Enhancing Spatial Transcriptomics with Vision Transformers for Single-Cell Resolution Mapping

The overall goal of this project is to improve current spatial transcriptomics methods by taking low-resolution data and transforming it to single-cell resolution by leveraging deep learning architecture, specifically vision transformers.

Authors
Anisha Chatterjee
anishac@andrew.cmu.edu

Young Je Lee
Professor Jose Lugo-Martinez
Carnegie Mellon University

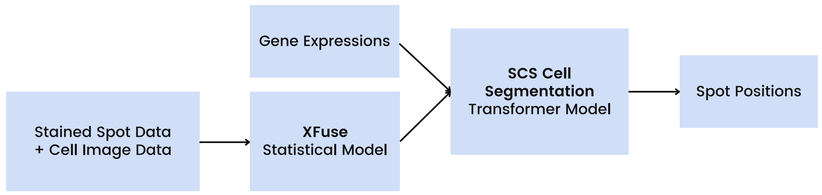
Introduction

- Spatial transcriptomics allows scientists to measure the gene activity in a given tissue sample and map where the activity is occurring.
- It has greatly improved our ability to analyze gene expression within the context of different tissue structures, providing insights into cell-cell interactions and structural organization.

Background

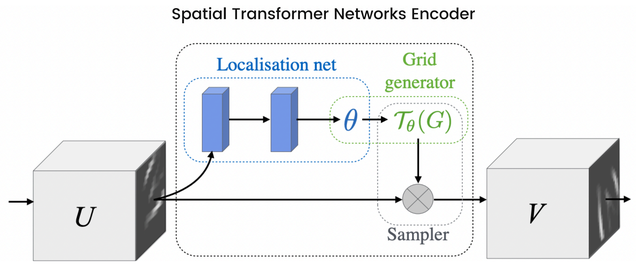
Current spatial transcriptomics method:

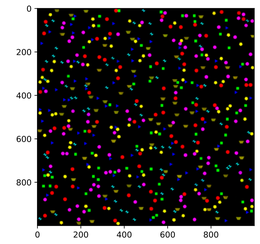
- The pixel-level gene expression profile is created by applying XFuse, which relies on statistical modeling, on stained spot data.
- Using the pixel-level gene xpression profile and additional cell image data, cells can then be spatially mapped and matched to their gene expression.
- The main challenge for this step is cell segmentation, essentially separating cells and understanding which cells belong in which spots.
- A popular method, SCS, first identifies cell nuclei from staining images using the Watershed algorithm. Then, the transformer model infers for each spot its relative position with respect to the nuclei of each cell.

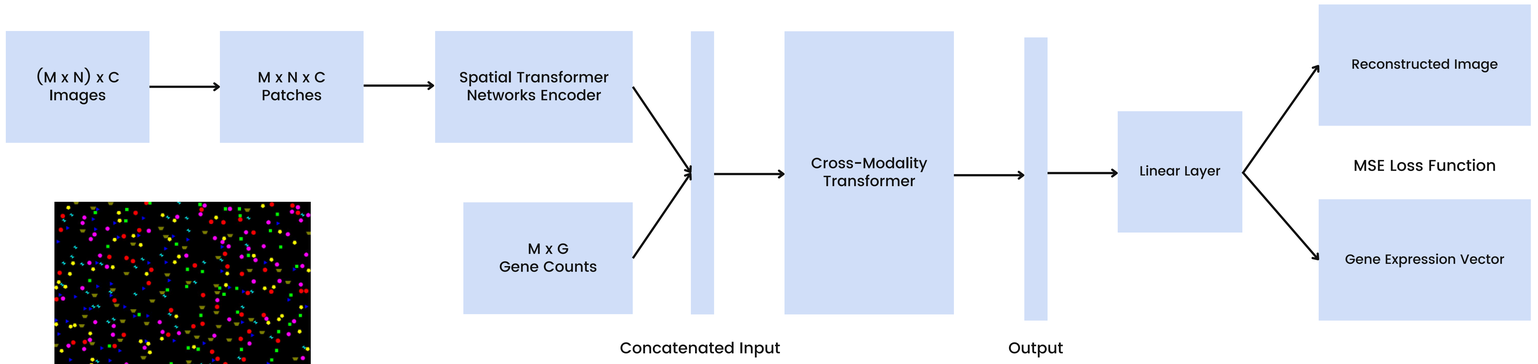


Methodology

M x N = Image Size
C = # Channels
G = # Genes



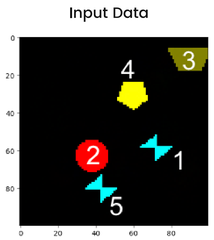


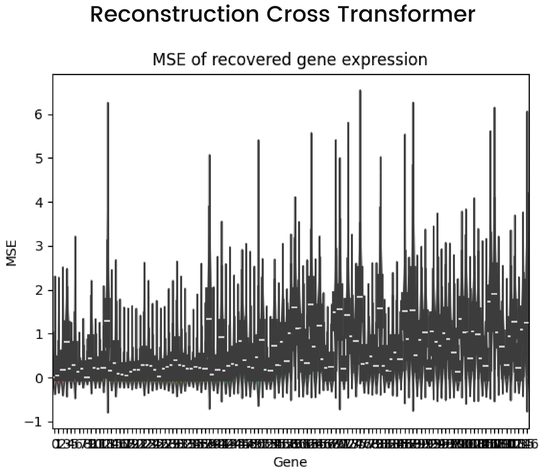


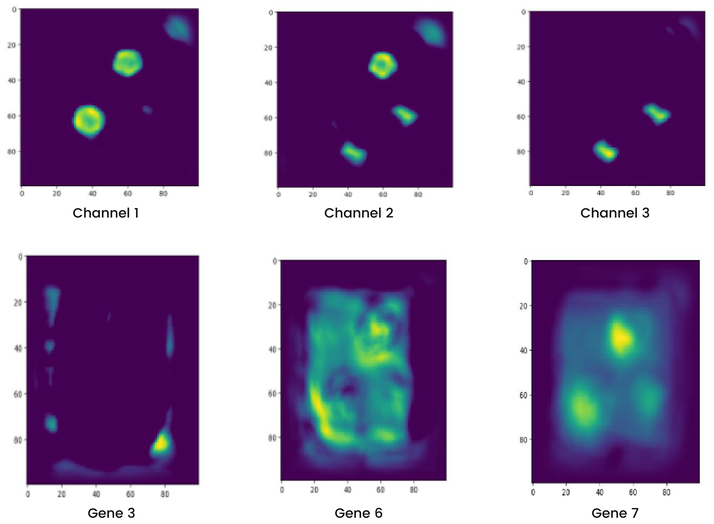
Simulated Dataset

The current architecture is a combination of experiments trying different input sizes, ways to combine the two different types of data, different backbone transformer models, and different decoder structures.

Preliminary Results







Further Work

- We are still working to find better results with the use of the spatial transformer networks encoder. It has previously shown to extract the strongest features of images and significantly reduce computation, a main goal of our project. We are looking at different decoder architectures and pooling techniques.
- Our model currently uses a cross-modality transformer and experimenting with different cross-modality transformers is another method we might explore.
- Another factor that has helped previous papers do well with spatial transcriptomics tasks is to use labeled data where the cell type is considered as well.

References

Jaderberg, M., Simonyan, K., Zisserman, A., & Kavukcuoglu, K. (2016, February 4). Spatial Transformer Networks. arXiv.org. https://arxiv.org/abs/1506.02025