FPT UNIVERSITY

Leveraging Deep Learning in Biological Image Analysis

by

Nghia Tra[1,2], Ha Nguyen[1,2], Nhat Nguyen[1,2]

Supervisor: Le Thanh Hai[2], Ph.D; Son Pham[1], Ph.D

(1 Bioturing Inc, San Diego, CA, US)
(2 FPT University, HCM, VN)

THE FPT UNIVERSITY HO CHI MINH CITY

Leveraging Deep Learning in Biological Image Analysis

by

Nghia Tra[1,2], Ha Nguyen[1,2], Nhat Nguyen[1,2]

Supervisor: Le Thanh Hai[2], Ph.D; Son Pham[1], Ph.D

(1 BioTuring Inc, San Diego, CA, US)
(2 FPT University, HCM, VN)

A final year capstone project submitted in partial fulfillment of the requirement for the Degree of Bachelor of Artificial Intelligence in Computer Science

DEPARTMENT OF AI

THE FPT UNIVERSITY HO CHI MINH CITY

April 2024

ACKNOWLEDGMENTS

We would like to express our sincere gratitude to the following individuals and organizations who contributed to the successful completion of this project:

1. Son Pham, Ph.D.:
    - For his invaluable guidance and continuous support throughout this project.
    - His expertise, insights, and constructive feedback have been instrumental in shaping the direction of our research and enhancing the quality of the outcomes.
2. Le Thanh Hai, Ph.D.:
    - Our research supervisor, whose guidance and expertise throughout this study.
    - His insightful feedback significantly improved the quality of my work.
3. BioTuring AI and Algorithm Team:
    - We are thankful to the BioTuring AI and Algorithm Team for providing the necessary resources and facilities required to conduct this research.
    - The conducive environment and access to relevant materials have greatly facilitated the progress of this project.
4. We would also like to extend our special thanks to Mr. Nguyen Quoc Trung and Mr. Le Phu Nguyen. Their insightful reviews and feedback throughout the project were instrumental in strengthening the overall quality of our work.

Lastly, we want to acknowledge the broader scientific community for fostering an environment of curiosity and collaboration. Thank you all for being part of this journey.

AUTHOR CONTRIBUTIONS

The contributions of the authors to this work are outlined as follows: Nhat Nguyen conducted research on Large Image Representation. Quality control of the dataset was performed by all authors. Nhat Nguyen, Nghia Tra, and Ha Nguyen were responsible for training and benchmarking the Cell Segmentation model. Spatial Analysis was researched collectively by all authors, with Nhat Nguyen focusing on Spatial Analysis in Disease Research and Graph Compass, Nghia Tra on constructing the Spatial Graph, and Ha Nguyen on Spatial Autocorrelation. Benchmarking was conducted by Nhat Nguyen, while the demonstration was completed by Ha Nguyen and Nghia Tra. All authors reviewed and approved the final version of the manuscript.

ABSTRACT

Biological image analysis is crucial for advancements in fields like drug discovery and biomedical research. Traditional methods, relying on manual feature extraction, are laborious and prone to errors. While deep learning has revolutionized automated image analysis, many challenges remain. These include the need for vast amounts of high-quality labeled data for training, limited generalizability to unseen data, and the computational complexity requiring expensive hardware. This project investigates the application of deep learning for biological image analysis, focusing on overcoming these limitations in three key areas: large image representation, accurate cell segmentation, and spatial image analysis. By addressing these challenges, we aim to develop robust and accessible deep learning tools that can significantly enhance biological image analysis, ultimately accelerating research progress.

Keywords: bioinformatics; spatial; deep learning; spatial analysis; cell segmentation

CONTENTS

List of Figures

List of Tables

1. INTRODUCTION

The field of biological research is undergoing a data revolution. High-throughput technologies, like next-generation sequencing and advanced microscopy, are churning out massive datasets (often referred to as "big data") at an unprecedented rate. These datasets hold immense potential to unlock new frontiers in our understanding of biological processes, from understanding how each cell works to characterize the interaction of a complex ecosystem. However, analyzing this vast amount of information presents significant computational challenges. A major challenge is the scarcity of labeled data. Unlike many other data analysis domains, where vast amounts of pre-labeled data may exist (think image recognition with millions of pre-labeled cat pictures), biological data often lacks this critical element. Identifying and meticulously categorizing the objects of interest within these complex datasets requires the expertise of trained scientists – a time-consuming and expensive process. This scarcity of labeled data severely limits the applicability of traditional supervised learning techniques, which rely heavily on pre-labeled data to train models for accurate analysis. Adding another layer of complexity is the inherent nature of biological data itself. Biological systems are intricate, with intricate spatial relationships between different molecules and diverse cell types existing within a single image. Capturing these nuances and accurately analyzing such data demands sophisticated computational approaches that can navigate the complexities of biological organization.

First, we provide an overview of **encoder-decoder architectures** (Figure 1.1). This powerful and versatile neural network architecture is designed for tasks involving sequence-to-sequence learning. Imagine an encoder as a powerful compressor, taking a complex input sequence and condensing it into a more compact representation that captures its essential features. This compressed representation acts as a hidden code summarizing the image's key elements while ignoring noise. The decoder then takes this hidden code and expands it back into a new output sequence, aiming to reconstruct the original input but with additional information or a different format.

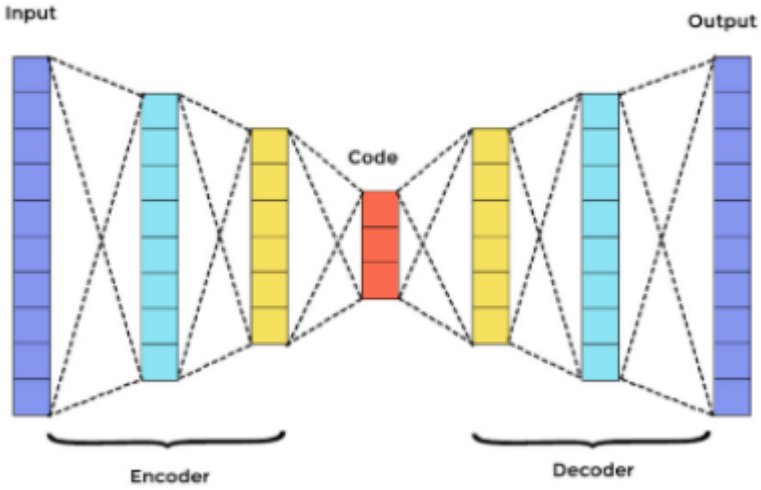

***Figure 1.1**. Encoder-decoder architectures.*

However, standard encoders capture only complex features through downsampling. **U-net**[1] addresses this by adding "skip connections" that act like information highways. These connections bridge the gap between the encoder and decoder, allowing the decoder to directly access the high-resolution details the encoder captured before compression. This is crucial for precise segmentation tasks. **U-net** decoders can not only understand the big picture but also meticulously recreate the fine details.

Spatial Biology: Traditional biology often studies molecules and processes in isolation. However, a cell's behavior and function are heavily influenced by its location and interaction with neighboring cells within a tissue. This is where **spatial biology** steps in. It's a revolutionary field that investigates the distribution and organization of molecules within tissues, providing a much more comprehensive understanding of biological systems. Imagine zooming in on a tissue sample and not just seeing different cell types, but also visualizing their precise locations and how they interact with each other through signaling molecules or physical connections. This intricate spatial organization holds the key to unlocking the secrets of development, disease progression, and fundamental biological processes.

A question is how we can analyze this complex spatial information. This is where **cell segmentation** comes into play. It is an important technique to identify individual cells with a tissue image. By accurately segmenting cells, scientists can quantify their properties, analyze their gene expression patterns, and understand how they contribute to the overall tissue function. Traditional segmentation methods often rely on image processing techniques or readily available labeled datasets. However, creating these labeled datasets can be time-consuming and expensive, especially in spatial biology where tissues are intricate and cell types diverse.

While cell segmentation is crucial for identifying individual cells, its true power lies in understanding how these cells interact. This is especially important for scientists studying the interaction between tumors and immune cells, for diagnosis and choosing immunotherapy treatment approaches. Therefore, spatial analysis becomes essential. By adding spatial context to gene expression data, researchers can tackle problems beyond the capabilities of traditional analysis, revealing how cells are positioned and interact within the tissue, and providing a more comprehensive picture of disease processes.

In this project, we focus on a critical challenge hindering spatial biology research: the scarcity of labeled data. While cell segmentation is vital for unraveling spatial information, creating meticulously labeled datasets is a laborious and time-consuming process. This bottleneck restricts research progress and limits the applicability of segmentation methods. In this thesis, we developed a novel object segmentation model that combines the power of unsupervised and supervised learning.

2. RELATED WORK

2.1. U-Net:

In the field of medical imaging, U-Net stands out as a cleverly designed convolutional neural network (CNN) architecture. Introduced in 2015 by Olaf Ronneberger and colleagues in their paper "U-Net: Convolutional Networks for Biomedical Image Segmentation," this architecture specifically tackles the challenge of segmenting biomedical images.

The power of the U-Net model lies in its clever two-pronged approach:

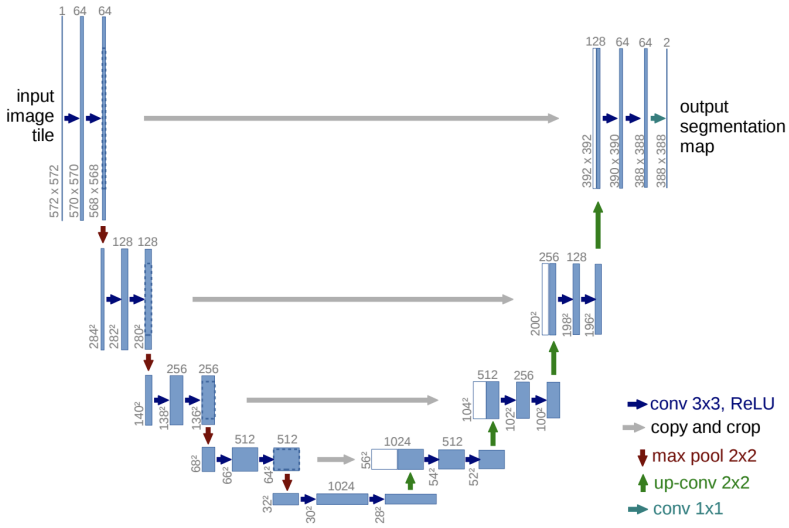

Figure 2.1. *U-Net architecture.*

- Encoder-decoder structure: U-Net follows a typical encoder-decoder architecture. The encoder part captures essential features by progressively downsampling the input image. The decoder part then up-samples the captured features and refines them to produce a segmentation map with the same resolution as the input image.
- Skip connections: A key feature of U-Net is the use of skip connections. These connections bridge the encoder and decoder paths, directly feeding spatial

information from the contracting path to the expanding path. This helps the model retain precise localization details and improves segmentation accuracy.

By combining these elements, U-Net effectively addresses the challenge of limited datasets in medical image segmentation, making it a valuable tool for this field.

2.2. CellPose:

CellPose[2, 3], introduced in the paper "CellPose: A Versatile Trainable Method for Cell Segmentation" by Annemijn Pannekamp et al. (2019), is a deep learning framework designed specifically for cell segmentation in microscopy images. It joins a growing field of related works that leverage deep learning for this task. CellPose offers a general-purpose solution that can handle various cell types and staining methods, without the need for extensive training on specific datasets.

Cellpose leverages the power of convolutional neural networks, specifically the U-Net architecture, for cell segmentation. However, it introduces several modifications to enhance efficiency and performance:

- Efficient Upsampling: Instead of the usual feature concatenation during upsampling in U-Net, Cellpose employs direct summation. This reduces the number of parameters in the model, making it more lightweight.
- Residual Power: Standard U-Net building blocks are replaced with residual blocks. Renowned for improving performance in deep networks, these blocks allow Cellpose to achieve better results. To further capitalize on this approach, the network depth is also doubled, aligning with best practices for residual networks.
- Incorporating Image Style: CellPose integrates a global average pooling layer. This layer captures the overall "style" of the image, summarizing its key characteristics. This style information is then fed into later processing stages. The idea is that by understanding the image's style, the network can potentially adapt its computations to better analyze that specific image.

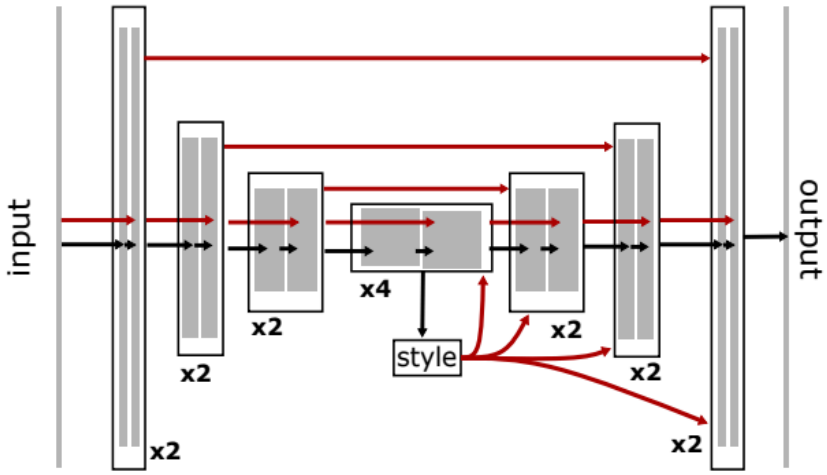

Figure 2.2. *Cellpose architecture.*

Cellpose predicts three outputs (Figure 2.3):

- Flow Fields (X & Y): These indicate the direction (X and Y coordinates) in which pixels should be grouped to form complete cell objects. The flow fields essentially guide the model in constructing the final cell segmentation masks.
- Pixel Confidence: This represents the probability of a pixel belonging to a cell.

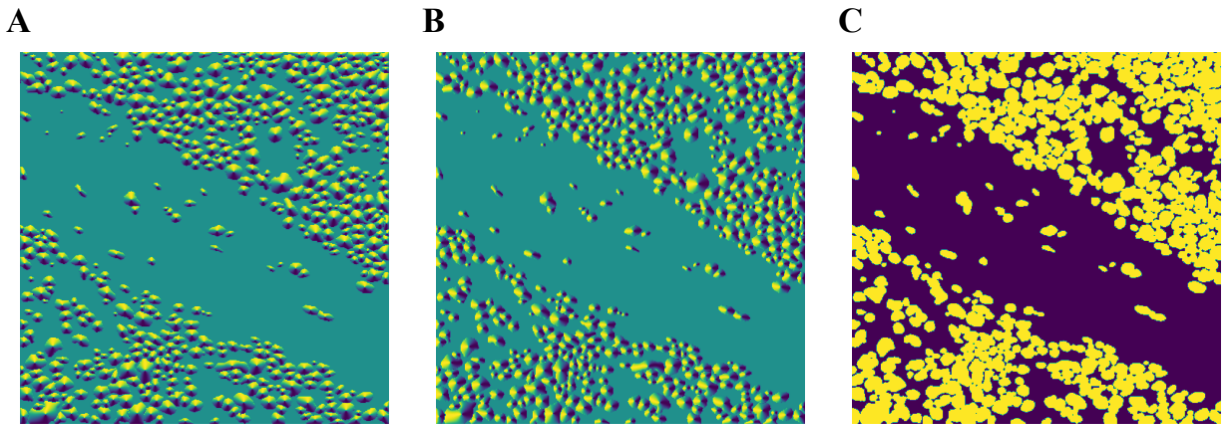

Figure 2.3. *Three layers of output. (**A**) Horizontal gradient. (**B**) Vertical gradient. (**C**) Probability.*

All the 3 channels are used to construct the final cell segmentation masks:

- Thresholding the probability channel (channel 3) to identify likely cell pixels.
- Utilizing the gradient maps (channels 1 and 2) to iteratively refine the segmentation based on the direction of intensity change. This approach is inspired by dynamic systems, where each pixel iteratively updates its location based on the surrounding gradient information (Figure 2.4).
- Constructing a final segmentation mask by grouping pixels based on their "stop locations" obtained through the dynamic system loop.

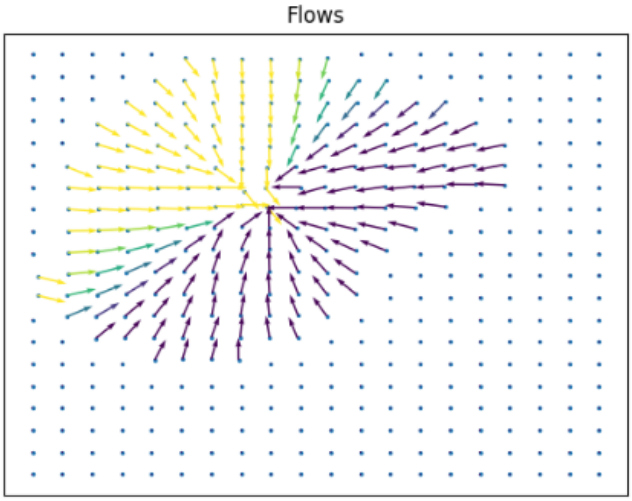

Figure 2.4. *Cellpose's Post-processing: Using the flows to detect the cell center.*

3. PROJECT MANAGEMENT PLAN

Table 1. Project Plan

Task name	Priority	Owner	Start date	End date	Status	Issues
Literature review	High	All	01/01/2024	07/01/2024	Done	
Analyze dataset	Low	All	08/01/2024	21/01/2024	Done	The TissueNet dataset has wrong labels.
Explore Encoder-Decoder based models	High	All	22/01/2024	31/01/2024	Done	A native encoder-decoder model is underperformance in almost all cases.
Explore the Cellpose model	High	All	01/02/2024	15/02/2024	Done	The Cellpose model gives low performance on unseen datasets.
Improve the Cellpose model	Medium	All	16/02/2024	01/04/2024	Done	Hard to improve from the Cellpose model - the well-performing model.
Writing report	Medium	All	15/03/2024	15/04/2024	Done	
Benchmarking	Medium	Nhat	02/04/2024	15/04/2024	Done	
Demonstration	Low	Nghia, Ha	02/04/2024	15/04/2024	Done	

Table 2. Datasets

Items	Link	Description
Data	DeepCell Datasets	The second release of the TissueNet dataset from Greenwald, Miller et al. (Detailed in 4.2.1)
Data	Data Release Program | Vizgen	Vizgen public data release

4. MATERIALS AND METHODS

4.1. Large images presentation:

4.1.1. Large single-cell images:

In contemporary biological research, scientists need to unravel the fundamental biological mechanisms of living organisms. Given that cells serve as the fundamental units of all living organisms, it is only natural for researchers to place significant emphasis on comprehending the internal workings of cells. Consequently, molecules responsible for cellular functions have become a focal point of intensive study across nearly every field of biology today. Moreover, the demand for observing and quantifying information at high resolutions is exceedingly high.

High-resolution images captured through microscopy or other imaging techniques, often referred to as spatial datasets, enable researchers to scrutinize individual cells in intricate detail. Various technologies such as Merscope[5], Cosmx[6], Visium, and Xenium[7], etc., have been developed to generate diverse studies encompassing different types of single-cell spatial data. Notably, Merscope and Cosmx stand out as technologies capable of facilitating studies with exceptionally large single-cell images, reaching up to several billion pixels and requiring several terabytes of storing all of them.

As a result, analyzing and fully utilizing these studies present considerable challenges, yet human resources are insufficient to address all aspects. Consequently, artificial intelligence (AI) emerges as a crucial tool for undertaking this demanding task.

4.1.2. Storage and Accession:

With many advanced technologies, we have obtained diverse datasets encompassing various spatial images. However, this has resulted in a significant volume of data in multiple formats. Storing this data for easy access poses a major challenge, particularly when dealing with large amounts of single-cell spatial images. Therefore, we have chosen a unified format across all technologies to systematically store both the original data and analysis results. This approach ensures seamless access and interaction with the data, especially for crucial single-cell spatial images. This format is referred to as SpatialData[8].

4.1.2.1. Zarr structure:

Zarr is a chunked, compressed, N-dimensional array (Figure 4.1). It is particularly useful for storing large datasets that don't fit into memory, as it allows for efficient access to chunks of data on disk.

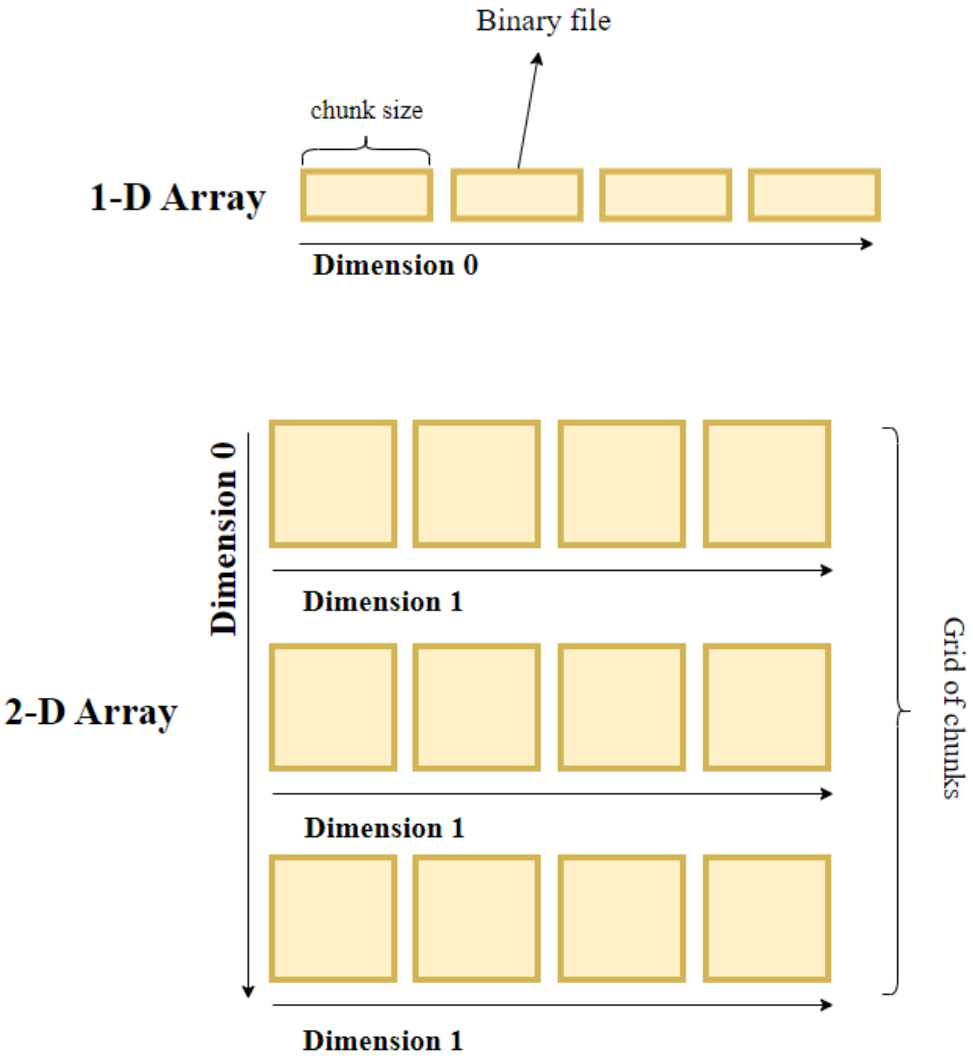

Figure 4.1. *Zarr structure.*
An array is stored in binary files by dividing it into fixed-size chunks and saving them as separate groups of files on the disk.

Zarr stores data in a hierarchical format, similar to a file system, where arrays are organized into groups and subgroups. Each array is divided into fixed-size chunks, which are stored as separate blocks on disk. This chunked storage allows for efficient reading and writing of data, as only the necessary chunks need to be loaded into memory at any given time.

4.1.2.2. SpatialData format:

SpatialData is a framework that establishes a unified and extensible multi-platform file format, lazy representation of larger-than-memory data, transformations, and alignment to common coordinate systems. This format is represented using five primitive elements: Images (raster images), Labels (e.g. raster segmentation masks), Points (e.g. molecular probes), Shapes (e.g., polygon regions of interests, array capture locations, etc.), and Tables (e.g., molecular quantifications and annotations) (Figure 4.2).

- **Images**: images are raster data that store high-resolution microscopy images. They are stored as Zarr arrays (OME-NGFF[9]) and are represented in memory as (multiscale) SpatialImage.
- **Labels**: labels are raster data that contain regions of interest such as segmentation masks. They are stored similarly to images as Zarr arrays (OME-NGFF[9]) on disk and represented in memory as (multiscale) SpatialImage.
- **Shapes**: shapes are polygon data that contain regions of interest such as cell segmentations, capture locations of array-based spatial transcriptomics data, or other types of ROIs. They are stored as a series of arrays that contain coordinates and offsets of the polygons as Zarr arrays (OME-NGFF[9]) on disk and represented in memory as Shapely objects in the GeoPandas DataFrame[10].
- **Points**: points contain large collections (typically in the order of millions or billions) of coordinates and annotations such as transcript locations and their associated metadata. They are stored as a parquet file on disk and represented in memory as a lazy object with a DaskDataFrame[12].
- **Tables**: tables store molecular profile information (gene expression, protein expression, etc.) and associated metadata for observations and variables. It also stores the adjacency matrix of spatial graphs as well as any relevant additional metadata. It is stored on disk and represented in memory as AnnData[11].

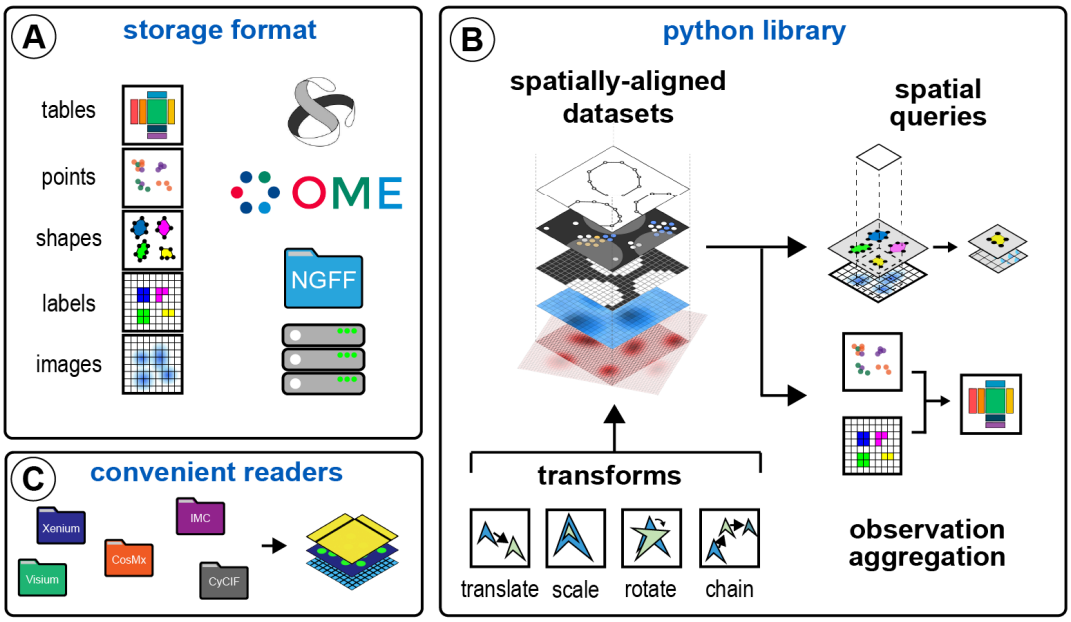

Figure 4.2. *Design overview and core functionality of SpatialData*[8].

(A) The SpatialData storage format provides uniform storage for raw and derived data of diverse spatial omics technologies. The format builds on five primitive elements (SpatialElements), stored to Zarr in an OME-NGFF-compliant manner.

(B) The SpatialData Python library implements core operations for data access, alignment, queries, and aggregation of spatial omics data. Transformations align multiple spatial elements to a common coordinate system (CCS). The CCS allows unified spatial queries and aggregation operators to be deployed across datasets.

(C) The SpatialData provides access to various data formats, including vendor-specific file formats. Multiple datasets can be stored in a single file and together are represented as a SpatialData object.

4.1.2.3. Advantage of SpatialData Format:

Inherited OME-Zarr, using Zarr arrays as a backend, SpatialData contains many advantages of findability, accessibility, interoperability, and reusability (FAIR).

SpatialData is organized either as a tree of folders or files, with each element being entirely independent. This ensures that there are no fragments left behind when data is deleted or new data is appended, not as in formats like HDF5 or TileDB. Since only some small tiles of spatial data need to be read simultaneously, discontinuous data storage not only poses no issues but also has a significant benefit in parallel data transfer across the network.

Moreover, the components of SpatialData are delineated, simplifying access. Presently, a majority of image data is stored as TiffFile, though the existence of numerous formats of a

TiffFile necessitates human intervention for data reading and processing. Additionally, TiffFile is governed by a Header, complicating interactions with the file (such as adding, deleting, editing, transferring, and so on). In contrast, Zarr is managed by a JSON metadata file, rendering similar interactions considerably easier. It can be asserted that Zarr represents a significant advancement towards better alignment with image data.

Zarr seamlessly integrates with leading cloud storage platforms like Amazon S3, Google Cloud Storage, and Azure Blob Storage. Its storage of independent elements with extremely small storage units enables rapid data transfer, facilitating seamless remote visualization of SpatialData.

4.1.3. Significance of Cell Segmentation:

Cell segmentation (Figure 4.3) stands as a significant step within single-cell image processing. Within the outcomes furnished by image processing technologies, a vital data type known as transcripts emerges. Transcripts encapsulate details concerning the spatial positioning of mRNA transcripts, each carrying information regarding a distinct gene. By cell segmentation, we discern the respective cells to which these transcripts pertain, thereby facilitating the construction of an expression matrix for the segmented cells.

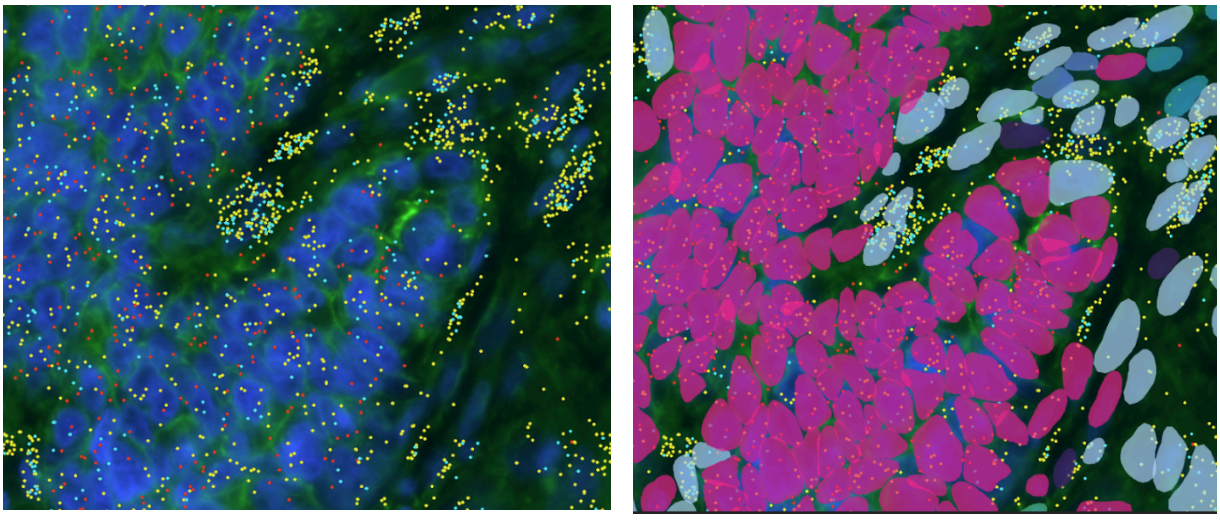

Figure 4.3. *Detecting transcripts based on cell segmentation.*

Transcripts are small points in the figure. Each color of transcripts represents a gene. Segmentation masks (in the right figure) represent cells. Each color of cells is a cell type that is determined by genes.

The expression matrix is a mRNA-seq dataset. mRNA-seq dataset in single-cell refers to a collection of data generated through single-cell mRNA sequencing (mRNA-seq) techniques. This dataset typically contains gene expression profiles from individual cells, capturing the mRNA transcripts present within each cell. Each entry in the dataset represents the expression levels of various genes in a single cell, providing valuable information about the transcriptional landscape of heterogeneous cell populations. These datasets are fundamental in single-cell analysis, enabling scientists to explore cellular heterogeneity, uncover gene regulatory networks, and investigate cellular responses to various stimuli or conditions. Moreover, dimensionality reduction techniques are often employed for visualization or clustering purposes with these datasets. This aids scientists in identifying cell types or enables AI algorithms to predict cell types.

4.2. Cells segmentation:

4.2.1. Materials:

4.2.1.1. Algorithm:

In our research, we leveraged the power of deep learning for the task of cell segmentation in grayscale biological images. This involved employing a deep learning framework, which essentially acts as a powerful software toolbox specifically designed for these types of applications. Within this framework, we utilized the well-regarded Cellpose model, known for its effectiveness in cell segmentation across various image types.

While the Cellpose model achieved good results, we sought to push the boundaries of performance for our specific needs. Therefore, we developed a novel enhancement to the original Cellpose model, which we refer to as the **self-identification** model. This model is described in detail within section 4.2.2.2 of this paper.

4.2.1.2. Datasets:

We utilized two datasets. First, we chose 400 labeled images from the comprehensive TissueNet v1.1 (nearly 200.000 over 1 million labeled cells). Second, we leveraged 160 unlabeled images extracted from Vizgen's Public Data Release (large immunofluorescence cell images, 322MB over 300 GB). Both labeled data were used for all models, but unlabeled images were especially important for training the *Identifying step* for the **self-identification** model.

A. B.

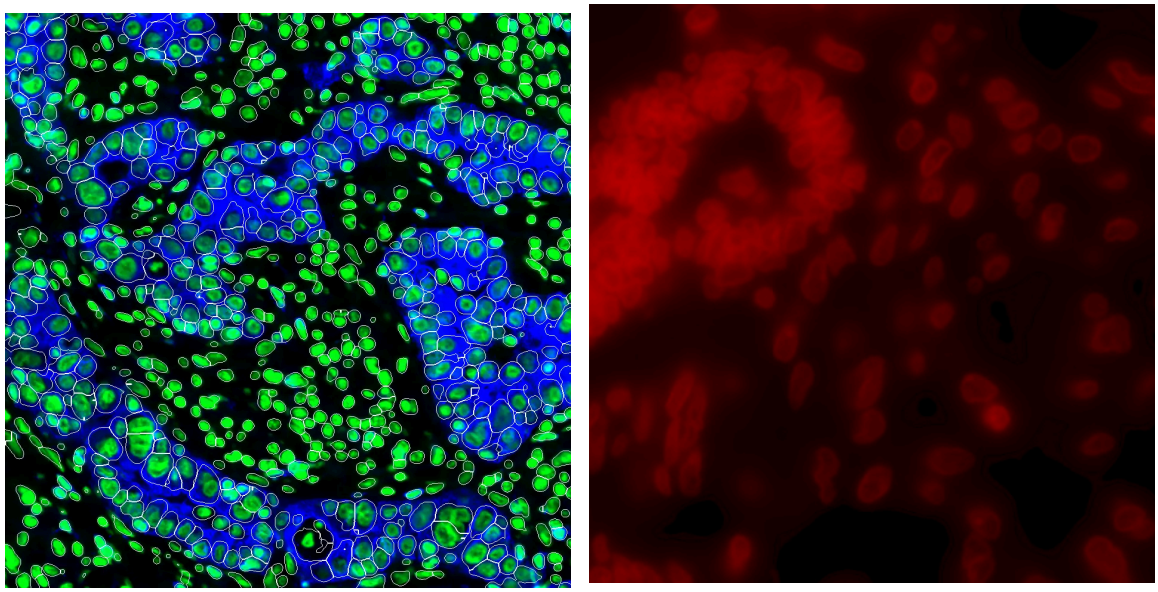

Figure 4.4*. Dataset.*
A. TissueNet Dataset B. Vizgen's Public Data Release Dataset.

4.2.2. Methods:

4.2.2.1. Preprocessing:

- Inputs are converted to correct shapes (224 x 224 x 3) for training and rescaled so that 0.0 is the 1st percentile and 1.0 is the 99th percentile of image intensities in each channel.
- Creating vector flows from labels: The vector gradient flows (Figure 4.6) are derived from a heat diffusion simulation (Figure 4.5), which guides pixels toward the centers of cells while avoiding boundary intersections.

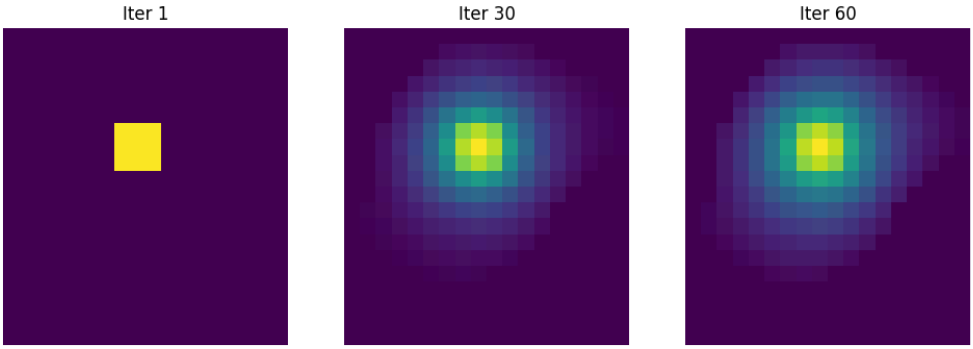

Figure 4.5. *The diffusion of the cell center.*

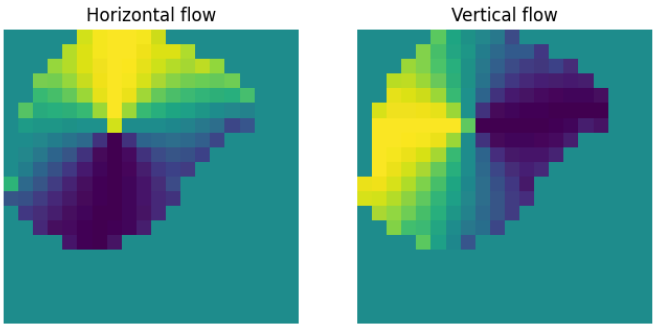

Figure 4.6. *The horizontal and vertical flows.*
The flows are the horizontal and vertical gradients of the diffusion.

4.2.2.2. Self-Identification Model

In the Cellpose model, the way to get results goes directly from images, so it needs to train with many different shapes of cells to define all the ways to segment them. However, in the era of spatial transcriptomics technology, the resolution and number of datasets are increasing rapidly day by day, so labeling all incoming data cannot catch up. Therefore, we try to reduce all cases to a more general pattern, similar to what we are taught about cells before segmenting them. Training segmentation needs labeled data, but training patterns not, this solution may be a good way to handle new incoming data.

Our model builds upon the U-Net architecture by incorporating a self-supervised learning component. This additional architecture aims to reconstruct the original input image, effectively learning by introducing an "information bottleneck." The model leverages the mean squared error (MSE) loss function to guide this reconstruction process.

For the “identifying” step, we utilize the Cellpose model as the backbone. For detail, we’ve adapted Cellpose by employing its Downsampling block as our Encoder and designed a corresponding Decoder block (Figure 4.7). In this way, the identifying step is capable of detecting cellular patterns. Notably, this step is pre-trained on unlabeled data to refine its ability to recognize these patterns.

The information bottleneck of the architecture has the same shape as the lowest layer of Cellpose, we use it to run on the Upsampling block of Cellpose for segmentation. So the computing cost of inference of the self-identification model is the same as the Cellpose model, the images only run through Downsampling and Upsampling blocks to define cells. However, Identifying blocks helps our models detect patterns of cells first, the model will work well on incoming datasets. To keep these advantages, we use the combination of Cellpose’s loss and Identifying loss to define the loss function for the segmentation step.

While segmenting data based on general patterns might lead to slightly lower accuracy compared to directly segmenting for specific features, the significant advantages of

segmenting cells within strange shapes are clearly shown. When the labeled data becomes more diverse, we believe this solution only contains advantages.

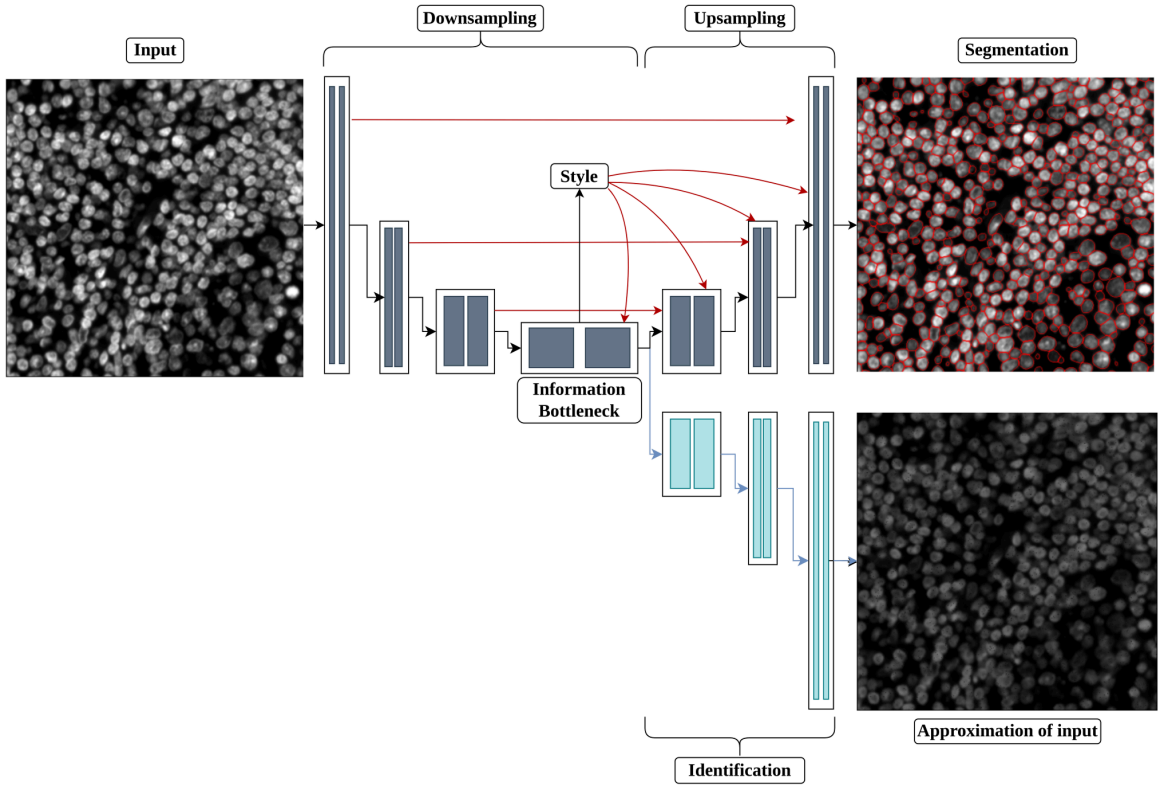

Figure 4.7. *Self-Identification model.*

Cellpose model with unsupervised elements to identify cell patterns.

4.2.2.3. Post Processing:

We do the segmentation step using the following post-processing procedures:

- Thresholding the probability channels:
  - Choose a threshold t, typically 0.5.
  - Only consider pixels with a probability higher than t (belongs to cells).
  - Each pixel is a starting point for the segmentation process.
- The gradient maps (both horizontal and vertical) indicate the direction of change in intensity across the image.
- Dynamic system loop:
  - Each starting point runs for 200 iterations.
  - In each iteration:
    - With each pixel p_{ij}:

        - Find the next grid location $p_{i'j'}$ based on the vector flow (vector created from horizontal and vertical gradients).
        - Update the vector flow of p_{ij} to the vector flow of the $p_{i'j'}$.
    - After this step, the value at location *(i, j)* is called "stop location".
- Construct a mask:
    - Find all stop locations.
    - Slice the image into boxes (shape n×n). In each box, the pixel with the most frequent intensity of stop locations is assigned as a "cell center". Each cell center has a different integer value.
    - The value of each pixel is the value of its stop location.
    - Some pixels inside the cell may be assigned the wrong value, we fill these "holes".

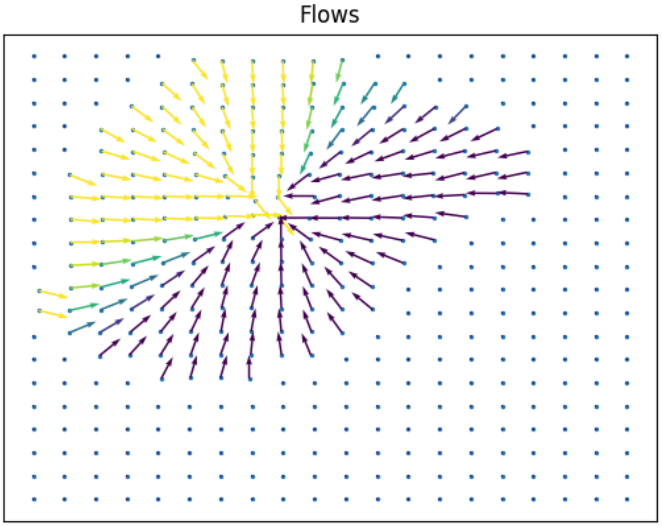

Figure 4.8. *The gradient of the diffusion image.*

4.2.2.4. Parameters:

- Number of epochs: 500
- Optimizer: stochastic gradient descent
- Training parameters: Using default parameters of Cellpose.

4.2.2.5. Loss function:

- Mean squared error (MSE) loss for the gradient channels (horizontal and vertical flows) and self-identification layer.

$$MSE = \frac{1}{n}\sum_{i=1}^{n}(Y_i - \widehat{Y}_i)^2$$

- Binary-cross-entropy (BCE) loss for the probability channels.

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}(Y_i \times log(\widehat{Y}_i) + (1 - Y_i) \times log(1 - \widehat{Y}_i))$$

- Training loss:

$$Training\ Loss = \frac{Loss_{horizontal\ flow} + Loss_{vertical\ flow}}{2} + Loss_{probability} + 2 \times Loss_{self-identification}$$

4.2.3. Evaluation metrics:

We use the F1-score to evaluate the performance of the model:
The F1-score is calculated by:

$$F1 = \frac{precision \times recall}{2 * (precision + recall)}$$

The precision and recall are calculated by:

$$precision = \frac{TP}{TP + FP}$$
$$recall = \frac{TP}{TP + FN}$$

The prediction is quantified by matching each predicted mask to the ground-truth mask that is most similar, as defined by the intersection over union metric (IoU). A detected object is considered a match (true positive TP) if a ground truth object exists whose IoU is greater than a given threshold $\tau \in [0, 1]$. Unmatched predicted objects are counted as false positives (FP), and unmatched ground truth objects are considered as false negatives (FN). We evaluated the predictions at various IoU thresholds (from 0.1 to 1 with step 0.1). The final average precision is averaged over the AP for each image in the test set, using the *"matching_dataset"* function from the Stardist package.
Another metric to evaluate the model is the accuracy score:

$$accuracy = \frac{TP + TN}{TP + FN + TN + FP}$$

In summary, the accuracy score is used to measure how often the model makes correct predictions at the pixel level while the F1 score provides a more nuanced view by considering both false positives and false negatives, making it more suitable for evaluating models.

4.3. Spatial analysis:

4.3.1. Spatial Analysis in Disease Research:

Single-cell transcriptomics focuses on identifying individual cells via profiling gene expression, allowing the discovery of rare cell types and forecasting future development patterns by deducing state transitions of pseudotime. Meanwhile, spatial transcriptomics explores gene expression within tissue contexts, providing information about the spatial arrangement of transcriptomes within cells and tissues. It facilitates the generation of high-resolution relationships of gene expression patterns within their natural spatial context, which holds an important position in disease research.

In disease research, tumor biology, which analyzes the microenvironment and heterogeneity of tumors, is an important part. In tumor biology, analyzing the immune cell topography of solid tumors is a trend nowadays, it is an important predictive factor for the progression of disease and response to immunotherapy. Tiwari, Ankur, et al.[17] mention three patterns of immune cells in solid tumors “Immune inflamed”, “Immune desert” and “Immune excluded” to describe conditions of anti-tumor activity.

“Immune inflamed” is the response of the immune system to harmful stimuli, such as tumors in this context; “Immune desert” is defined as the lack of lymphocytes in either the tumor parenchyma or the periphery of the tumor, and “Immune excluded” is characterized by being confined to the stroma of the tumor and can not penetrate the parenchyma of the tumors. The definition of these patterns so that we can visualize their distribution and

monitor the patient's progress becomes a challenging problem. Previously, some scientists tried to define them by manual annotation, there are several factors in spatial context concerned such as cell densities or cell proportion:

- Galon et al.[13] determined the threshold of cell densities for each cell type in the center of the tumor (CT) and in the invasive margin (IM) and assigned each region as high or low status. Then they classified tumors based on CT/IM status, one of their results, they showed that in CD3 cell type, the best survival is in case high/high, the opposite is in case low/low.
- Desbois et al.[14] developed a method to transfer information from counts of CD8+ T-cells in tumor and stroma region to the number of cells and spatial distribution of them, in there, the quantity of cells describes the number of CD8+ T-cells in tumor and stroma and spatial distribution is determined by the ratio of this two counts (Figure 4.9).

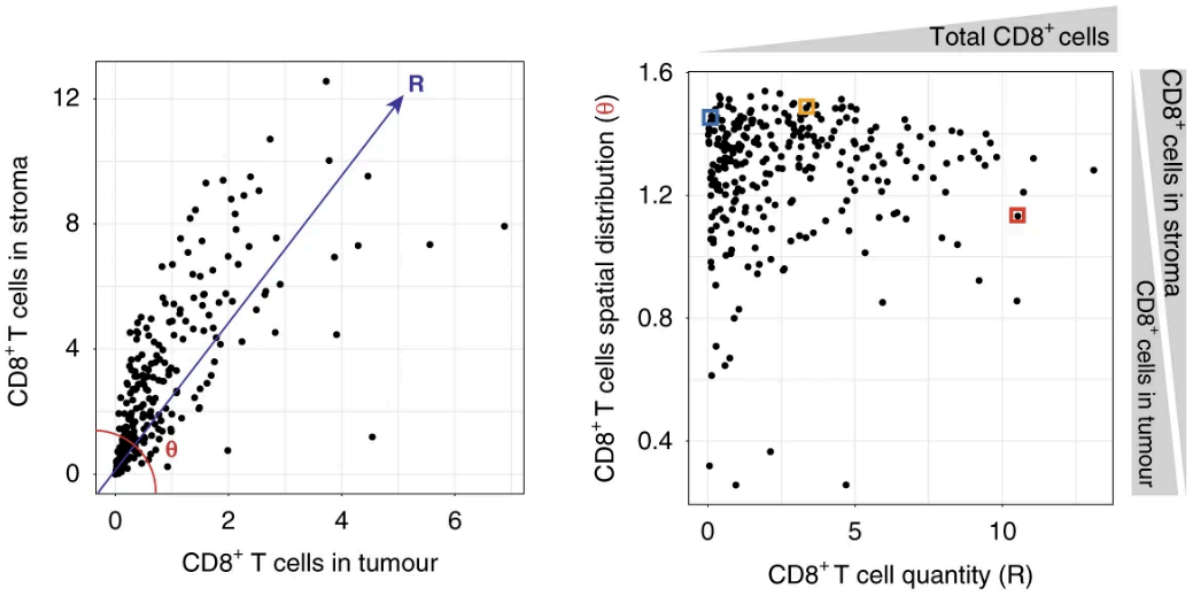

Figure 4.9. *Spatial quantification of CD8+ T cells with a digital image analysis algorithm*[14].

- Hammerl et al.[15] determined anti-tumor activity by measuring CD8+ T-cell at the tumor border and center, for details:

- Immune-inflamed (inflamed): *"almost equal frequencies of CD8+ T cells at the border and center"*
- Immune excluded (excluded): *">10 times more CD8+ T cells at the border compared to center"*
- Immune desert (ignore): *"hardly any CD8+ T cells present at the border and center."*

Hence, they found a significant relation between the phenotypes and survival, tumors with an inflamed phenotype had the best prognosis, and the ignored phenotype was the opposite.

However, spatial data is increasing rapidly, as of April 2024, the TCGA data portal has 11,732 cases with 19.55 Terabytes of data. This moment is when human resources are insufficient, marking the onset of deep learning's entrance into spatial analysis. Besides manual annotation, there are some unsupervised methods used to determine tumor regions for finding out anti-tumor activity status. For instance, Dong, K., & Zhang, S.[4] designed a Graph attention auto-encoder, which network integrates gene expression and spatial information to identify spatial domains. The network uses gene expression as input and clustering information, then builds a Cell type-aware attention layer from a spatial graph and the clustering information.

In conclusion, spatial analysis is a powerful tool for disease research. Along with the rapid increase of data, the utilization of deep learning for analysis is gaining popularity steadily. This seems like an inevitable trend that cannot be stopped.

4.3.2. Spatial Analysis with Spatial Neighborhood Graph:

4.3.2.1. Construct a spatial neighborhood graph:

In many cases of spatial image analysis, we want to explore the relationship between objects in an image. A spatial graph is an approach that assumes objects near others have a

relationship with them. A spatial graph is a graph of spatial neighbors with observations as nodes and neighbor-hood relations between observations as edges. Two well-known graphs are the k-nearest neighbors graph (k-NN graph), Delaunay, and the geometric graph.

Let $V = \{ v_1, v_2 ..... v_n \}$ be a set of points in $\boldsymbol{R^n}$. The nearest neighbor of v_i is a point v_j *(j != i)* with minimum distance *d(i, j)* from v_i. The distance *d(i, j)* is usually Euclidean distance. To make the nearest neighbor unique we choose the point v_j with minimum index in case of the same distance, and denote it by $nn(v_i)$. For any *v*, we define the directed edge *e(v) = (v, nn(v))*. The nearest neighbor graph of *V*, denoted by *NNG(V)*, is the directed graph *{V, E}* where *E = {e(v) | v in V}*.

4.3.2.1.1. K-nearest neighbors graph (k-NN graph):

In a k-NN graph, each set V_i contains exactly *k* nearest neighbors of v_i. In total, *NNG(V)* contains $n \times k$ elements. In the pairwise approach, the computational complexity is $O(n^2)$.

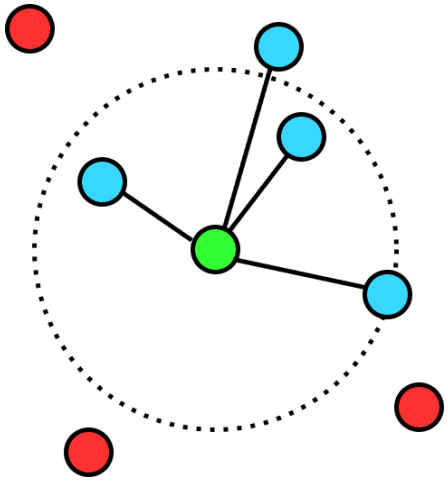

Figure 4.10. *k-NN graph (k = 4).*

The green point is the point of interest, blue points are the nearest neighbors of the point of interest. red points are not the neighbors of the point of interest.

In some approximate approaches, the accuracy is calculated from the number of intersections of the approximate graph and the exact graph divided by k.

One simple idea to find an approximate graph is to calculate the "neighbors of neighbors" from a random state. This method is called NN-Descent. The following are the steps of the NN-Descent method:

1. Start with a random graph (connect each node to random nodes)
2. For each node:
3. Measure the distance from the node to the neighbors of its neighbors
4. If any are closer then update the graph accordingly, and keep only the closest
5. If any updates were made to the graph then go back to step 2, otherwise stop

This is an efficient approach to finding nearest neighbors because we do not need to calculate the pairwise distance of 2 points in the graph, and it is easy to parallel. The empirical cost of this algorithm is around $O(n^{1.14})$.

4.3.2.1.2 Geometric graph:

In a geometric graph, each set V_i is a set that contains all v_{ij} such that $d(i, j) < r$. r is called the *"radius"* of the graph. In the pairwise approach, the computational complexity is $O(n^2)$.

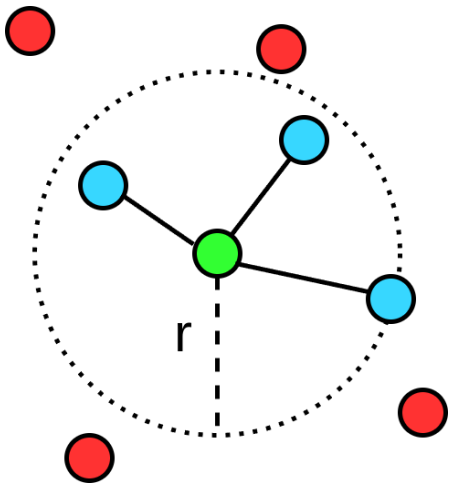

Figure 4.11. *Geometric graph.*

The green point is the point of interest, blue points are the nearest neighbors of the point of interest. red points are not the neighbors of the point of interest.

4.3.2.2. Spatial auto-correlation (Moran's I statistics):

Spatial autocorrelation refers to the phenomenon where the values of a variable at one location are related to the values of the same variable at nearby locations.

In simpler terms, the spatial autocorrelation tells the clustering tendency of similar things. There are two main types of spatial autocorrelation:

- Positive spatial autocorrelation: similar values cluster together
- Negative spatial autocorrelation: dissimilar values cluster together.

Moran's I test is a popular tool to measure this clustering. The Moran's I statistic for matrix X and weight W is computed as:

$$I = \frac{n}{\sum_{i=1}^{n}\sum_{j=1}^{n} W_{i,j}} \times \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}\left(W_{i,j} \times Z_i \times Z_j\right)}{\sum_{i=1}^{n} Z_i^2}$$

In which:

- Z_i: the deviation of an attribute for a feature i from its mean $\left(X_i - \overline{X}\right)$.
- n: total number of features.

The test with confidence interval CI with a permutation picked uniformly at random:

- H_0: No spatial autocorrelation (p-value: P = 0).
- H_1: Spatial autocorrelation exists (P ≠ 0).

Moran's I score z_I for the statistics I is computed as: $z_I = \frac{I - E[I]}{\sqrt{V[I]}}$

where:

- $E[I] = -\frac{1}{n-1}$
- $V[I] = E\left[I^2\right] - E[I]^2$

P is calculated based on the normal continuous random variable. When P < 1 - CI:

- z_I < 0:positive spatial autocorrelation.
- z_I < 0: positive spatial autocorrelation.

Else there is no spatial autocorrelation.

4.3.2.3. Portrait of Graph and Graph Compass:

4.3.2.3.1. Portrait of Graph:

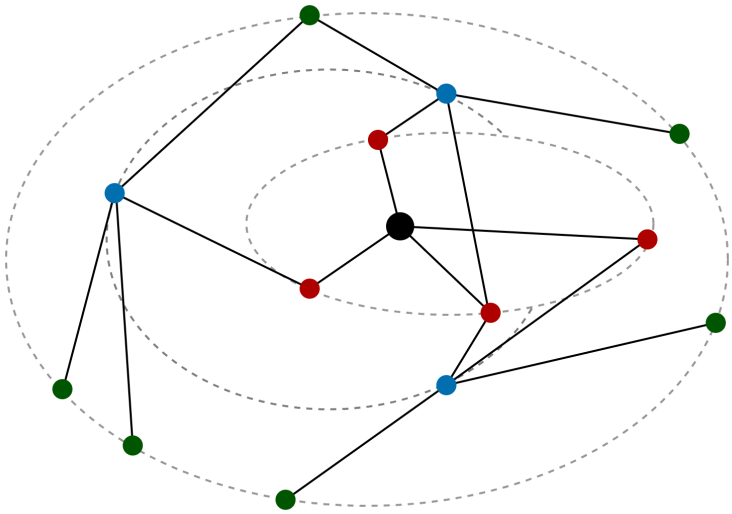

Figure 4.12. *l-shell a node and its connections.*

In this case, the node will be counted into $B_{0,1}$,$B_{1,4}$,$B_{2,3}$,$B_{3,5}$ of the portrait graph. Distances of the node to others are only used to construct neighborhood graphs and the length of shortest paths is defined as the minimum number of steps required for traversal.

A portrait of a graph[16] represents the overall structure of a graph as a way to visualize, and encode many structural properties by information about the distribution of distances between nodes and degree distribution of nodes.

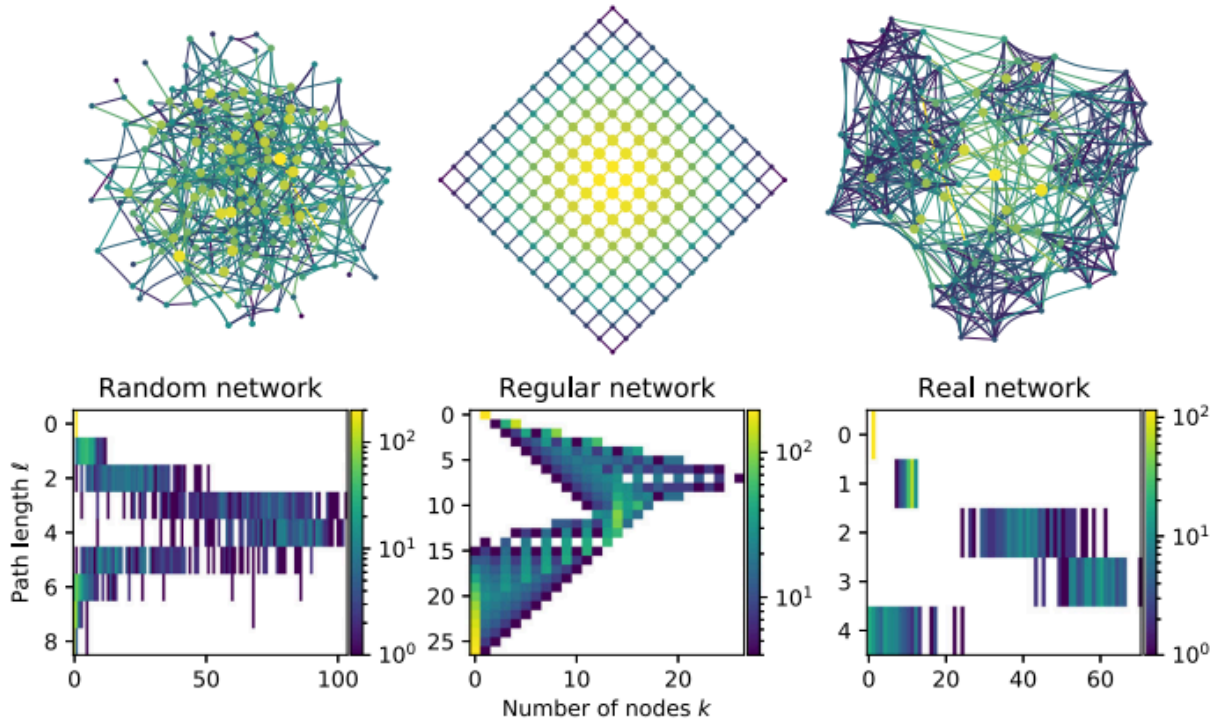

Figure 4.13. *Example networks and their portraits[16].*

Given two graphs G and G', portrait B of each graph is defined as an array with $l \times k$ elements such that $B_{l,k}$ is the number of nodes who has length of shortest path to k nodes is l ($l - shell$). As noted above, we don't compare two graphs directly, instead, we compare their portrait B and B'.

Instead of comparing two graphs, we compare two portrait graphs. First, this method defines some probabilities with N the number of nodes, and n_c is the number of nodes in connected component c:

- Selecting two nodes at random, probabilities they are connected:

$$\frac{\sum_c n_c^{\,2}}{N^2}$$

- The probabilities length of the shortest path (distance of two nodes) is l:

$$\frac{\#\ paths\ with\ length\ is\ l}{\#\ all\ paths} = \frac{\sum_{k'=0}^{N} k' B_{l,k'}}{\sum_c n_c^{\,2}}$$

- Selecting two nodes at random, probabilities one of two nodes has $k-1$ nodes at distance l:

$$\frac{kB_{l,k}}{\sum_{k'=0}^{N} k' B_{l,k'}}$$

Combining those probabilities gives us a probability of choosing two nodes at a distance l and one of them has k nodes at distances l:

$$P(k,\ l) = \frac{kB_{l,k}}{N^2}$$

and similarly with $Q(k,\ l)$ for B'. However, this distribution is only normalized when the neighborhood graph is connected, or

$$\sum_k \sum_l kB_{l,k} = \sum_c n_c^{\,2} = N^2$$

so this distribution is conditioned on the two randomly chosen nodes being connected:

$$P(k,\ l) = \frac{kB_{l,k}}{\sum_c n_c^{\,2}}$$

The similarity of the two graphs was defined Jensen-Shannon divergence

$$D_{JS}(G,\ G') = \frac{1}{2} KL(P \mid\mid M) + \frac{1}{2} KL(Q \mid\mid M)$$

with KL is Kullback-Leibler divergence and $M = \frac{1}{2}(P + Q)$ is the mixture distribution of P and Q.

4.3.2.3.2. Graph Compass:

Single-cell transcriptomics focuses on identifying individual cells via profiling gene expression, allowing the discovery of rare cell types and forecasting future development patterns by deducing state transitions of pseudotime. Meanwhile, spatial transcriptomics explores gene expression within tissue contexts, providing information about the spatial arrangement of transcriptomes within cells and tissues. It facilitates the generation of high-resolution relationships of gene expression patterns within their natural spatial context. However, the similarities and dissimilarities of the spatial context under each specific condition can not be identified by simple subtraction, we need to define the interactions of cells to another in spatial and compare relationships between them via a graph to demonstrate complex connections within separate cells in the same spatial context.

Assuming that we constructed spatial neighborhood graphs of each specific condition, the questions are finding how differences in each condition affect cells. The spatial context is very special, the structure of the connection of a cell is rarely similar to another structure before for the same cell and the same condition, so a comparison metric to detect the most similarity will give us the answer about the interaction of cells within a specific condition is necessary. Portrait of graphs is one of the solutions for this problem, the difference of portraits approximates the difference of their graphs. So instead of comparing two graphs, we compare their portraits.

Ali, Mayar, et al.[18] used GraphCompass to analyze three datasets derived from three different technologies and spatial systems and gave results based on many different comparison methods. By portrait distance method, they generated cell type-specific subgraphs and observed changes in certain conditions over continuous time:

- TOF breast cancer dataset[19]: To investigate the downstream effects of myoepithelial disruption on breast tissue architecture at different scales, they used GraphCompass to compare the spatial organization of myoepithelial cells (MYOEP) in normal breast tissue with progressor tissue and non-progressor tissue, and significantly more similar is found in progressor tissue.
- 10x Genomics Visium heart dataset[20]: To study the effects of ischemic injury at the cellular organization level, they compare samples taken from three regions: ischaemic zone (IZ), remote zone (RZ), and control cardiac tissue. They noticed that the organization of cardiac muscle cells in the RZ differs from that of normal tissue samples and IZ, it is somewhat similar to cardiac muscle cells in control samples. This supports the idea that the damage caused by local vascular injury to myeloid cells is only local to the site of injury.
- Stereo-seq axolotl dataset[21]: To elucidate the molecular events preceding the regeneration of axolotl, researchers removed a part of the brain and collected spatial transcriptomics data focused on the last two regenerative stages 30 and 60 days post-injury (30 DPI and 60 DPI). According to the portrait graph, multiple cell types in the 60 DPI sample resemble the 30 DPI organization more than the adult brain organization, indicating incomplete regeneration 60 days after injury. However, certain cell types in the 60 DPI sample include dorsal pallium excitatory neurons (dpEX) and Sfrp+ ependymal glial cells (sfrpEGC) exhibit spatial organization similar to that of adult cells (ependymoglial cells), implying having progress in the regeneration process (Figure 4.14).

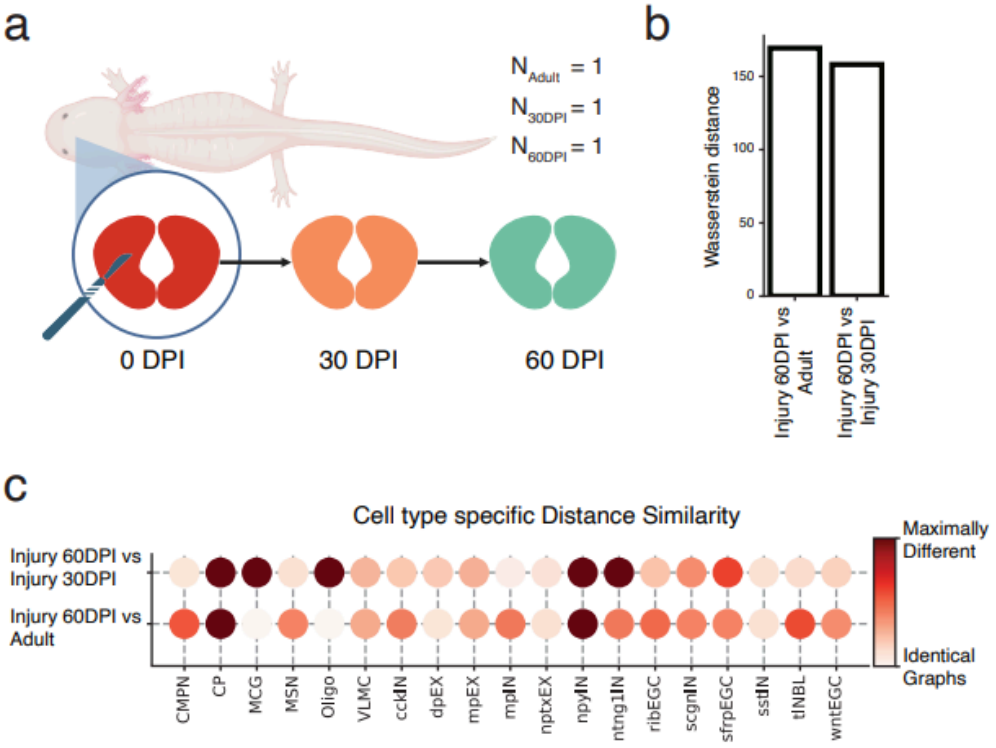

Figure 4.14. *Stereo-seq dataset studying the axolotl brain during development and regeneration.*[19]

a. *Schematic figure describing the subset of regeneration stages we investigated. On the day of injury (0 days post-injury, DPI), a section of the brain was removed. We compared a tissue sample collected 30 days post-injury (30 DPI) with a section obtained 60 days post-injury (60 DPI) and a control sample from an unharmed adult axolotl.*

b. *Comparing entire tissue samples, using Weisfeiler-Lehman Graph Kernels, to show the overall similarity in spatial organization across two stages (Injury 60 DPI versus Adult and Injury 60 DPI versus Injury 30 DPI). The smaller the Wasserstein distance, the more similar the spatial organization is under the two compared conditions.*

c. *Cell-type-specific subgraphs comparison, using the portrait method, across condition pairs (Injury 60 DPI versus Adult, Injury 60 DPI versus Injury 30 DPI).*

5. RESULTS

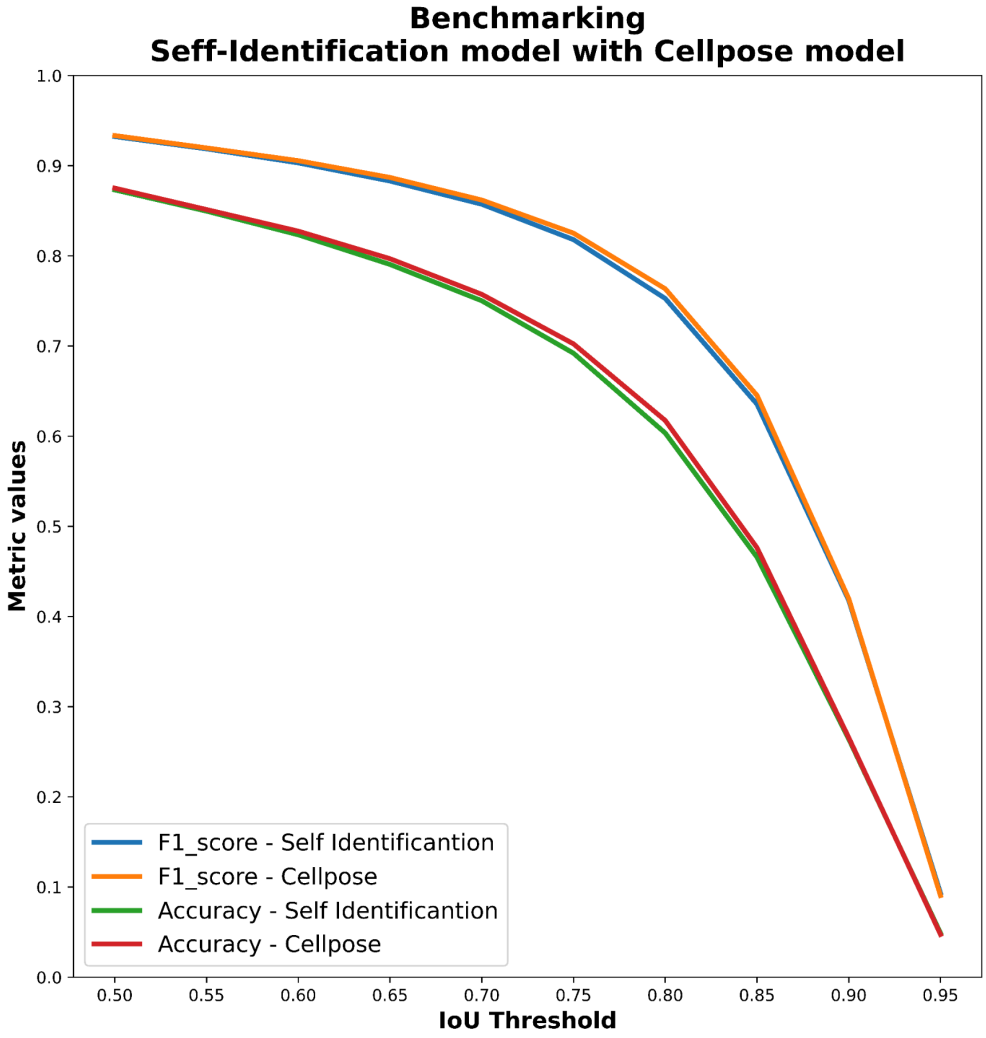

Figure 5.1. *Benchmark models trained with 100 labeled training samples by IoU threshold.*

For the established cell segmentation dataset, both algorithms show similar F1 scores and Accuracy.

Table 3. *Benchmark F1-score and Accuracy for models trained with 100 labeled training samples by IoU threshold.*

	F1 - Score		**Accuracy**	
IoU Threshold	Cellpose model	Self - Identification model	Cellpose model	Self - Identification model
0.50	0.933	0.932	0.875	0.873
0.55	0.92	0.919	0.851	0.85
0.60	0.906	0.903	0.827	0.823
0.65	0.887	0.883	0.797	0.79
0.70	0.862	0.857	0.757	0.75
0.75	0.825	0.818	0.702	0.692
0.80	0.763	0.753	0.617	0.603
0.85	0.645	0.636	0.476	0.466
0.90	0.42	0.418	0.266	0.264
0.95	0.091	0.093	0.047	0.049

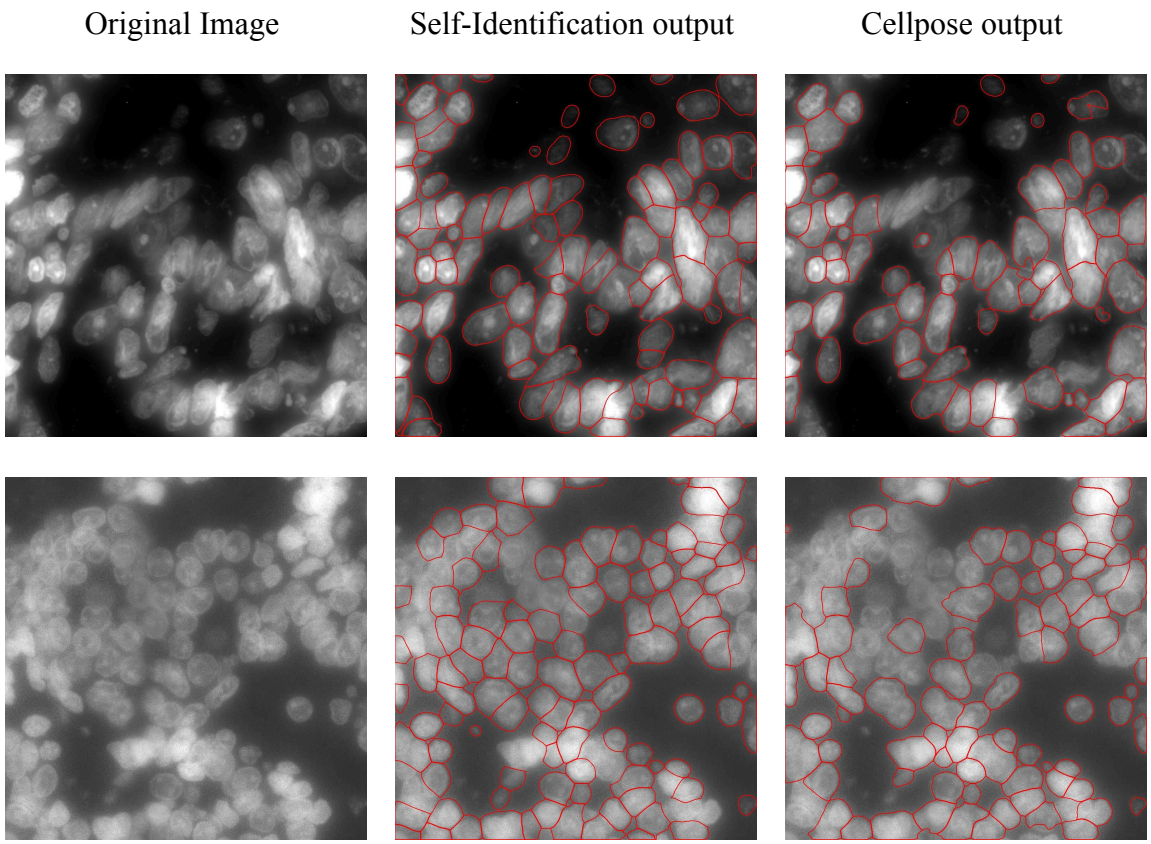

Figure 5.2. *Ability to predict unlabeled datasets.*
For new and unseen datasets (not appearing in training), the Self-Identification model shows significant advantages compared to the Cellpose model.

We benchmark the F1 and accuracy scores of both Self-Identification and Cellpose models by increasing the number of training samples. Figure 5.1 shows that both models obtained similar scores across the number of labeled samples. In some cases, we observe the Self-identification model has a slightly lower score compared to the original version.

In Figure 5.1, we benchmarked the F1 and accuracy scores for the models that were trained using 100 samples across the other IoU threshold. The figure shows that both models have similar scores for all thresholds.

However, when we benchmarked the two models on the out-of-training dataset, and observed a significant improvement in the Self-Identification model compared to the original model in both F1 and Accuracy scores (See Figure 5.2). In particular, the Self-Identification model works well for complex images where cells are located densely and may overlap.

The marginal decrease in performance of the Self-Identification (SI) model compared to the Cellpose model can be attributed to the SI model's unique approach of balancing between the unlabeled and labeled datasets, whereas the Cellpose model prioritizes optimization solely for labeled dataset performance. The SI model's design inherently aims to strike a balance between utilizing both labeled and unlabeled data, potentially leading to a nuanced trade-off in performance compared to the Cellpose model, which focuses primarily on maximizing performance on labeled data.

However, this nominal decrease in performance with the SI model is overshadowed by its notable advantage when faced with out-of-training datasets. The SI model's ability to outperform the Cellpose model in such scenarios underscores its capacity for enhanced generalization and adaptability to real-world applications. Consequently, this slight reduction in performance can be viewed as an acceptable trade-off, underscoring the SI model's contribution as a novel approach to addressing the challenges posed by out-of-training datasets.

6. DISCUSSION

Our capstone project explored the development of a machine learning model for image segmentation, with a focus on adaptability to new datasets. The results presented in Section 5 indicate that our model achieved a slightly lower F1-score compared to the original Cellpose. This is particularly evident when examining the performance at high IoU thresholds.

While a lower F1 score might suggest a less accurate model overall, it's important to consider the context of our project. The fact that our model achieves comparable performance to Cellpose on the original dataset, but surpasses it in adaptability to new datasets, signifies a potential strength. This suggests that our model is more generalizable and less prone to dependence on the labeled training data.

6.1 Leveraging Limited Data:

Traditional machine learning algorithms often hit a roadblock when faced with limited labeled data. This is particularly true in medical imaging, where the labeling process requires significant time and specialized expertise from medical professionals. The scarcity of labeled data can lead to a phenomenon where the model performs well on the training data but fails to generalize to unseen data. This ultimately hinders its real-world applicability.

Our proposed model, however, demonstrates promising results in scenarios with limited labeled data. This suggests that it has the potential to excel in various applications where acquiring large amounts of labeled data is either expensive or impractical. This could be particularly beneficial in the medical field, where access to labeled medical images can be restricted due to privacy concerns or limited resources. Additionally, our model's ability to

perform well with less data could expedite the development of machine learning models in other domains where data labeling is a bottleneck.

6.2. Future work:

Future work could involve investigating how the model performs with even smaller amounts of labeled data. This would further strengthen the case for its effectiveness in scenarios with limited resources. Additionally, exploring multiple techniques, such as transfer learning or augmentation data, could potentially improve the F1 score while maintaining the model's strength in handling limited labeled data.

6.3. Limitation:

It's important to acknowledge that a lower F1 score might still indicate limitations in certain situations. Analyzing the types of cells or image features where the model performs less well compared to Cellpose could help to identify areas for improvement.

7. CONCLUSIONS AND PERSPECTIVES

Spatial biology is transforming our understanding of life by revealing the complex organization of molecules within tissues. But to unlock its true potential, we need to overcome the challenge of limited labeled data. This thesis proposes a new object segmentation model that combines the power of unsupervised and supervised learning, specifically designed to address this bottleneck. By reducing the need for time-consuming manual labeling, this model has the potential to significantly accelerate research in spatial biology. This could lead to groundbreaking discoveries in areas like disease progression, development, and the very fundamentals of how life works.

8. REFERENCES

1. Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In Medical image computing and computer-assisted intervention–MICCAI 2015: 18th international conference, Munich, Germany, October 5-9, 2015, proceedings, part III 18 (pp. 234-241). Springer International Publishing.
2. Stringer, C., Wang, T., Michaelos, M., & Pachitariu, M. (2021). Cellpose: a generalist algorithm for cellular segmentation. Nature methods, 18(1), 100-106.
3. Pachitariu, M., & Stringer, C. (2022). Cellpose 2.0: how to train your own model. Nature methods, 19(12), 1634-1641.
4. Dong, K., & Zhang, S. (2022). Deciphering spatial domains from spatially resolved transcriptomics with an adaptive graph attention auto-encoder. Nature communications, 13(1), 1739.
5. Chen, K. H., Boettiger, A. N., Moffitt, J. R., Wang, S., & Zhuang, X. (2015). Spatially resolved, highly multiplexed RNA profiling in single cells. Science, 348(6233), aaa6090.
6. Merritt, C. R., Ong, G. T., Church, S. E., Barker, K., Danaher, P., Geiss, G., ... & Beechem, J. M. (2020). Multiplex digital spatial profiling of proteins and RNA in fixed tissue. Nature biotechnology, 38(5), 586-599.
7. Ståhl, P. L., Salmén, F., Vickovic, S., Lundmark, A., Navarro, J. F., Magnusson, J., ... & Frisén, J. (2016). Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. Science, 353(6294), 78-82.
8. Marconato, L., Palla, G., Yamauchi, K. A., Virshup, I., Heidari, E., Treis, T., ... & Stegle, O. (2023). SpatialData: an open and universal data framework for spatial omics. bioRxiv, 2023-05.
9. Moore, J., Basurto-Lozada, D., Besson, S., Bogovic, J., Bragantini, J., Brown, E. M., ... & Swedlow, J. R. (2023). OME-Zarr: a cloud-optimized bioimaging file format with international community support. bioRxiv, 2023-02.

10. Jordahl, K., Van den Bossche, J., Fleischmann, M., Wasserman, J., McBride, J., Gerard, J., & Leblanc, F. (2020). geopandas/geopandas: v0. 8.1 (Version v0. 8.1); Zenodo.
11. Virshup, I., Rybakov, S., Theis, F. J., Angerer, P., & Wolf, F. A. (2021). anndata: Annotated data. BioRxiv, 2021-12.
12. Daniel, J. (2019). Data science with Python and Dask. Simon and Schuster.
13. Galon, J., Costes, A., Sanchez-Cabo, F., Kirilovsky, A., Mlecnik, B., Lagorce-Pagès, C., ... & Pagès, F. (2006). Type, density, and location of immune cells within human colorectal tumors predict clinical outcome. Science, 313(5795), 1960-1964.
14. Desbois, M., Udyavar, A. R., Ryner, L., Kozlowski, C., Guan, Y., Dürrbaum, M., ... & Wang, Y. (2020). Integrated digital pathology and transcriptome analysis identifies molecular mediators of T-cell exclusion in ovarian cancer. Nature communications, 11(1), 5583.
15. Hammerl, D., Martens, J. W., Timmermans, M., Smid, M., Trapman-Jansen, A. M., Foekens, R., ... & Debets, R. (2021). Spatial immunophenotypes predict response to anti-PD1 treatment and capture distinct paths of T cell evasion in triple negative breast cancer. Nature communications, 12(1), 5668.
16. J. P. Bagrow and E. M. Bollt. An information-theoretic, allscales approach to comparing networks. Applied Network Science, 4(1):1–15, 2019
17. Tiwari A, Oravecz T, Dillon LA, Italiano A, Audoly L, Fridman WH, Clifton GT. Towards a consensus definition of immune exclusion in cancer. Front Immunol. 2023 Mar 22;14:1084887.
18. Ali, M., Kuijs, M., Hediyeh-zadeh, S., Treis, T., Hrovatin, K., Palla, G., ... & Theis, F. J. (2024). GraphCompass: Spatial metrics for differential analyses of cell organization across conditions. bioRxiv, 2024-02.
19. Risom, T., Glass, D. R., Averbukh, I., Liu, C. C., Baranski, A., Kagel, A., ... & Angelo, M. (2022). Transition to invasive breast cancer is associated with progressive changes in the structure and composition of tumor stroma. Cell, 185(2), 299-310.

20. Kuppe, C., Ramirez Flores, R. O., Li, Z., Hayat, S., Levinson, R. T., Liao, X., ... & Kramann, R. (2022). Spatial multi-omic map of human myocardial infarction. Nature, 608(7924), 766-777.
21. Wei, X., Fu, S., Li, H., Liu, Y., Wang, S., Feng, W., ... & Gu, Y. (2022). Single-cell Stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. Science, 377(6610), eabp9444.