Spot On: Unraveling Patterns in Colorectal Cancer Tumors with Spatial Transcriptomics and Image-based Deep Learning

Viraj Mehta
Stanford University
viraj28m@stanford.edu

Avash Shrestha
Stanford University
avash@stanford.edu

Priyanka Shrestha
Stanford University
shrestp@stanford.edu

June 6, 2024

Abstract

Colorectal cancer is the second deadliest cancer in the U.S [6]. Effective treatment of cancers such as colorectal cancer (CRC) is limited by the intrinsic heterogeneity of tumors, particularly between patients. Understanding the heterogenous distribution of cell types within tumors is an important clinical problem in developing personalized treatment options. Towards this goal, spatially-resolved transcriptomics (SRT) has emerged as an important method for visualizing the spatial distribution of RNA transcripts within tissues and cells, thus allowing for cell-type identification.

Our project aims to improve upon the existing methodology of the SpaGCN model, which integrates spatial transcriptomics and gene expression data along with histology images to predict clusters within tissue samples. In addition to a baseline model that harnesses SpaGCN for CRC tumor tissue clustering, we implement several downstream extensions that improve upon SpaGCN: (1) modifying the loss function by harnessing a JS Divergence rather than KL Divergence, (2) implementing a distance decay function for the weight of pixel values in histology images, and (3) integrating pathologist annotations into a semi-supervised learning model.

We find that our JS Divergence models outperforms or perform equivalently to SpaGCN on 7/14 samples, and our distance decay function outperforms or perform equivalently on 4/14 samples, demonstrating improvement over the baseline and potential for expansion to additional architectural modifications.

1. Introduction

Colorectal cancer tumors can exhibit significant genetic, epigenetic, and morphological heterogeneity, making personalized treatment difficult and leading to drug resistance. Further, it is extremely time- and cost- intensive to identify heterogenous components manually via methods such as pathologist annotation of cell types. Thus, automatic identification of differential expression of genes or pathways in tumor microenvironments (TME) is an important problem that remains to be solved in cancer treatments.

One emergent method that can be applied towards solving this problem is spatially-resolved transcriptomics (SRT), which enables the localization of RNA transcripts within spatial domains such as tissue samples [4]. Through methods such as next-generation sequencing (NGS), one is able to precisely identify coordinates for transcriptomic data obtained from existing methods such as single-cell RNA sequencing (scRNA-seq). Knowing the exact location of differential gene expression is important towards targeting specific genes or pathways in treatment development.

In approaching this problem, Hu et al. developed SpaGCN, a multimodal graph convolutional network that integrates three data components: (1) gene expression data, (2) spatial location of gene expression, and (3) histology images (microscopic images of disease tissues) [3]. We further describe the architecture for SpaGCN below. SpaGCN harnesses a clustering algorithm to identify clusters of similar spots within the histology images, each of which corresponds with certain differential gene expression. In order to identify these genes, SpaGCN also enables identification of spatially-variable genes (SVGs) expressed within each cluster.

We harnessed SpaGCN as a baseline to identify spatial clusters for 14 colorectal cancer tumor samples. SpaGCN trains on a single sample at a time. The inputs to our model were spatial gene expression data in the form of a matrix indicating gene expression at every location in the histology image, along with the histology image itself. Applying SpaGCN resulted in several spatial clusters, where each location or spot in the histology image is assigned to a domain. Downstream, these domains assignments are used to identify spatially-variable genes associated with each cluster, essentially indicating which genes of interest are most expressed in each cluster.

Following our baseline, we implemented several improvements to the SpaGCN model, which are described in close detail in Methods. For our first extension, we aimed to modify SpaGCN's vision component, which calculates similarity of locations by comparing pixel values. We im-

plemented a distance decay function that weights pixel values based on how far away they are from the current location. At a high level, our second extension involved modifying the loss function to harness Jensen-Shannon Divergence rather than Kullback-Leibler Diveregence. Lastly, we also trained a semi-supervised learning model to incorporate pathologist annotations of each location as a "ground truth."

2. Related Works

Previously, Wang et al. showed the importance of charting TME for colorectal cancer using SRT, given the degree of heterogeneity uncovered [7]. Here, we discuss additional previous works in related domains.

Hunter et al. harness SRT to identify an "interface" where melanoma tumor cells contact neighboring tissues; they succesfully demonstrate enrichment in cilia genes and proteins at the tumor boundary, which could be targets for melanoma treatment development [4]. For CRC specifically, Wood et al. evaluated SRT data, finding enrichment of immune cell populations (regulatory T cells and neutrophils) in patients with poor prognoses, demonstrating the significant clinical relevance of this type of analysis [8]. One exemplary paper in this category is that of Valdeolivas et. al, who harness SRT data for CRC to cluster cell populations within tumor slices, making use of SRT data as well as histology image slices [6]. Valdeolivas et al., unlike their peers, perform cell-type clustering based on gene expression, comparing these clusters with pathologist annotations of tumor components, thus inspiring our use of pathologist annotations as described in Methods below [6].

However, one limitation of the previous papers is that they do not harness deep learning or computer vision methods to analyze the images themselves in conjunction with SRT data. As discussed by Toninelli et al., it is important to harness both SRT data and image data, as image-based algorithms perform clustering based on segmentation of cell or nuclear boundaries, but physical features alone are not enough to identify complex cell-to-cell interactions [5]. Conversely, computer vision/deep learning methods have the potential to identify patterns in gene expression or physical features that are important towards understanding differential gene expression patterns at various locations.

In this vein, various architectures have begun to emerge, such as SpaGCN from Hu et al. as mentioned above [3]. Similarly, Xu et al. develop SPACEL, a deep learning-based architecture for spatial domain identification harnessing a similar graph convolutional network approach [9]. Though this is a similar SOTA approach, SPACEL's sequential modularity can be a limitation, with distinct modules for handling cell type deconvolution, spatial domain identification, and 3D alignment. Additionally, SPACEL focuses on a different task, constructing 3D tissue architectures, while SpaGCN is tailored towards identifying spatial domains within individual histology images.

Zhao et al. also introduce TransformerST, a model harnessing a transformer architecture to discern latent image-gene representations, using multi-head attention to build a spatial graph from the feature set [1]. One strength of TransformerST is that it achieves super-resolved gene expression without scRNA-seq data; however, it also has a much larger computational overhead than SpaGCN, and is better suited for tasks with larger/more complex datasets.

Uniquely, Fatemi et al. harness Inceptionv3 to predict gene expression from local histology image patches; however, this presents a separate prediction task that could be an interesting exploration if we did not have access to gene expression data already [2]. Finally, Xu et. al also introduce SEDR, harnessing a deep autoencoder along with self-supervised learning for the same task as Zhao et al., Xu et al., and Hu et al.; though they achieve good latent representations, these are solely based off of SRT data, and do not harness histology images in a computer vision sense [10].

3. Methods

For our baseline, we use SpaGCN to cluster spots into spatial clusters and generate spatially-variable genes within each domain. Relevant to our ablation tests, this baseline model uses just the Kullback-Leibler Divergence as the loss function with an averaging window to capture the 3rd dimension of the histological (image) information.

3.1. Histology Image Averaging Modifications

In the original SpaGCN paper, the authors use a Graph Convolutional Network (GCN) which naturally entails the construction of an undirected weighted graph $G = (V, E)$ (where each vertex $v \in V$ is a spot) for their spot data, which comes in the form of gene expressions and a histology image [3]. This graph is fully connected, and the weights on the edges between any two vertices in V is calculated based on the physical locations as well as the histological information of the two vertices. The weights are meant to measure "the degree of relatedness between spots u and v and is negatively associated with their distance" [3].

The physical locations are used to determine how far the spots are from each other, while the histological information is used to capture the differences in the average RGB values around a spot, i.e. how much one spot's surrounding image window differs visually from another's. This averaging involves constructing a square window (with the center pixel being the spot they are creating the representation for) and calculating the mean of the RGB values. This is analogous to average pooling, but without a sliding window. Instead, the window is centered on each individual spot. Importantly, this means that an pixel at the top left is weighted the same as a pixel right next to the center pixel.

Thus, we concerned ourselves making modifications to the histological information when calculating the edge weights of G, because this measurement is not punishing the distance pixels are from the central spot. A key assumption of spatial transcriptomics analysis is that spots or cells closer to each other should have a greater affect on neighboring biological processes. This phenomena is not captured in the current integration of histological information, thus we propose an alternative averaging of this window. Specifically, we propose a square root decay for this window to penalize the influence of spots that are further away.

We wanted the window to scale to the quality of the image, so that the window would better capture the physical distance between the spots, since, for example, the distance between pixels for an image with an area of 100 pixels and an image with 10000 pixels, where both images capture the exact same real-world area, differs significantly. So, we defined a "spotlight of interest" such that the area of this circular spotlight is approximately 0.3% the area of the entire histology image. Any pixels outside of this "spotlight" would not contribute to the pixel at the center, i.e. the weight of these outer pixels would be 0. This was chosen so that the area of the inscribed square (our alternative window) would have approximately 0.2% the area of the entire histology image. The math for this can be seen in the Appendix. Within this spotlight, the further away from the center of the circle a point is, the less weight is assigned to its RGB values when determining the average, where the minimum and maximum weights are 0 and 1. This allows for distance within the image window itself to be captured and downweighted appropriately. This also allows for the inscribed square to have a sidelength of $n = 85$, whereas the suggested window size in the SpaGCN paper had a sidelength of $n = 50$ [3]. We included this sidelength as a hyperparamter for our model, beta, but having a sidelength larger than the slidelength of the inscribed square is not advised, since the outer pixels would then have a weight of 0.

The formula for the weight of pixel p_i using decaying average is as follows:

$$D_i = \frac{r^{\frac{1}{2}} - ||p_s, p_i||_2^{\frac{1}{2}}}{r^{\frac{1}{2}}} \tag{1}$$

where r is the radius of the "spotlight of interest," p_s is the center pixel (the spot), and the distance metric is L_2 distance. To get the alternative decaying average of a window, we sum up all of the weighted pixels, $\sum_i^{n^2} D_i p_i$ and divide by the the sum of the weights (to normalize), $\sum_i^{n^2} D_i$, where n^2 is the total number of pixels in the window.

3.2. Loss Function Modifications

In the original SpaGCN paper, the authors used Kullback-Leibler (KL) Divergence as their loss function. This is because SpaGCN conducts an unsupervised clustering task, so they iteratively constructed 2 distributions for each spot and passed them into KL Divergence [3]. The predicted distribution q is constructed as follows:

$$q_{ij} = \frac{(1 + h_i - \mu_j^2)^{-1}}{\sum_{j'}^{K}(1 + h_i - \mu_j^2)^{-1}} \tag{2}$$

where q_{ij} is the predicted probability of assigning spot i to cluster j, h_i is the embedding of spot i and μ_j is centroid μ_j [3]. The target distribution p is constructed as follows:

$$p_{ij} = \frac{\frac{q_{ij}^2}{\sum_{i'}^{N} q_{i'j}}}{\sum_{j'}^{K} \frac{q_{ij}^2}{\sum_{i'}^{N} q_{i'j}}} \tag{3}$$

Note that p is constructed from q [3].

For background, KL Divergence of two probability distributions p_i and q_i for a single spot i is as follows:

$$D_{KL}(p_i, q_i) = \sum_{j \in K} p_{ij} \log(\frac{p_{ij}}{q_{ij}}) \tag{4}$$

where K is all possible clusters and p_{ij}/q_{ij} is the probability of spot i being in cluster j for p_i/q_i respectively. KL Divergence of p and q measures how different q is from p, where the minimum value is 0 (if $p = q$) and the maximum is ∞ (for example, consider an arbitrary $q_{ij} \approx 0$) [11]. The specific loss function they use is:

$$L_{KL} = D_{KL}(P, Q) = \sum_{i}^{N} \sum_{j}^{K} p_{ij} \log(\frac{p_{ij}}{q_{ij}}) \tag{5}$$

where N is the # of spots, both P and Q are the probability distributions p and q, respectively, for all N spots, and everything else is as defined previously. The dimensions of both P and Q are NxK. Importantly, the authors frame q_i as the predicted distribution and p_i as the target or "ground-truth" distribution for a spot i, and thus want to improve q_i (the predictions) by shrinking its differences from p_i.

Jensen-Shannon Divergence Our modification involved changing this loss to Jensen-Shannon (JS) Divergence. For background, JS Divergence of two probability distributions p_i and q_i for a single spot i is as follows:

$$D_{JS}(p_i, q_i) = \frac{1}{2}(D_{KL}(p_i, m_i) + D_{KL}(q_i, m_i)) \tag{6}$$

$$m_i = \frac{1}{2}(p_i + q_i) \tag{7}$$

JS Divergence measures the similarity between p_i and q_i using the mixture distribution of the two probability distributions, m_i, by averaging the KL dDvergence of the two distributions with m_i. For JS Divergence, the minimum value is 0 (if $p = q$) and the maximum is 1 [11]. In this way, JS Divergence instead frames both p_i and q_i as both predictions and 'ground-truth' targets.

We decided to modify the architecture in this way, because, as noted above, the construction of p_i doesn't come any ground-truth information (since this is unsupervised learning). Thus, we thought treating both p_i and q_i as target distributions, in a sense, could inform a better clustering

than assuming p_i to be the true distributions of the spots into various clusters. We define the JS Divergence loss function to be:

$$L_{JS} = D_{JS}(P, Q) = \sum_{i}^{N} D_{JS}(p_i, q_i) \quad (8)$$

where everything is as defined previously. In general, we define L_D to be the loss term responsible for the specific Divergence measurement. Thus, $L_D \in \{L_{KL}, L_{JS}\}$

3.3. Semi-Supervised Learning: Pathologist Annotation

The baseline model relies completely on unsupervised learning to cluster the SRT data into spatial clusters. However, our dataset also contains pathologist annotations labeling each spot with the tissue region at that location. These annotations show that certain cell types and regions cluster with other cell types and regions in a spatially dependent manner. In the real world, there are often some pathologist annotations of tumor slices available to reference when analyzing the heterogeneity of tumors. In order to capture this additional information in our model, we modified the SpaGCN architecture to have the option of using pathologist annotations during training in a semi-supervised manner.

Using the pathologist annotations as the ground truth labels for each spot, we added an additional cross entropy loss to L_D capturing the unsupervised learning. We assumed that within a given cluster, each spot's annotation can be drawn from the same probability distribution. Thus, given K spatial domains identified by SpaGCN, we create k-dimensional logit vectors for each spot by counting the abundance of each type of annotation within that spatial cluster. Our semi-supervised loss function follows as such,

$$\mathrm{L}_S = L_D + CE = L_D - \sum_{i=1}^{M} \sum_{j=1}^{K} \log(\frac{e^{s_l}}{\sum_z^K e^{s_z}}) \quad (9)$$

Where M is the number of labeled samples,s_z is the count of annotation type z, with l being the positive annotation class, and everything else is as defined previously.

4. Dataset and Features

At a high level, our dataset consisted of 14 samples of tumors from colorectal cancer patients (7 patients, 2 samples per patient). The average image resolution was 1973 x 1985. These samples were obtained from the paper by Valdeolivas et al., who made their data publicly available [6]. Each of the 14 samples consisted of an annotated data matrix mapping gene expression data to "spots," which are essentially spatial locations. These spots are defined by 10x Genomics VISIUM, a spatial transcriptomics technology outputting a mapping from genes to spatial locations. Each spot has an associated coordinate in the histology image, and gene expression for each gene is defined in the matrix at every spot (Figure 1). Along with this matrix, we

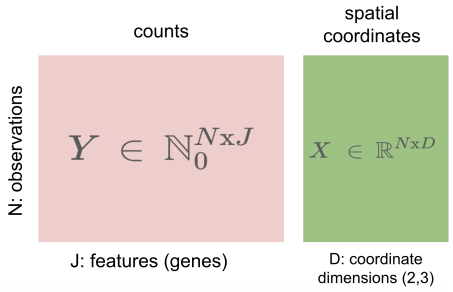
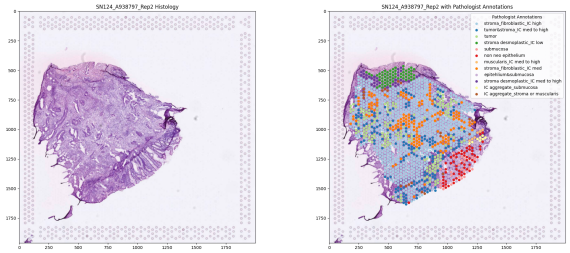


Figure 1: Spatial transcriptomics data. D can be in 2 or 3 dimensions

had access to all 14 original histology images of the tumors (Figure 2a). Lastly, we also had access to annotations made by a pathologist for every spot defining which tissue region each spot belongs to (Figure 2b).



(a) Tumor slice imaged on H&E stained slide (b) Tumor slice overlayed with pathologist annotations

Figure 2: Histology Input Data

We took several preprocessing steps for our data. First, we filter out cells that have fewer than 200 genes detected, as low gene counts are often indicative of dead of poor-quality cells that simply introduce noise. Next, we filtered out genes that are detected in fewer than 3 cells, as genes that are not widely expressed are likely not biologically relevant, and similarly only function to introduce noise into the detection of spatially variable genes. We then proceeded to normalize the gene expression matrix, where gene expression is indicated by counts of mRNA transcripts for a gene. We normalized the counts to sum to 10,000 to account for variations in sequencing depths between cells, and finally applied a log normalization to make the data normally distributed and account for variance.

5. Experiments/Results/Discussion

First, we performed hyperparameter selection using grid search for the hyperparameters defined in Table 1 below. 's' is a hyperparameter defined by SpaGCN controlling the relative importance of histology information. 'beta' defines the side length of an image patch used to calculate z, the 3rd

Hyperparameter Name	Grid Search Values
s	[0.5, 1.0, 1.5, 2]
beta	[25, 35, 45, 50, 55, 65, 75]
p	[0.5, 1]
init	['louvain', 'kmeans']
resolution	[0.2, 0.4, 0.6, 0.8, 1.0]
num neighbors	[5, 10, 15, 30, 50, 70, 90]
num clusters	[5, 9, 11, 13, 15]
learning rate	[1e-2, 1e-3, 1e-4]
weight decay	[0, 1e-1, 1e-3, 1e-5]

Table 1: Hyperparameters used in grid search

dimension from histology, as described in Methods above. 'p' is the percentage of total expression controlled by the neighboring domains. 'init' defines whether K-means or the Louvain method are used for clustering; if Louvain is used, 'resolution' defines how fine-grained the clusters are and 'num neighbors' defines whether a more global or local structure is captured. if K-means is used, 'num clusters' defines how many clusters should be created. Finally, learning rate and weight decay (for the optimizer) maintain their traditional definitions. This grid search was conducted on the baseline model, but these hyperparameters were used across all model types due to compute/time constraints.

For our evaluation metrics, we primarily used Adjusted Rand Index (ARI) to evaluate the clustering and Moran's I (M.I.) to evaluate the SVG generation. ARI measures agreement between predicted clusters and ground truth clusters; for our purposes, we defined ground truth using the pathologist annotations of tumor subregions. ARI ranges from a scale of 0 to 1. Moran's I is a metric for evaluating spatial autocorrelation, measuring how similar or dissimilar spots are to neighboring spots. Moran's I ranges from -1 (perfect dissimilarity) to 1 (perfect similarity). We used Moran's I to evaluate the extent of spatial autocorrelation of gene expression at spots, thus validating our selection of spatially-variable genes.

The Adjusted Rand Index (ARI) is calculated by first calculating the rand index (RI) and then applying an adjustment:

$$\text{RI} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{FP} + \text{FN} + \text{TN}} \tag{10}$$

$$\text{ARI} = \text{RI} - \frac{\mathbb{E}[\text{RI}]}{max(\text{RI})} - \mathbb{E}[\text{RI}] \tag{11}$$

where (let P = predicted clustering and GT = ground truth clustering):
TP: True Positives (# of pairs of elements in the same cluster in **both** P and GT).
TN: True Negatives (# of pairs of elements **not** in the same cluster in **both** P and GT).
FP: False Positives (# of pairs of elements in the same cluster in P but **not** in GT).
FN: False Negatives (# of pairs of elements **not** in the same cluster in P but in GT).

For Moran's I (M.I.):

$$\text{M.I.} = \frac{N \sum_{i,j}(x_i - \bar{x})(x_j - \bar{x})}{\sum_{i,j} d_{ij} \sum_i (x_i - \bar{x})^2} \tag{12}$$

where:
x_i and x_j: Gene expression values at spots i and j.
$\bar{x}$: Mean gene expression value across all spots.
d_{ij}: Distance between spots i and j
N: Total number of spots.

We performed ablation tests, training a model for every combination of model extension described in the Methods section. These results are displayed below, with Table 2 showing the ARI for generated clusters for each sample and Table 3 showing the average Moran's I for each gene identified as an SVG per sample. Visually, we can see the similarities between the clusters generated by the best model for each sample and the ground truth pathologist annotations in Figure 3. We can also see the average Moran's I for each cluster for every sample across model types in Figure 4.

	ARI by Model Type			
Sample	KL, avg.	KL, decay	JS, avg.	JS, decay
1	0.297	0.304	**0.306**	0.300
2	0.261	**0.302**	0.270	0.284
3	**0.101**	0.101	0.101	0.101
4	**0.140**	0.136	0.138	0.139
5	0.210	0.211	**0.236**	0.226
6	**0.413**	0.261	0.412	0.262
7	**0.217**	0.210	**0.217**	**0.217**
8	0.288	0.287	**0.303**	0.300
9	**0.342**	0.209	0.339	0.205
10	0.302	0.297	**0.303**	0.298
11	0.234	0.158	**0.238**	0.187
12	**0.180**	0.170	0.152	0.177
13	0.456	**0.462**	0.456	**0.462**
14	**0.278**	0.248	0.251	0.244

Table 2: Comparison of ARI (Adjusted Rand Index) for different model types. Best model for each sample bolded. KL = Kullback-Leibler Divergence, JS = Jensen-Shannon Divergence, avg. = average pixel weight function, decay = distance decay weight function. Refer to Appendix for sample names corresponding with sample numbers (Table 4).

Of note is that we exclude results from the extension of semi-supervised learning with pathologist annotations, as these ARIs and M.I.s were largely the same when including pathologist annotations as without. However, upon analysis, this can be attributed to the goal of the pathologist annotations' loss term. Unfortunately, from our debugging,

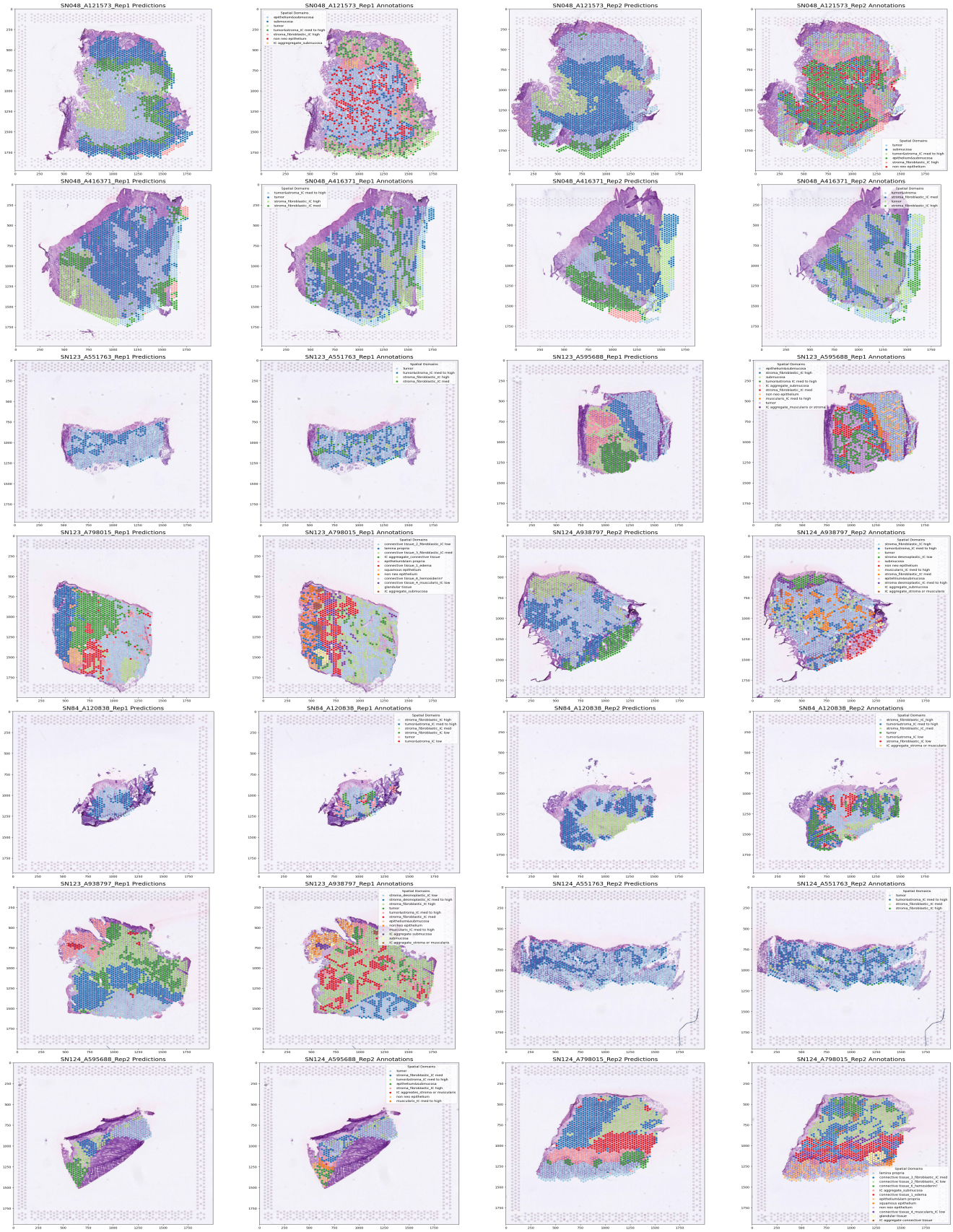
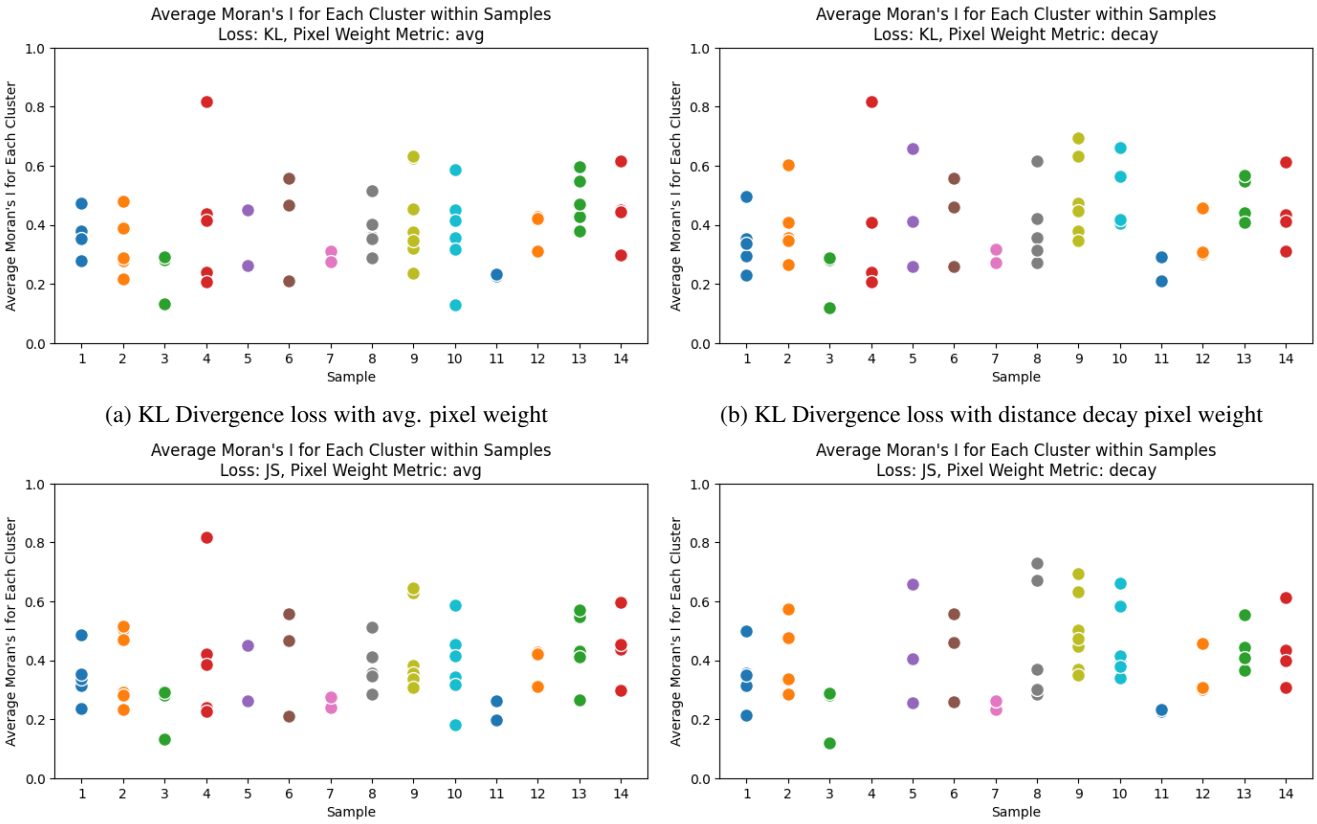


Figure 3: Best model clustering results compared to ground truth annotations for all 14 samples



(a) KL Divergence loss with avg. pixel weight

(b) KL Divergence loss with distance decay pixel weight

(c) JS Divergence loss with avg. pixel weight

(d) JS Divergence loss with distance decay pixel weight

Figure 4: Average Moran's I for each cluster within samples for all 4 model types

we realized that the membership of clusters did not vary much between iterations of training from the initially generated ones during the initialization step (either Louvain's method or K-means). This means that spots typically got assigned a cluster to begin with and stuck to them. Thus, it was difficult for the cross entropy loss for the pathologist annotations to impact the model, since that loss term would be pretty much the same value from the starting to ending iteration of the training step. Thus, this added loss term did not have a recognizable impact on our various models. However, the domain knowledge gained from being able to add an experts annotations into a clustering model is still an important biological insight that biological clustering models should have but that is currently lacking in the field. Our proposed method, though ineffective due to our choice of baseline model, could still be generalized and applied to other unsupersived biological methods where some domain expert annotations may exist.

As shown quantitatively, the JS Divergence loss outperforms or performs equivalently on 7/14 samples on the clustering task, and on 6/14 samples on the SVG generation task, indicating significant improvement over the baseline KL Divergence. We believe that this can largely be attributed to a more nuanced approach for the two probability distributions generated iteratively during training, since we no longer assume that exactly one of them is the target, or "true," distribution and the other is just the predicted distribution.

Further, the distance decay pixel weighting function demonstrated improvement on 4/14 samples on the clustering task and on 5/14 samples on the SVG generation task, indicating its ability to provide a better metric for incorporating image data into this multimodal model. Biologically, this is intuitive given spatial autocorrelation of genes, as one would expect cells that are spatially correlated to express similar genes and have similar function.

We visualize the generated clusters against the ground truth annotations for all samples in Figure 4. Qualitatively, can see that while the generated clusters are not as granular as the pathologist annotations, they are generally separated along the same boundaries for many samples. For example in SN048_A121573_Rep2, pathologists identify differential regions along the outside of the tumor slice versus in the inner portion. The outputted SpaGCN clusters follow a similar delineation with clear separation of spatial domains into ones that fall towards the edges of the tumor slice ver-

Sample	Moran's I by Model Type			
	KL, avg.	KL, decay	JS, avg.	JS, decay
1	**0.388**	0.385	0.377	0.376
2	0.309	0.339	0.321	**0.349**
3	**0.290**	0.286	**0.290**	0.286
4	**0.265**	0.259	0.255	——-
5	**0.280**	0.276	0.268	0.256
6	**0.370**	0.339	0.365	0.348
7	0.404	0.420	0.385	**0.422**
8	0.358	**0.365**	0.356	0.362
9	0.338	**0.353**	0.339	0.351
10	**0.480**	0.474	**0.480**	0.474
11	0.321	**0.424**	0.324	0.402
12	0.229	0.230	**0.245**	0.229
13	**0.339**	0.337	**0.339**	0.337
14	**0.455**	0.447	0.375	0.447

Table 3: Comparison of Moran's I for different samples. Best model for each sample bolded. Sample 4 for JS, decay excluded due to SpaGCN architecture error. Refer to Appendix for sample names corresponding with sample numbers (Table 4).

sus ones that fall towards the inner portion.

On the other hand, the SpaGCN generate clusters for SN048_A416371_Rep1 struggled to match the pathologists annotations, with an ARI of 0.101. This is likely becacuse this sample had very busy ground truth annotations without many clearly separated regions, a level of granularity that SpaGCN is not able to cluster to.

In examining the SVGs, we can see a high average Moran's I for sample SN123_A938797_Rep1 of 0.480 for KL loss with average pixel weighting. Upon qualitative examination of the clusters for this sample, we can see that there are distinct subregions, which biologically would correlate with distinct cell types and a high spatial autocorrelation. For one of the lowest Moran's I samples, SN048_A416371_Rep2 for JS loss with average pixel weighting, we can see that there is more intermingling of the generated clusters, which we would expect to have a lower spatial autocorrelation.

For further downstream biological analysis, we chose sample 9 (SN123_A798015_Rep1) that had both a high ARI of 0.342 and a high Moran's I of 0.356. A visual inspection shows that the generated clusters are able to differentiate between connective tissue annotated regions towards the right and epithelium regions towards the left. Based on the heatmap in Figure 5, we can see that cluster 1 strongly houses squamous epithelium cells while cluster 6, 7, 8 show high amounts of lamina propia (type of connective tissue) and mixtures of connective tissue and fibroblasts. From these findings we can see that connective tissue cells seem to have high spatial correlation. These types of cells are often associated with cancer metastasis and further characterization into these initial patterns could lead to a greater understanding of the mechanisms of cancer proliferation and how this can inform treatment.

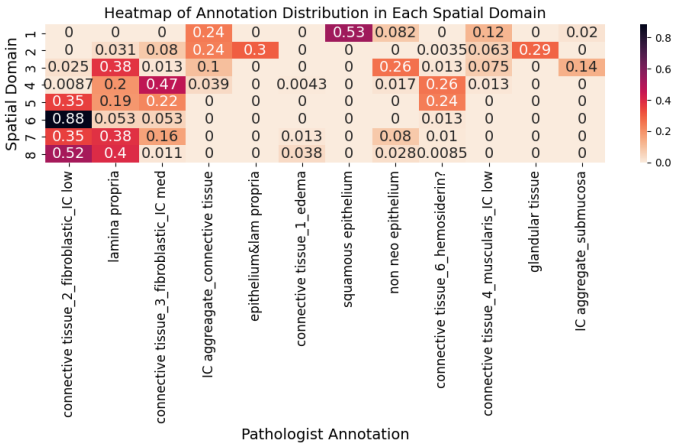


Figure 5: Heatmap of the proportion of different tumor regions in generated clusters for sample 9

6. Conclusion/Future Work

Identifying tumor subregions is an important problem in treating heterogenous cancers such as colorectal cancer. Further, state-of-the-art methods such as spatially resolved transcriptomics could benefit from integration with deep learning or computer vision methods. Utilizing SpaGCN as a baseline, we demonstrate success in making architectural modifications that show improvement in both cluster generation and detection of spatially-variable genes. Our modifications specifically focus on image data, involving incorporating distance decay pixel weighting and a JS Divergence loss function harnessing multimodal inputs. Downstream, this type of methodology could offer unique insights into genes or cell types to target with therapeutic development.

In the future, we would like to run an extensive hyperparameter search on all combinations of our model modifications to determine the best ones per model, rather than using those that worked best at baseline. We suspect that this search would allow for clearer trends in the varying performances of the different models, since it's quite likely that the best hyperparameters for the baseline model are not the best for other models. In addition, we would like to consider other ways of clustering the spots that could allow for more movement between the spots. This should allow us to better investigate the value of our extension of including ground-truth pathologist annotations to make the training semi-supervised instead of unsupervised, which we believe should help performance, as discussed above.

7. Appendix

7.1. Spotlight of Interest Work

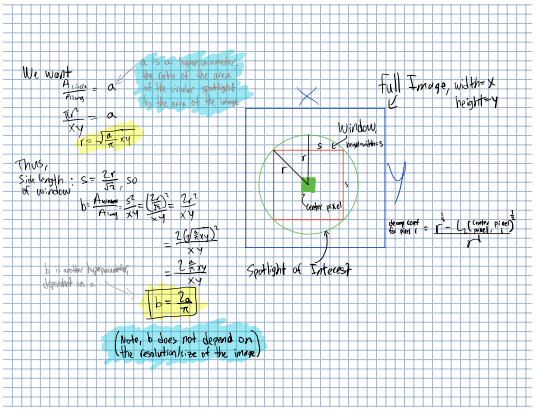


Figure 6: Spotlight of Interest diagram

7.2. Map of Sample Identifier to Sample Name

Sample Identifier	Sample Name
1	SN048_A121573_Rep1
2	SN048_A121573_Rep2
3	SN048_A416371_Rep1
4	SN048_A416371_Rep2
5	SN84_A120838_Rep1
6	SN84_A120838_Rep2
7	SN123_A551763_Rep1
8	SN123_A595688_Rep1
9	SN123_A798015_Rep1
10	SN123_A938797_Rep1
11	SN124_A551763_Rep2
12	SN124_A595688_Rep2
13	SN124_A798015_Rep2
14	SN124_A938797_Rep2

Table 4: Mapping from sample identifier to sample name.

8. Contributions and Acknowledgments

We harnessed the SpaGCN architecture created by Hu et al. for our baseline for this paper, and built modifications on top of their original code, linked here.

Individual team member contributions:

- Avash: Ideated and came up with histology image decay function and loss modification of JS Divergence for downstream extensions. Wrote code for these extensions, hyperparameter search, and ablation tests. Collated all results and contributed equally to writing paper.
- Priyanka: Ideated and came up with semi-supervised learning approach with pathologist annotations. Wrote code for this extension, post-processing of images/clusters, ARI calculation, and biological analyses of tumor subregions. Contributed equally to writing paper.
- Viraj: Wrote code for preprocessing data, baseline model, hyperparameter search, cluster generation, spatially-variable gene identification, and calculation of Moran's I metric. Conducted to ideating downstream extensions and background/related works research. Contributed equally to writing paper.

References

[1] Chongyue Zhao, Zhongli Xu, Xinjun Wang, Shiyue Tao, William A MacDonald, Kun He, Amanda C Poholek, Kong Chen, Heng Huang, Wei Chen, Innovative super-resolution in spatial transcriptomics: a transformer model exploiting histology images and spatial gene expression, Briefings in Bioinformatics, Volume 25, Issue 2, March 2024, bbae052, https://doi.org/10.1093/bib/bbae052.

[2] Fatemi et al., Inferring spatial transcriptomics markers from whole slide images to characterize metastasis-related spatial heterogeneity of colorectal tumors: A pilot study, Journal of Pathology Informatics, Volume 14, 2023, 100308, ISSN 2153-3539, https://doi.org/10.1016/j.jpi.2023.100308.

[3] Hu, J., Li, X., Coleman, K. et al. SpaGCN: Integrating gene expression, spatial location and histology to identify spatial domains and spatially variable genes by graph convolutional network. Nat Methods 18, 1342–1351 (2021). https://doi.org/10.1038/s41592-021-01255-8.

[4] Hunter, M.V., Moncada, R., Weiss, J.M. et al. Spatially resolved transcriptomics reveals the architecture of the tumor-microenvironment interface. Nat Commun 12, 6278 (2021). https://doi.org/10.1038/s41467-021-26614-z

[5] Mattia Toninelli, Grazisa Rossetti, Massimiliano Pagani, Charting the tumor microenvironment with spatial profiling technologies, Trends in Cancer, Volume 9, Issue 12, 2023, Pages 1085-1096, ISSN 2405-8033, https://doi.org/10.1016/j.trecan.2023.08.004.

[6] Valdeolivas, A., Amberg, B., Giroud, N. et al. Profiling the heterogeneity of colorectal cancer consensus molecular subtypes using spatial transcriptomics. npj Precis. Onc. 8, 10 (2024). https://doi.org/10.1038/s41698-023-00488-4

[7] Fei Wang et al. ,Single-cell and spatial transcriptome analysis reveals the cellular heterogeneity of liver metastatic colorectal cancer.Sci. Adv.9,eadf5464(2023).DOI:10.1126/sciadv.adf5464

[8] Wood, Mitchell, et al. "Spatially Resolved Transcriptomics Deconvolutes Prognostic Histological Subgroups in Patients with Colorectal Cancer and Synchronous Liver Metastases." Cancer Research, vol. 83, no. 8, 2023, pp. 1329-1342. American Association for Cancer Research, https://doi.org/10.1158/0008-5472.CAN-22-2794.

[9] Xu, H., Wang, S., Fang, M. et al. SPACEL: deep learning-based characterization of spatial transcriptome architectures. Nat Commun 14, 7603 (2023). https://doi.org/10.1038/s41467-023-43220-3.

[10] Xu, H., Fu, H., Long, Y. et al. Unsupervised spatially embedded deep representation of spatial transcriptomics. Genome Med 16, 12 (2024). https://doi.org/10.1186/s13073-024-01283-x.

[11] Lin, Jianhua. “Divergence measures based on the Shannon entropy.” IEEE Trans. Inf. Theory 37 (1991): 145-151.

Libraries used: NumPy, pandas, ScanPy, SquidPy, PyTorch, SciPy, OpenCV, scikit-learn, SpaGCN