Dr. Jeff Regier, Dr. Jackson Loper, and Roman Kouznetsov on Spatial Transcriptomics

Reading Group 02/09

Presentation by: Roman Kouznetsov

University of Michigan Department of Statistics

Table of Contents

On Genetic Information in Cells

Q: If I wanted see the sequencing information of a cell, how would I do it?

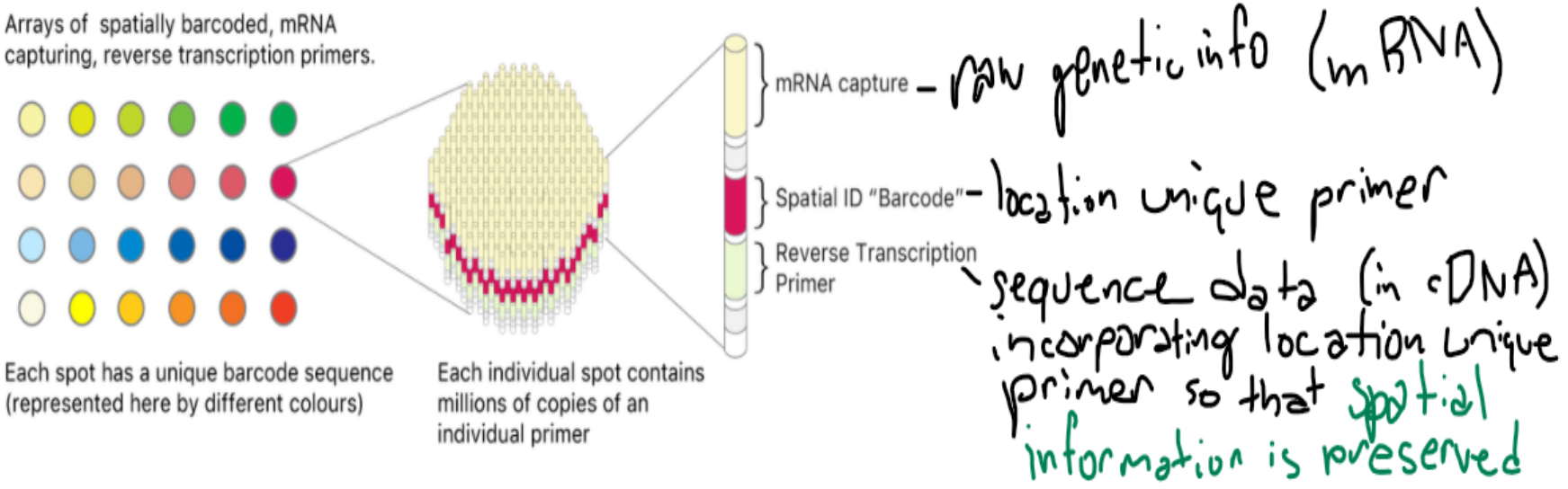
A:

- single cell sequencing
- bulk RNA extraction
- spatial transciptomics

How does ST work?

My Understanding:

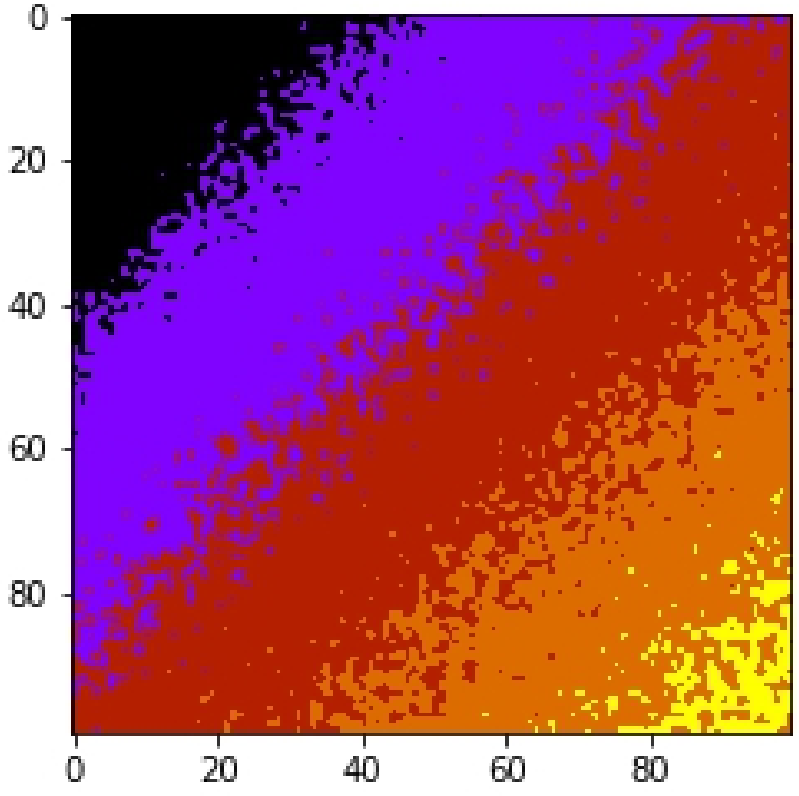
How does ST work?



Original Image Source: James Chell

Importance Sampling

Given a gene and its expressions in a cell, we could learn its orientation θ and intensity v and choose values that maximize a likelihood (likely Poisson).



Problems with Importance Sampling

- Only works well with one directional spatial data. e.g. how would this handle...
  - spatially bimodal data?
  - alternating patterns?
  - identical expression everywhere?
- No direct incorporation of neighbor behavior.

Current Method (enter MoNeT)

MoNeT is a way to perform Gaussian mixture model CNNs in a graph setting.

- Nodes: Cells
- Edges: Neighboring Cells
- Node Attributies: Gene Expression Levels

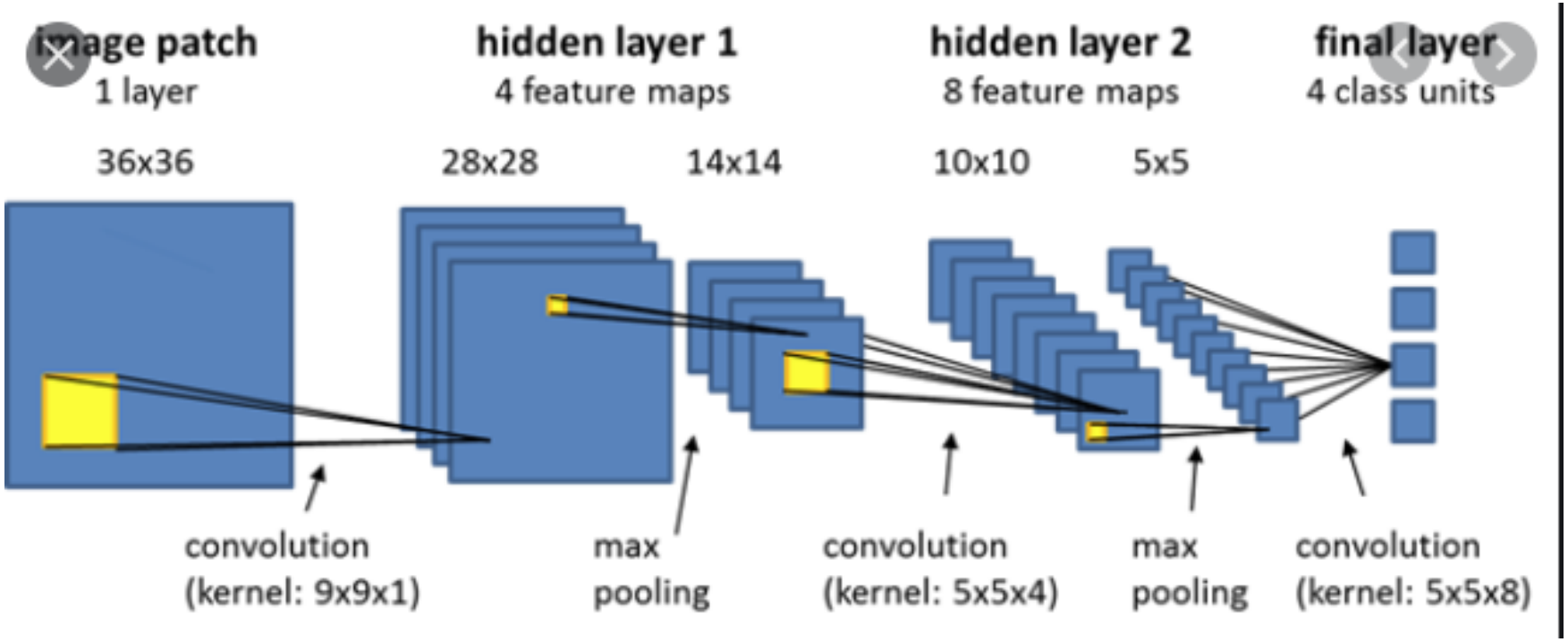


Image Source: Trimble Inc.

Autoencoder Framework

- Tissue Sample $\rightarrow$ Low-Dim Representation $\rightarrow$ Recreation of Tissue Sample
- The low-dimensionality applies to the gene expressions (node attributes) but can also be applied to the size of the graph (number of nodes and edges).
- Evaluated on Negative Poisson Likelihood Loss

Autoencoder Framework

- Tissue Sample $\rightarrow$ Low-Dim Representation $\rightarrow$ Recreation of Tissue Sample
- The low-dimensionality applies to the gene expressions (node attributes) but can also be applied to the size of the graph (number of nodes and edges).
- Evaluated on Negative Poisson Likelihood Loss

Probabilistic Model (VAE)

The current goal of the project is to create a probabilistic model in the form of a VAE.

This allows us to create a probabilistic model where the latent dimension is easily observable and easily sampled from.

Why does this work matter?

- Spatial Transcriptomics was invented in 2016, so not much novel deep learning on genetic spatial data has been done yet.
- Low dimensional data allows us to give some visualization of the spatial genetic data.
- Graph convolutional networks allow us to *directly* use neighboring cells as relevant information for learning.
- Once a probabilistic model is built, we want some measurement of the correlation of gene expressions between neighboring cells (and maybe why it is happening).
  - This leads to an interesting idea of "probabilistically condition on all the things scientists already know about, and see if there's still a strong effect present even after conditioning on everything we know." - Jackson