173*, 2024.

[19] Charlie Saillard, Rodolphe Jenatton, Felipe Llinares-López, Zelda Mariet, David Cahané, Eric Durand, and Jean-Philippe Vert. H-optimus-0, 2024. URL https://github.com/bioptimus/releases/tree/main/models/h-optimus/v0.

[20] Guillaume Jaume, Paul Doucet, Andrew Song, Ming Yang Lu, Cristina Almagro Pérez, Sophia Wagner, Anurag Vaidya, Richard Chen, Drew Williamson, Ahrong Kim, et al. Hest-1k: A dataset for spatial transcriptomics and histology image analysis. *Advances in Neural Information Processing Systems*, 37:53798–53833, 2024.

[21] Ioannis Gatopoulos, Nicolas Känzig, Roman Moser, Sebastian Otálora, et al. eva: Evaluation framework for pathology foundation models. In *Medical Imaging with Deep Learning*, 2024.

[22] Daiwei Zhang, Amelia Schroeder, Hanying Yan, Haochen Yang, Jian Hu, Michelle Y. Y. Lee, Kyung S. Cho, Katalin Susztak, George X. Xu, Michael D. Feldman, Edward B. Lee, Emma E. Furth, Linghua Wang, and Mingyao Li. Inferring super-resolution tissue architecture by integrating spatial transcriptomics with histology. *Nature Biotechnology*, pages 1–6, 2024.

[23] Shuailin Xue, Fangfang Zhu, Jinyu Chen, and Wenwen Min. Inferring single-cell resolution spatial gene expression via fusing spot-based spatial transcriptomics, location, and histology using gcn. *Briefings in Bioinformatics*, 26(1), 2025.

[24] Manzil Zaheer, Satwik Kottur, Siamak Ravanbakhsh, Barnabas Poczos, Russ R Salakhutdinov, and Alexander J Smola. Deep sets. *Advances in neural information processing systems*, 30, 2017.

[25] Bruce Alberts, Dennis Bray, Karen Hopkin, Alexander D Johnson, Julian Lewis, Martin Raff, Keith Roberts, and Peter Walter. *Essential cell biology*. Garland Science, 2015.

[26] Fabian Hörst, Moritz Rempe, Lukas Heine, Constantin Seibold, Julius Keyl, Giulia Baldini, Selma Ugurel, Jens Siveke, Barbara Grünwald, Jan Egger, and Jens Kleesiek. Cellvit: Vision transformers for precise cell segmentation and classification. *Medical Image Analysis*, 94:103143, 2024. URL https://www.sciencedirect.com/science/article/pii/S1361841524000689.

[27] Kalin Nonchev, Sonali Andani, Joanna Ficek-Pascual, Marta Nowak, Bettina Sobottka, Tumor Profiler Consortium, Viktor H Koelzer, and Gunnar Rätsch. Representation learning for multi-modal spatially resolved transcriptomics data. *medRxiv*, pages 2024–06, 2024.

[28] Amanda Janesick, Robert Shelansky, Andrew D. Gottscho, Florian Wagner, Stephen R. Williams, Morgane Rouault, Ghezal Beliakoff, Carolyn A. Morrison, Michelli F. Oliveira, Jordan T. Sicherman, Andrew Kohlway, Jawad Abousoud, Tingsheng Yu Drennon, Seayar H. Mohabbat, and Sarah E.B. Taylor. High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. *Nature Communications 2023 14:1*, pages 1–15, 2023.

[29] 10x Genomics. Ffpe human breast using the entire sample area, 2023. URL https://www.10xgenomics.com/datasets/ffpe-human-breast-using-the-entire-sample-area-1-standard.

[30] 10x Genomics. Ffpe human breast with pre-designed panel, 2023. URL https://www.10xgenomics.com/datasets/ffpe-human-breast-with-pre-designed-panel-1-standard.

[31] 10x Genomics. Pancreatic cancer with xenium human multi-tissue and cancer panel, 2024. URL https://www.10xgenomics.com/datasets/pancreatic-cancer-with-xenium-human-multi-tissue-and-cancer-panel-1-standard.

[32] 10x Genomics. Ffpe human pancreas with xenium multimodal cell segmentation, 2024. URL https://www.10xgenomics.com/datasets/ffpe-human-pancreas-with-xenium-multimodal-cell-segmentation-1-standard.

[33] Annika Vannan, Ruqian Lyu, Arianna L. Williams, Nicholas M. Negretti, Evan D. Mee, Joseph Hirsh, Samuel Hirsh, David S. Nichols, Carla L. Calvi, Chase J. Taylor, Vasiliy. V. Polosukhin, Ana PM Serezani, A. Scott McCall, Jason J. Gokey, Heejung Shim, Lorraine B. Ware, Matthew J. Bacchetta, Ciara M. Shaver, Timothy S. Blackwell, Rajat Walia, Jennifer MS Sucre, Jonathan A. Kropski, Davis J McCarthy, and Nicholas E. Banovich. Image-based spatial transcriptomics identifies molecular niche dysregulation associated with distal lung remodeling in pulmonary fibrosis. *bioRxiv*, 2023.

[34] Si-Jie Hao, Yuan Wan, Yi-Qiu Xia, Xin Zou, and Si-Yang Zheng. Size-based separation methods of circulating tumor cells. *Advanced drug delivery reviews*, 125:3–20, 2018.

A DeepSpot2Cell training details

DeepSpot2Cell was trained on an NVIDIA RTX 4090 using the Adam optimizer with a learning rate of 10^{-4} and a batch size of 256 spots. Early stopping was used based on validation loss. Dropout with rate 0.3 was applied to the ϕ and ρ_{gene} MLPs to reduce overfitting. We optimize the model by minimizing the mean squared error (MSE) loss between predicted and observed spot expressions:

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^{N} \|\hat{\mathbf{s}}_i - \mathbf{s}_i\|_2^2 ,$$

where N is the number of spots. The code is available at https://github.com/ratschlab/DeepSpot.

B Pathology foundation models

Tile embeddings were extracted from pretrained pathology foundation models (PFM), with their weights obtained from Hugging Face.

1. UNI: https://huggingface.co/MahmoodLab/UNI
2. Phikon v2: https://huggingface.co/owkin/phikon-v2
3. H-optimus-0: https://huggingface.co/bioptimus/H-optimus-0

We benchmarked their performance to assess their contribution and found that Phikon v2 produced more accurate expression predictions relative to the ground truth (Figure 4). To isolate the effect of the PFM in our benchmarks, we kept Phikon v2 fixed as the underlying pathology model in DeepSpot2Cell and scstGCN, ensuring fair comparability. Notably, these models are PFM-agnostic in practice, allowing the pathology foundation model to be replaced with a more suitable one depending on tissue characteristics.

C MLP baseline

The MLP baseline is a two-layer network designed to isolate the contributions of DeepSpot2Cell's core components: (1) spatial multi-level context integration, and (2) the DeepSets architecture, which handles variable numbers of cells per spot. This baseline provides a direct strategy for inferring single-cell expression from H&E images using PFM features, serving as a reference for evaluating the trade-off between architectural complexity and predictive performance.

Training followed the procedure described in Appendix A, with the MLP optimized to predict spot-level expression from spot-tile PFM features. During inference, cell-level PFM features were provided, and the outputs were interpreted as cell-level predictions.

D Super-resolution models

Super-resolution methods were trained using default hyperparameters from their respective official implementations.

D.1 Code availability

1. scstGCN: https://github.com/wenwenmin/scstGCN
2. iStar: https://github.com/daviddaiweizhang/istar

D.2 Superpixel map expression post-processing details

These methods generate continuous high-dimensional expression grids corresponding to the original patch regions: iStar produces 16×16 node grids while scstGCN outputs 14×14 grids. During evaluation, to obtain cell-level predictions from the grid outputs, cell bounding boxes were manually downscaled to grid coordinates and intersected with the grid nodes. Then, cell expression values were computed as the average of all nodes intersecting with each cell's downscaled bounding box. The resulting cell-level predictions were then normalized to 10,000 counts per spot, followed by log1p transformation.

E Data

We utilized organ-specific Xenium datasets, focusing on three cancer types for which sufficiently high-quality samples were available: 20 lung [28], 7 breast [29, 30], and 2 pancreatic [31, 32] cancer samples (Table 2). We downloaded the datasets using the HEST-1k preprocessing pipeline [20].

Table 2: HEST-1k datasets used in this study.

Organ	Split	Dataset	Samples	Spots	Cells
Lung	Training	Image-based spatial transcriptomics identifies molecular niche dysregulation associated with distal lung remodeling in pulmonary fibrosis [33]	20	76,023	998,856
Breast	Training	High-resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in-situ analysis [28]	3	26,568	415,320
	Training	FFPE human breast using the entire sample area [29]	2	177,885	1,766,768
	OOD Eval.	FFPE human breast with pre-designed panel [30]	2	65,944	1,144,523
Pancreas	Training	Pancreatic cancer with Xenium human multi-tissue and cancer panel [31]	1	20,842	189,736
	Training	FFPE human pancreas with Xenium multimodal cell segmentation [32]	1	5,827	139,581

F Data preprocessing

F.1 Pseudospots definition

10x Visium spots contain a tissue area of 55µm, capturing between 1-10 cells, depending on the cell size [4]. To account for the cell size variability [25], cells were considered inside the spot if their nucleus fell at least 10µm within the spot boundary, otherwise considered outside the spot. While this approach may be suboptimal—since cancer cells are generally larger [34] and their membranes could extend beyond the spot boundary—we calculated with this setup that the average number of cells per spot across the three datasets to be between 1-10 cells (Figure 3), which alligns with 10x Visium reported characteristics [4].

Specifically, H&E images were available at 20x magnification and were divided into non-overlapping 224×224 pixel tiles. Each tile contained a central circular pseudospot with a 160-pixel diameter, and spot centroids were spaced 224 pixels apart. Spot-level expression was computed as the sum of all cells located within the pseudospot. The distribution of cell counts per spot across all datasets is shown in Figure 3.

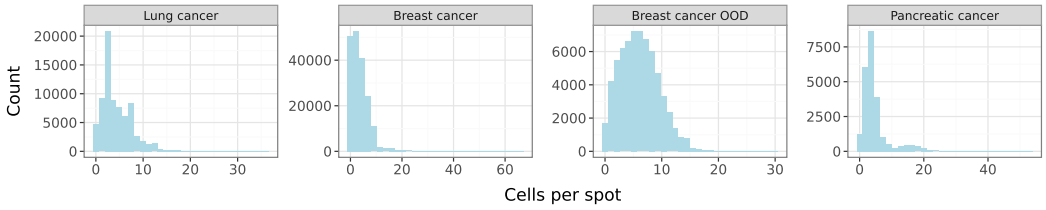

Figure 3: Distribution of cell counts per spot across datasets.

F.2 Gene expression preprocessing and quality control

Gene counts preprocessing followed a standardized pipeline. Genes expressed in fewer than 20 cells across the sample, as well as blank and negative control genes, were removed. For normalization, spot-level counts were scaled to sum to 10,000 transcripts per spot and subsequently log1p-transformed. Cell-level normalization was performed based on spatial context (inside or outside a spot): cells within a spot were normalized to sum to 10,000 counts, whereas outside-spot cells were normalized using the total counts of the inside-spot cells from the corresponding spot. This strategy ensured that outside-spot cells did not affect the normalization of inside-spot cells, while still undergoing consistent preprocessing.

F.3 Feature extraction

Individual cell tiles were defined as the smallest square that fully contains the segmented area of the cell, which was obtained by CellViT [26] and was provided with the HEST-1k dataset. Cell and spot tiles were transformed in accordance with the recommended preprocessing of each particular pathology foundation model.

G Evaluation details

Model performance was assessed using per-gene Pearson correlation.

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}}$$

where x_i denotes the predicted value for gene i, y_i denotes the observed value for gene i, $\bar{x}$ and $\bar{y}$ are the mean predicted and observed values, respectively, and n is the number of cells.

The Pearson correlation measures how well the predicted values agree with the observed values across samples for each gene. Cross-validation employed patient-level data partitioning to ensure validation splits contained only samples from distinct patients. The number of folds was set to *min(5, n_patients)*. We bootstrapped 10,000 times from the median Pearson correlation across the test folds and reported the resulting median Pearson correlation along with its standard error.

H Ablation details

All ablations were compared on the lung cancer dataset using the area under the Pearson correlation gene curve computed on cells from within spots (IS_{in}), as shown in Figure 4.

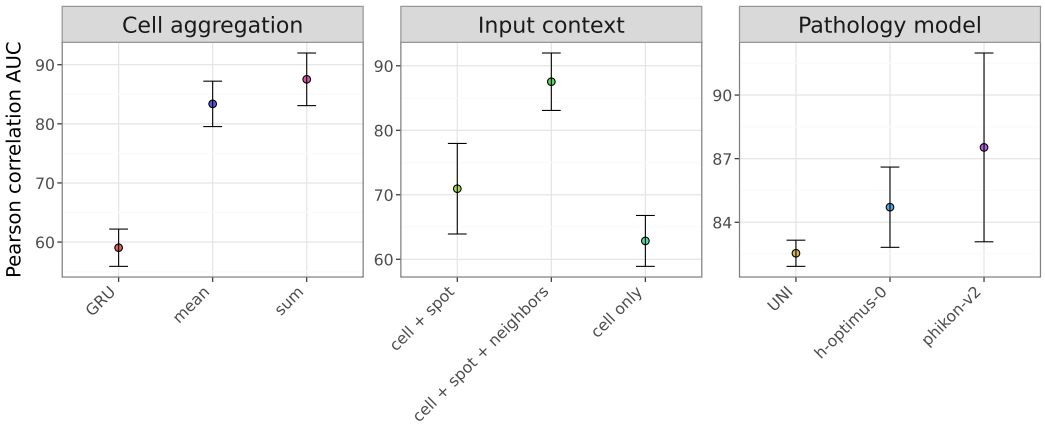

Figure 4: Different DeepSpot2Cell components compared based on the area under the Pearson correlation gene curve computed on the cells from within spots (IS_{in}).

I Extended evaluation results

Figure 2 qualitatively illustrates the deconvolution of *MSLN*, a known non-small cell lung cancer marker gene, on slide NCBI867 from the lung cancer dataset. Predictions from iStar ($r = 0.21$) and scstGCN ($r = 0.30$) are noisy and spatially diffuse, whereas DeepSpot2Cell ($r = 0.45$) produces coherent, spatially structured patterns that align better with the ground truth.

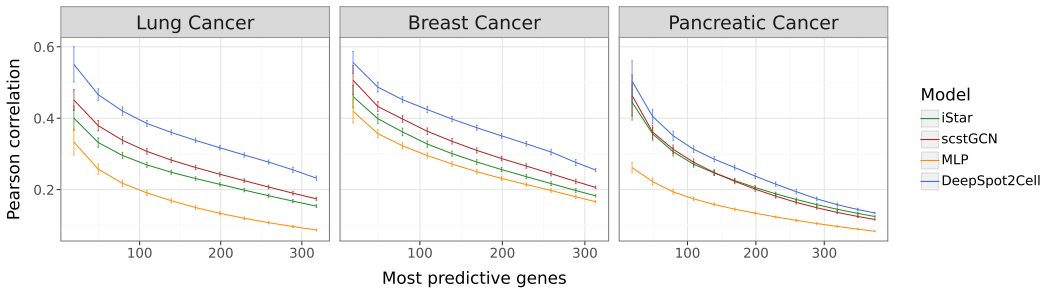

Figure 5: Deconvolution benchmark across lung, breast, and pancreatic cancer datasets. Sets of X most predictive genes for each model on the x-axis are sorted by the descending Pearson correlation on the y-axis. Correlations computed per-gene on the cells from within spots (IS_{in}).

Table 3: Benchmark of single-cell expression prediction across lung, breast, and pancreatic cancer datasets. Average Pearson correlation between predicted and ground-truth single-cell gene expression is reported for the top 50 most predictive genes.

	Lung cancer (n=20)			Breast cancer (n=7)				Pancreatic cancer (n=2)		
Model	IS_{in}	IS_{out}	OOS	IS_{in}	IS_{out}	OOS	OOD	IS_{in}	IS_{out}	OOS
MLP	0.26 (0.01)	0.29 (0.01)	0.24 (0.01)	0.36 (0.01)	<u>0.40</u> (0.01)	0.19 (0.01)	<u>0.33</u> (0.01)	0.22 (0.01)	0.11 (0.00)	**0.15** (0.01)
iStar	0.33 (0.01)	0.33 (0.01)	0.19 (0.01)	0.40 (0.01)	0.31 (0.01)	0.13 (0.01)	0.23 (0.01)	0.35 (0.01)	0.17 (0.01)	**0.15** (0.01)
scstGCN	<u>0.38</u> (0.01)	<u>0.41</u> (0.01)	<u>0.29</u> (0.01)	<u>0.43</u> (0.01)	0.39 (0.01)	<u>0.22</u> (0.01)	0.31 (0.01)	<u>0.36</u> (0.01)	<u>0.19</u> (0.01)	0.11 (0.01)
DeepSpot2Cell (ours)	**0.46** (0.01)	**0.47** (0.01)	**0.41** (0.01)	**0.49** (0.01)	**0.48** (0.01)	**0.35** (0.01)	**0.46** (0.01)	**0.40** (0.01)	**0.21** (0.01)	**0.15** (0.01)

Table 4: Benchmark of single-cell expression prediction across lung, breast, and pancreatic cancer datasets. Average Pearson correlation between predicted and ground-truth single-cell gene expression is reported for the top 200 most predictive genes.

	Lung cancer (n=20)			Breast cancer (n=7)				Pancreatic cancer (n=2)		
Model	IS_{in}	IS_{out}	OOS	IS_{in}	IS_{out}	OOS	OOD	IS_{in}	IS_{out}	OOS
MLP	0.13 (0.00)	0.17 (0.00)	0.14 (0.00)	0.23 (0.00)	<u>0.25</u> (0.00)	0.09 (0.00)	<u>0.17</u> (0.00)	0.13 (0.00)	0.06 (0.00)	**0.07** (0.00)
iStar	0.21 (0.00)	0.22 (0.00)	0.11 (0.00)	0.26 (0.00)	0.17 (0.00)	0.05 (0.00)	0.11 (0.00)	<u>0.21</u> (0.00)	0.08 (0.00)	<u>0.06</u> (0.00)
scstGCN	<u>0.24</u> (0.00)	<u>0.28</u> (0.00)	<u>0.18</u> (0.00)	<u>0.29</u> (0.00)	0.24 (0.00)	<u>0.12</u> (0.00)	0.16 (0.01)	0.20 (0.00)	<u>0.09</u> (0.00)	0.04 (0.00)
DeepSpot2Cell (ours)	**0.32** (0.00)	**0.33** (0.00)	**0.29** (0.00)	**0.35** (0.00)	**0.32** (0.00)	**0.19** (0.00)	**0.26** (0.01)	**0.24** (0.00)	**0.10** (0.00)	<u>0.06</u> (0.00)

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes] , [No] , or [NA] .
- [NA] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for NA).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will be asked to also include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While "[Yes] " is generally preferable to "[No] ", it is perfectly acceptable to answer "[No] " provided a proper justification is given (e.g., "error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering "[No] " or "[NA] " is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: **[TODO]**

   Guidelines: Please refer to the abstract and introduction part

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: Please refer to Limitations & Future work section.

Guidelines:

- The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [NA]

   Justification: The paper does not involve any theoretical results.

   Guidelines:

   - The answer NA means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We show fundamental experiment settings in Section 3.2, and more details for experiment settings in Appendix F and G. Besides, we provide the complete source code (model architecture, model training and reproducability pipelines) as supplementary materials https://anonymous.4open.science/r/DeepSpot2Cell-9CC1/.

   Guidelines:

- The answer NA means that the paper does not include experiments.
- If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [Yes]

   Justification: We provide all code to initialize and train the model, along with reproducibility code to run the benchmarking and notebooks tutorials https://anonymous.4open.science/r/DeepSpot2Cell-9CC1/.

   Guidelines:

   - The answer NA means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.

- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: Please refer to the Methods section and Appendix F and G.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [Yes]

   Justification: We used bootstrapping to compute standard errors.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
   - If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: For our experiments, we used an NVIDIA RTX 4090 with 24 GB of VRAM and 120 cores on a university-based high-performance cluster.

   Guidelines:

   - The answer NA means that the paper does not include experiments.

- The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We make sure the research conducted in the paper conforms, in every respect, with the NeurIPS Code of Ethics.

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: Although the work proposes potentially impactful methods, it remains at an early stage, and its wider implications and societal effects are difficult to evaluate.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.
    - If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

Answer: [NA]

Justification: We make use of existing deanonymized publicly available datasets.

Guidelines:

- The answer NA means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

Answer: [Yes]

Justification: We have cited necessary assets and conduct CC-BY for our codes.

Guidelines:

- The answer NA means that the paper does not use existing assets.
- The authors should cite the original paper that produced the code package or dataset.
- The authors should state which version of the asset is used and, if possible, include a URL.
- The name of the license (e.g., CC-BY 4.0) should be included for each asset.
- For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

Answer: [Yes]

Justification: The code and other supplementary materials are followed with readme and instructions.

Guidelines:

- The answer NA means that the paper does not release new assets.
- Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
- The paper should discuss whether and how consent was obtained from people whose asset is used.
- At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

Answer: [NA]

Justification: No direct interaction with subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [NA]

Justification: No direct interaction with subjects.

Guidelines:

- The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

Answer: [NA]

Justification: It is not required.

Guidelines:

- The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.