TinyH2GNet: Exploring Lightweight Models for Gene Expression Prediction from H&E Images

Maninder Kaur[1], Amit Kumar[1], Ankita Rani[3], Puneet Goyal[1,4], Raghvendra Mall[2], and Sukrit Gupta[3,4⋆]

[1] Department of Computer Science and Engineering, Indian Institute of Technology Ropar, India
[2] Qatar Center for Artificial Intelligence, Qatar Computing Research Institute, Hamad Bin Khalifa University, Doha, Qatar
[3] Department of Biomedical Engineering, Indian Institute of Technology Ropar, India
[4] School of Artificial Intelligence and Data Engineering, Indian Institute of Technology Ropar, India

Abstract. Predicting spatial gene expression from histopathology images has the potential to complement Spatial Transcriptomics (ST) while avoiding the costs and limited throughput associated with ST. Multiple recent methods rely on increasingly complex Deep Learning (DL) models. We hypothesized that high complexity is not essential for obtaining competitive performance for mapping histopathology spots to their corresponding gene expression. To test this hypothesis, we compared the image-to-gene expression prediction performance of several benchmarks on convolutional and transformer-based architectures on Visium and Xenium ST datasets. We did not find one clear winner across all settings. We tested several selected lightweight variants, termed as TinyH2GNet, from one of the benchmarks and show that they achieve comparable performance while using only a small fraction of the original parameters. The TinyH2GNet model size is $\leq$ 1 MB, its inference on a standard CPU is ten times faster than the smallest existing model. This makes TinyH2GNet suitable for efficient image-to-ST prediction on standard CPU hardware without requiring dedicated GPU resources. The results show that increasing model complexity yields diminishing returns and future research should prioritize biological fidelity over architectural scale. All the code for our experiments can be accessed via the Link.

Keywords: Deep Learning · Histopathology · Spatial Transcriptomics · Tissue Profiling · Low-Resource / Constrained Settings

1 Introduction

Spatial Transcriptomics (ST) technologies measure gene expression along with spatial context in tissue sections [4]. This enables the study of tissue architecture

⋆ Corresponding author: sukrit@iitrpr.ac.in

Table 1. Overview of representative DL approaches for spatial gene expression prediction from histology images. The table reports the core architectural components, **convolutional neural networks (CNN)**, **transformers (Trf)**, **graph neural networks (GNN)**, and **graph attention networks (GAT)**, together with the approximate number of learnable parameters (in millions) and the ST platforms used for evaluation. **V** and **X** indicate experiments conducted on **10x Visium** and **10x Xenium** datasets, respectively.

Work	**Architecture**	**# Params**	**Dataset**
ST-Net [5]	CNN	7.8	V
HisToGene [10]	Trf	68.9	V
DeepSpaCE [9]	CNN	120.9	V
BrST-Net [12]	CNN	6.2	V
TCGN [16]	CNN + GNN + Trf	22.9	V
THItoGene [7]	CNN + Trf + GAT	63.6	V
TinyH2GNet (ours)	**CNN**	**0.3**	**V+X**

and disease microenvironment at the molecular level [2]. However, ST remains costly, time-intensive (needs weeks), and needs specialized equipment thereby limiting its routine clinical use [13]. Conversely, hematoxylin and eosin (H&E) stained histology images, routinely collected in hospitals, capture tissue morphological features that reflect underlying molecular states [3]. This has motivated the use of Deep Learning (DL) methods to predict spatial gene expression directly from histology images bypassing ST's cost barrier.

Varied architectures (based on CNN, GNN, and Transformer models) with different complexity (parameter range 6.2 to 120.9 million) were proposed and tested on different ST technologies (refer Table 1). While the primary focus of these methods has been on obtaining more accurate gene expression predictions, two key questions remain unanswered: (a) Given the large gamut of complex models, does prediction performance scale with model complexity?; and (b) For computationally constrained environments, where access to GPUs or even high-end CPUs may be limited, can lightweight models perform as well as the large ones?

We selected these baselines because they cover the main architecture families (CNNs, transformers, and graph-based models) used for image-to-ST prediction. Several recent methods either extend these families or rely on large pretrained pathology backbones. Our goal is not to benchmark every published model, but to study how prediction performance changes across representative architectures with different model sizes under the same training and evaluation protocol.

To address these questions, we conduct a systematic study of how model complexity impacts ST prediction performance. We benchmark the existing models under identical training protocols across three diverse datasets (Hist2ST [14], HER2 breast cancer [1], and 10x Xenium [6]). We find that no single architecture consistently outperforms others across all metrics, models and datasets. In the evaluated breast image-to-ST settings, the smallest CNN backbone achieves per-

formance statistically indistinguishable from much larger CNN and transformer hybrids models, suggesting that architectural sophistication provides diminishing returns. Motivated by this, we select a compact CNN baseline backbone and construct its pruned variants. We systematically evaluate how reducing model capacity from 100% down to 3% of original parameters affects prediction performance and inference time for CPU hardware. Surprisingly, models with just 5-20% of the original parameters give a comparable performance as the full model. The TinyH2GNet model (∼300k parameters) runs on standard hospital CPUs with less than 4GB RAM peak memory usage, achieving speedups ranging from 26× (on HER2) to 103× (on 10x Xenium) in comparison to the most complex model.

To our knowledge, this is the first systematic study demonstrating that sub-million parameter models achieve comparable gene expression prediction performance across varied datasets. Our results provide the first comprehensive accuracy-efficiency Pareto frontier for image-based spatial transcriptomics prediction. The developed TinyH2GNet models provide an efficient alternative for image-to-ST prediction in resource-constrained computational environments.

2 Methods

2.1 Problem Formulation

We formulate spatial gene expression prediction as a supervised multi-output regression problem. Given a training set $\mathcal{D} = \{(\boldsymbol{X}_i, \boldsymbol{y}_i)\}_{i=1}^{N}$ where $\boldsymbol{X} \in \mathbb{R}^{h \times w \times c}$ denotes an input image patch extracted from a tissue section, h and w represent the spatial dimensions and c denotes the number of channels. The corresponding gene expression profile is represented as a vector $\boldsymbol{y} \in \mathbb{R}^p$, where p is the number of genes. The objective is to learn a mapping function

$$f_\Theta : \mathbb{R}^{h \times w \times c} \to \mathbb{R}^p,$$

parameterized by Θ, such that the predicted gene expression $\hat{\boldsymbol{y}} = f_\Theta(\boldsymbol{X})$ is close to the ground-truth expression $\boldsymbol{y}$.

2.2 Model Architecture

The overall network consists of a backbone feature extractor followed by a regression head. The backbone feature extractor computes the feature embeddings from the input image:

$$\boldsymbol{z} = g_\Omega(\boldsymbol{X}),$$

where $g_\Omega(\cdot)$ denotes the backbone network and $\boldsymbol{z} \in \mathbb{R}^D$ is the resulting D dimensional feature embedding. A regression head then maps these features to gene expression values,

$$\hat{\boldsymbol{y}} = h_\Psi(\boldsymbol{z}),$$

where $h_\Psi(\cdot)$ consists of fully connected layers. The complete model is given by

$$f_\Theta(\boldsymbol{X}) = h_\Psi\big(g_\Omega(\boldsymbol{X})\big),$$

with $\Theta = \{\Omega, \Psi\}$. Multiple architectures proposed for the spatial gene expression prediction problem use different backbone models $g_\Omega(.)$. The regression head, $h_\Psi(.)$, is adapted to the number of genes, p, to be predicted for the corresponding dataset.

Model weights Θ are optimized using paired image patches $\boldsymbol{X}$ and their corresponding gene expression values $\boldsymbol{y}$ using the Mean Squared Error (MSE),

$$\mathscr{L}(\Theta) = \frac{1}{|B|} \sum_{\boldsymbol{X},\boldsymbol{y} \in B} \big(f_\Theta(\boldsymbol{X}) - \boldsymbol{y}\big)^2$$

where B denotes the batch of samples. We use the same hyperparameters for all the experiments (as described in Section 3.1). The backbone and regression head are trained jointly in every setting.

2.3 Compound Scaling in EfficientNet-B0 (EffNet)

We used the `EffNet` [15] architecture for our experiments with scaled down models. `EffNet` backbone gave similar performance to the remaining models with a smaller set of parameters ($\approx 27\%$ fewer parameters than the next smallest model), making it a strong compact baseline. More importantly, `EffNet` provides a principled compound scaling framework [15] that allows controlled adjustment of network depth, width, and input resolution using a single scaling coefficient ϕ. The scaling factors are defined as:

$$d(\phi) = \alpha^\phi, \quad w(\phi) = \beta^\phi, \quad r(\phi) = \gamma^\phi,$$

where $d(\cdot)$, $w(\cdot)$, and $r(\cdot)$ scale the number of layers, channels, and the input resolution, respectively. The coefficient ϕ controls the overall size of the network. For `EffNet`, $\phi = 0$, and positive values of ϕ increase model capacity, while negative values of ϕ prune the number of layers and channels of the networks. The values of $\alpha > 1$, $\beta > 1$, and $\gamma > 1$ are selected such that $\alpha \cdot \beta^2 \cdot \gamma^2 \approx 2$ which keeps the increase in computational cost balanced when scaling the model. We construct models with approximately 3%, 5%, 10%, 20%, and 100% of the `EffNet` parameters by selecting negative values of ϕ to study the effect of model size on performance. Figure 1 shows the overall architecture of our model.

3 Results

3.1 Experimental Details

All models are implemented in Python 3.10 using PyTorch [11]. CPU inference experiments are performed on a system with 4 physical cores (8 logical threads)

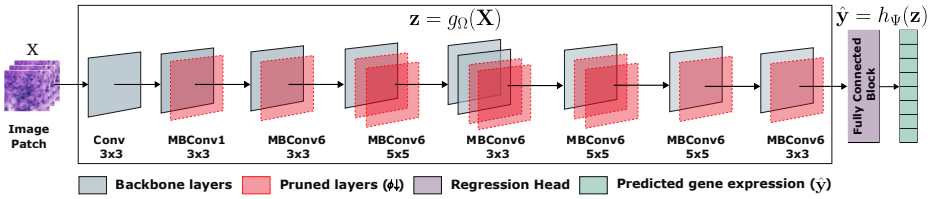

Fig. 1. Architecture of the TinyH2GNet, `EffNet`-based regression model for gene expression prediction. The network follows the tiny architecture of `EffNet`, where red dashed blocks indicate reduced capacity layers obtained via compound scaling ($\phi \downarrow$).

and 32 GB RAM. Model training was performed on NVIDIA A100 GPU (40 GB). To avoid data leakage, we use patient-level cross-validation, where samples from the same patient are kept in the same split. All models are trained and evaluated on identical data splits to ensure a fair comparison. We report L1 error, Pearson Correlation Coefficient (PCC) and other performance measures to evaluate prediction performance. We also report the number of model parameters and inference time to analyze the trade-off between accuracy and model size. All models are optimized using *Adam* [8] (Learning Rate (LR) 1×10^{-4}, batch size 128) for up to 200 epochs with a fixed seed of 123. Early stopping is applied based on validation Spearman Rank Correlation Coefficient (SRCC) (patience = 30). We use an adaptive learning rate scheduler (*ReduceLROnPlateau*) with decay factor 0.1 (patience = 30, minimum LR 1×10^{-8}). MSE is used as the training objective (described in the Section 2.2).

3.2 Datasets

We used three publicly available ST datasets that cover different tissue types and measurement platforms. All datasets provided paired histology images and gene expression profiles.

Human breast cancer in situ capturing transcriptomics (Hist2ST) [14] contains 68 ST samples from 23 breast cancer patients. Standard pre-processing was done by extraction of 224×224 pixel histological patches centered around each sequencing spot. These patches are used as input to the Convolutional Neural Network (CNN)-based models. We trained the models to predict 845 Highly Variable Genes (HVGs) as used in the baselines.

HER2-positive Breast Tumors (HER2) [1] consists of 36 formalin-fixed paraffin-embedded (FFPE) tissue sections from 8 patients profiled using the ST platform. We used the same patch extraction pipeline as the Hist2ST dataset. The models predicted 785 HVGs for this dataset as used in the baselines.

10x Xenium Human Breast Cancer (10x Xenium) [6] provides single-cell spatial gene expression profiles from a human breast cancer sample. The dataset consists of 167,780 cells with 541 genes and a matched H&E image. To establish image-gene correspondence at the cell level, we extracted 128×128 pixel image patches centered at each cell location.

3.3 Comparison of Existing Models

We compare several CNN and transformer-based models for spatial gene expression prediction. All models are trained and evaluated using the same data splits and preprocessing. Quantitative results are reported in Table 2. Across datasets, no single architecture consistently achieves the best results across all metrics and datasets. Both CNN-based models (e.g., VGG, `EffNet`) and transformer-based models (e.g., TCGN, HisToGene, `THItoGene`) show similar performance trends in gene expression prediction. An important observation from Table 2 is that the comparatively small `EffNet` achieves performance close to larger models. In contrast, models such as VGG and transformer-based approaches rely on much larger backbones, but the gains in accuracy are limited. Based on this analysis, we use `EffNet` as the backbone for further experiments, as it provides a strong and compact baseline for studying the effect of model size on prediction performance. This motivated us to focus on parameter-efficient variants of `EffNet` in the following sections.

Table 2. Performance comparison of different models on ST datasets. The parameters are specified in millions. The best results are in bold, second best are underlined. From a Paired t-test we did not find a statistical difference between the best and the second best method across datasets. `THItoGene` is evaluated on Hist2ST and HER2 only, as direct application to 10x Xenium is not feasible due to high memory cost. Abbreviations used: PCC: Pearson Correlation Coefficient, SRCC: Spearman Rank Correlation Coefficient, L1: Mean Absolute Error (MAE), L2: Mean Squared Error (MSE), MAPE: Mean Absolute Percentage Error.

Dataset	Model	≈ Params	PCC↑	SRCC ↑	L1↓	L2↓	MAPE(%)↓
Hist2ST [14]	`STNet`	7.8	$0.542_{0.022}$	$0.466_{0.020}$	$1.969_{0.060}$	$8.402_{0.430}$	$74.747_{1.295}$
	`DeepSpaCE`	120.9	$\underline{0.552}_{0.017}$	$\underline{0.478}_{0.014}$	$\mathbf{1.937_{0.041}}$	$\underline{8.334}_{0.424}$	$75.049_{0.779}$
	`HisToGene`	68.9	$0.541_{0.012}$	$0.456_{0.027}$	$\underline{1.951}_{0.058}$	$8.413_{0.396}$	$75.234_{1.923}$
	`EffNet`	6.2	$0.542_{0.016}$	$0.462_{0.015}$	$1.970_{0.067}$	$8.450_{0.397}$	$\underline{74.559}_{0.847}$
	`TCGN`	22.9	$0.545_{0.014}$	$0.471_{0.013}$	$1.952_{0.064}$	$8.427_{0.362}$	$\mathbf{74.486_{0.579}}$
	`THItoGene`	63.6	$0.475_{0.160}$	$0.397_{0.151}$	$1.980_{0.056}$	$9.143_{2.197}$	$79.846_{6.363}$
	`TinyH2GNet`	0.3	$\mathbf{0.559_{0.021}}$	$\mathbf{0.483_{0.017}}$	$1.987_{0.053}$	$\mathbf{8.099_{0.369}}$	$74.712_{0.347}$
HER2 [1]	`STNet`	7.8	$0.461_{0.045}$	$0.405_{0.052}$	$2.297_{0.250}$	$9.555_{0.753}$	$\mathbf{68.089_{6.486}}$
	`DeepSpaCE`	120.7	$\mathbf{0.481_{0.039}}$	$\mathbf{0.423_{0.046}}$	$2.319_{0.173}$	$\mathbf{9.440_{0.970}}$	$69.983_{4.517}$
	`HisToGene`	69.0	$0.450_{0.035}$	$0.377_{0.027}$	$2.335_{0.146}$	$9.933_{1.093}$	$73.133_{2.478}$
	`EffNet`	6.1	$0.463_{0.045}$	$0.394_{0.051}$	$\mathbf{2.277_{0.236}}$	$9.745_{0.931}$	$73.612_{5.239}$
	`TCGN`	25.1	$\underline{0.473}_{0.039}$	$\underline{0.414}_{0.047}$	$\underline{2.284}_{0.211}$	$\underline{9.549}_{0.868}$	$\underline{69.833}_{4.852}$
	`THItoGene`	63.6	$0.373_{0.148}$	$0.331_{0.150}$	$2.383_{0.337}$	$10.849_{1.608}$	$76.623_{8.288}$
	`TinyH2GNet`	1.2	$0.451_{0.043}$	$0.385_{0.058}$	$2.291_{0.292}$	$9.895_{0.873}$	$71.270_{5.795}$
10x Xenium [6]	`STNet`	7.5	$\underline{0.674}_{0.002}$	$\underline{0.534}_{0.002}$	$\mathbf{0.960_{0.003}}$	$\mathbf{4.520_{0.016}}$	$80.231_{0.291}$
	`DeepSpaCE`	119.7	$\mathbf{0.674_{0.001}}$	$\mathbf{0.552_{0.001}}$	$\underline{0.953}_{0.004}$	$\underline{4.544}_{0.009}$	$80.305_{0.556}$
	`HisToGene`	68.6	$0.661_{0.001}$	$0.503_{0.001}$	$1.002_{0.004}$	$4.698_{0.013}$	$80.445_{0.150}$
	`EffNet`	5.9	$0.670_{0.001}$	$0.527_{0.001}$	$0.972_{0.004}$	$4.585_{0.004}$	$\mathbf{80.097_{0.202}}$
	`TCGN`	23.9	$0.668_{0.001}$	$0.514_{0.004}$	$0.965_{0.002}$	$4.606_{0.027}$	$\underline{80.201}_{0.199}$
	`TinyH2GNet`	0.3	$0.661_{0.001}$	$0.515_{0.003}$	$1.014_{0.005}$	$4.692_{0.019}$	$81.634_{0.250}$

3.4 Effect of Model Size on Performance

We select the `EffNet` architecture and construct smaller variants by reducing the number of parameters using the compound scaling strategy (described in Section 2.3). We used negative values of the scaling coefficient ϕ to obtain models that contain approximately 3%, 5%, 10%, and 20% of the parameters of the full `EffNet` model. In some datasets, further reduction beyond the 3% configuration was not feasible due to architectural constraints imposed by the compound scaling rule.

Model Performance vs Pruned Model Sizes Figure 2 shows the L1 error boxplots and PCC across five folds as a function of model size. Across datasets, reducing the number of parameters leads to small changes in mean L1 error that is below 0.02 on HER2 and Hist2ST, and below 0.04 on 10x Xenium. Even at 3% model size, the absolute increase in L1 remains below 0.18 across datasets and models with 5% of the parameters achieve performance close to the full backbone. Similarly, the PCC changes only slightly as the model size is reduced across datasets.

Further, we tested the performance of our smaller variants by comparing them with full models (refer Table 2). We found that models with 5% and 20% parameters (Hist2ST, 10x Xenium and HER2) give a similar performance to the full models. In fact, the pruned models rank higher than some of the fully parameterized benchmarks.

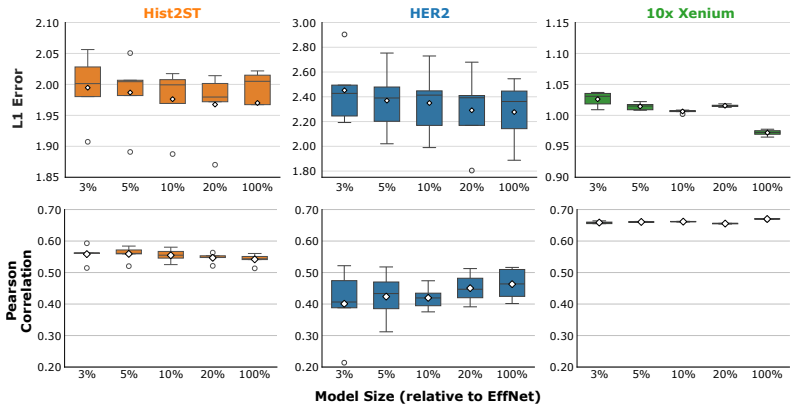

Fig. 2. Comparison of L1 error (top panel) and Pearson's Correlation (bottom panel) for different datasets across different `EffNet` model sizes.

Inference Time vs. Model Size. We report inference time for all model variants to study how model size affects runtime. Table 3 reports CPU inference time across datasets. Smaller models achieve lower latency in a consistent manner. Large backbones such as DeepSpaCE and HisToGene show much higher runtimes. On Hist2ST, the 3% model is around 4–5× faster than the full model.

Table 3. CPU inference time (**seconds**) for different models across datasets. `EffNet` tiny variants correspond to compound scaling ratios: 3%, 5%, 10%, and 20%. For **HER2** and **Hist2ST**, inference is extrapolated for 700 and for **10x Xenium** 167780 spots (covering the entire WSI). Results on 10x Xenium dataset were scaled by a factor of 10^{-3}. Our CPU was unable to perform inference for `TCGN` and `THItoGene` models on 10x Xenium dataset.

Model	Hist2ST	HER2	10x Xenium $* 10^{-3}$
`EffNet` (3%)	$\mathbf{4.117_{0.372}}$	$\mathbf{4.411_{0.527}}$	$\mathbf{0.994_{0.037}}$
`EffNet` (5%)	$\underline{5.369_{0.483}}$	$\underline{4.826_{0.495}}$	$\underline{1.140_{0.040}}$
`EffNet` (10%)	$6.204_{0.605}$	$6.766_{0.561}$	$1.369_{0.087}$
`EffNet` (20%)	$7.740_{0.685}$	$8.009_{0.837}$	$1.718_{0.054}$
`EffNet` (100%)	$20.418_{1.176}$	$19.494_{1.071}$	$3.684_{0.053}$
`STNet` (127%)	$48.530_{0.764}$	$47.380_{1.121}$	$8.517_{0.211}$
`DeepSpaCE` (1971%)	$167.270_{66.218}$	$117.975_{0.993}$	$102.264_{4.002}$
`HisToGene` (1125%)	$113.838_{4.866}$	$110.853_{2.926}$	$381.516_{48.942}$
`THItoGene` (1038%)	$62.707_{11.397}$	$28.071_{12.694}$	–
`TCGN (411%)`	-	-	-

On 10x Xenium, the speedup is around 3–4× when using the smallest variant. This shows that parameter reduction leads to large runtime gains while preserving prediction quality. These results show that tiny models offer a better trade-off between accuracy and runtime.

4 Discussion and Conclusion

Our systematic benchmarking shows that CNN, transformer, and hybrid architectures achieve similar performance on the evaluated Hist2ST, HER2, and 10x Xenium breast cancer datasets under the same training setup. While some models perform better on specific metrics, no architecture consistently outperforms the others across all datasets. These results suggest that increasing model size and architectural complexity does not always lead to proportional gains in prediction performance for the evaluated image-to-ST tasks.

Based on this observation, we focus on principled parameter scaling instead of designing another large backbone. The `EffNet` variants used in TinyH2GNet preserve most of the predictive performance even when reduced to a small fraction of the original model size. The L1 error and Pearson correlation remain close to the full model for several reduced variants, especially in the 5–20% parameter range. This suggests that compact backbones can learn useful histology-gene mappings in the evaluated datasets.

The CPU inference time results further show that reducing model size leads to clear runtime gains on standard hardware. This makes `TinyH2GNet` a practical option for resource-constrained computational settings where GPU access may be limited. At the same time, some genes and datasets remain hard to predict.

This suggests that image features alone may not fully explain all spatial gene expression patterns.

Overall, this work shows that parameter-efficient models can achieve performance close to larger backbones for image-based gene expression prediction in the evaluated breast cancer datasets. Future work should study pathway-level and gene-level behavior under model compression, and should evaluate lightweight models across broader tissues, institutions, and spatial transcriptomics platforms.

Disclosure of Interests

The authors have no competing interests to disclose.

References

1. Andersson, A., Larsson, L., Stenbeck, L., Salmén, F., Ehinger, A., Wu, S.Z., Al-Eryani, G., Roden, D., Swarbrick, A., Borg, Å., et al.: Spatial deconvolution of her2-positive breast cancer delineates tumor-associated cell type interactions. Nature communications **12**(1), 6012 (2021). https://doi.org/https://doi.org/10.1038/s41467-021-26271-2
2. Coutant, A., Cockenpot, V., Muller, L., Degletagne, C., Pommier, R., Tonon, L., Ardin, M., Michallet, M.C., Caux, C., Laurent, M., Morel, A.P., Saintigny, P., Puisieux, A., Ouzounova, M., Martinez, P.: Spatial transcriptomics reveal pitfalls and opportunities for the detection of rare high-plasticity breast cancer subtypes. Laboratory Investigation **103**(12), 100258 (2023). https://doi.org/https://doi.org/10.1016/j.labinv.2023.100258, https://www.sciencedirect.com/science/article/pii/S0023683723002015
3. Fischer, A.H., Jacobson, K.A., Rose, J., Zeller, R.: Hematoxylin and eosin staining of tissue and cell sections. Cold spring harbor protocols **2008**(5), pdb–prot4986 (2008). https://doi.org/10.1101/pdb.prot073411
4. Fu, X., Cao, Y., Bian, B., Wang, C., Graham, D., Pathmanathan, N., Patrick, E., Kim, J., Yang, J.Y.H.: Spatial gene expression at single-cell resolution from histology using deep learning with ghist. Nature Methods pp. 1–11 (2025). https://doi.org/https://doi.org/10.1038/s41592-025-02795-z
5. He, B., Bergenstrahle, L., Stenbeck, L., Abid, A., Andersson, A., Borg, A., Maaskola, J., Lundeberg, J., Zou, J.: Integrating spatial gene expression and breast tumour morphology via deep learning. Nature Biomedical Engineering **4**(8), 827–834 (2020). https://doi.org/10.1038/s41551-020-0578-x, https://doi.org/10.1038/s41551-020-0578-x
6. Janesick, A., Shelansky, R., Gottscho, A.D., Wagner, F., Williams, S.R., Rouault, M., Beliakoff, G., Morrison, C.A., Oliveira, M.F., Sicherman, J.T., et al.: High resolution mapping of the tumor microenvironment using integrated single-cell, spatial and in situ analysis. Nature communications **14**(1), 8353 (2023). https://doi.org/https://doi.org/10.1038/s41467-023-43458-x
7. Jia, Y., Liu, J., Chen, L., Zhao, T., Wang, Y.: Thitogene: a deep learning method for predicting spatial transcriptomics from histological images. Briefings in Bioinformatics **25**(1), bbad464 (2024). https://doi.org/https://doi.org/10.1093/bib/bbad464

8. Kingma, D., Ba, J.: Adam: A method for stochastic optimization. International Conference on Learning Representations (2014)
9. Monjo, T., Koido, M., Nagasawa, S., Suzuki, Y., Kamatani, Y.: Efficient prediction of a spatial transcriptomics profile better characterizes breast cancer tissue sections without costly experimentation. Scientific Reports **12**(1) (2022). https://doi.org/10.1038/s41598-022-07685-4, https://doi.org/10.1038/s41598-022-07685-4
10. Pang, M., Su, K., Li, M.: Leveraging information in spatial transcriptomics to predict super-resolution gene expression from histology images in tumors. bioRxiv (2021). https://doi.org/10.1101/2021.11.28.470212, https://www.biorxiv.org/content/early/2021/11/28/2021.11.28.470212
11. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Köpf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library (12 2019). https://doi.org/10.48550/arXiv.1912.01703
12. Rahaman, M.M., Millar, E.K., Meijering, E.: Breast cancer histopathology image-based gene expression prediction using spatial transcriptomics data and deep learning. Scientific Reports **13**(1), 13604 (2023). https://doi.org/https://doi.org/10.1038/s41598-023-40219-0
13. Rao, A., Barkley, D., França, G.S., Yanai, I.: Exploring tissue architecture using spatial transcriptomics. Nature **596**(7871), 211–220 (2021). https://doi.org/https://doi.org/10.1038/s41586-021-03634-9
14. Stenbeck, L., Bergenstråhle, L., Lundeberg, J., Borg, Å.: Human breast cancer in situ capturing transcriptomics. Mendeley Data **2** (2021)
15. Tan, M., Le, Q.: Efficientnet: Rethinking model scaling for convolutional neural networks. In: International conference on machine learning. pp. 6105–6114. PMLR (2019)
16. Xiao, X., Kong, Y., Li, R., Wang, Z., Lu, H.: Transformer with convolution and graph-node co-embedding: An accurate and interpretable vision backbone for predicting gene expressions from local histopathological image. Medical Image Analysis **91**, 103040 (2023). https://doi.org/https://doi.org/10.1016/j.media.2023.103040