Determining Spatial Distribution of Missing Proteins using CODEX Imaging and AI

Danni Chen, Zitian Tang, Benoit Marteau, Felipe Giuste, May Dongmei Wang Ph.D.

Abstract

CODEX, or co-detection by indexing, is an imaging approach that allows visualization of the spatial distribution of many proteins within tissue at a microscopic resolution. It uses a novel method of detecting antibody binding events using DNA barcodes and fluorescent analogs to achieve these spatial protein signals. CODEX can be used to label cells within tissue using known combinations of protein expressions specific to each cell type. This allows for deep characterization of cellular niches and dynamics important in a wide range of clinical pathology tasks, such as tumor grading. Recent studies suggest that a deep learning model, called UNET, which is frequently used to identify regions of fluid-filled lung in chest X-ray, can identify the distribution of proteins not directly detected via CODEX-generated images. Traditional cell labeling methods are time-consuming and expensive. With an innovative deep-learning algorithm, we aim to accurately estimate protein distribution within tissue samples and improve the efficiency of the cell-labeling process. We developed a UNET model to reconstruct single-channel protein signals based on the distribution of other proteins within the same microenvironment. During model testing, a total of 56 UNET models are implemented and tested, one for each protein signal. Protein signals which can be accurately reconstructed from others may be ignored in future experiments without significant loss of information. This novel model addresses the needs of researchers seeking to determine cell-level protein expression within tissue samples.

Background

- CO-Detection by indEXing
- Highly-mutiplexed cytometric imaging
- DNA barcodes and fluorescent analogs

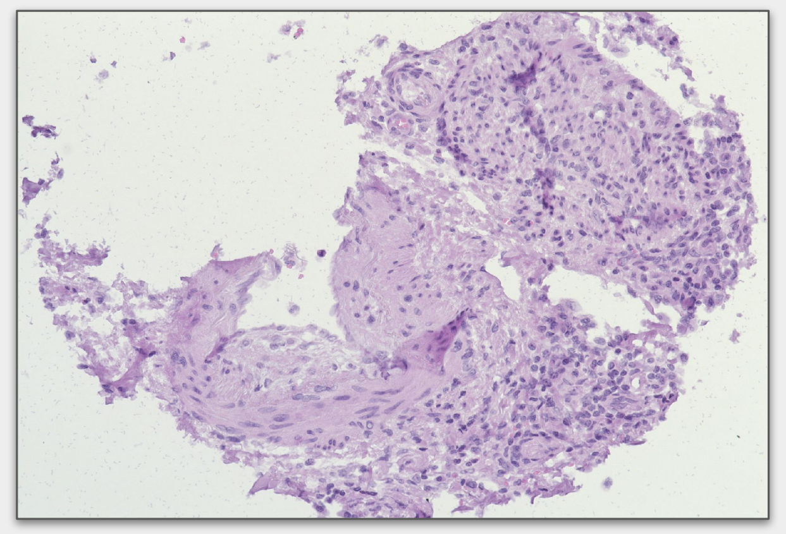

Figure 1. Image showing a specific region of a CLR colorectal cancer tissue region under microscope.

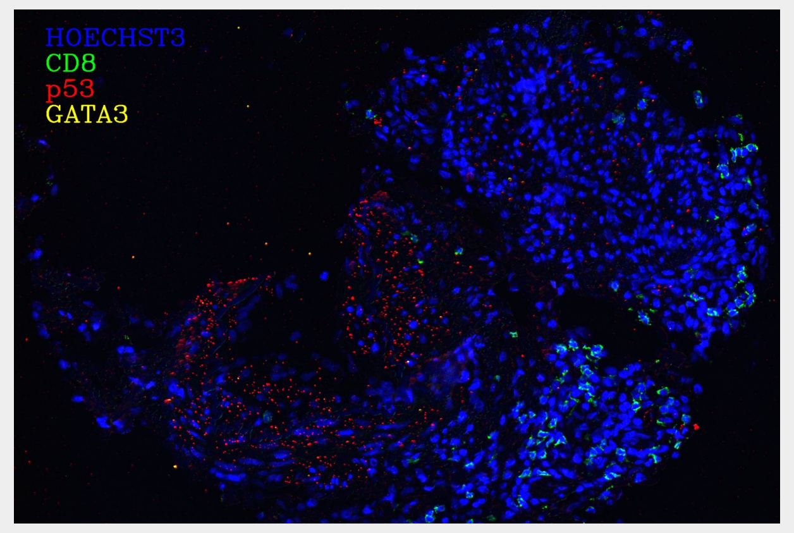

Figure 2. Image of the corresponding CODEX expression, where a nuclear stain, HOECHST3, is labeled in blue, and three protein markers, CD8, p53, and GATA3 are labeled in green, red, and yellow, respectively.

- Currently used for:
  - Characterize and quantify cellular traits
  - Intercellular environment analysis
- Cell labeling:
  - Of 60+ protein signals
  - In multiple cycles
  - 1 protein → 1 channel

Data Description

Dataset Size	35 patients (2 classes: 17 CLR, 18 DII), 140 regions
Data Source	(Schürch et al., 2020)
# of Protein Channels	56 protein markers/channels + 2 nuclear staining
Type of Tissue	Colorectal Cancer Tissue

Tabel 1. Description of the Dataset

Overview and Results

Hypothesis

We can extrapolate the distribution of non-imaged proteins using existing CODEX imaging data.

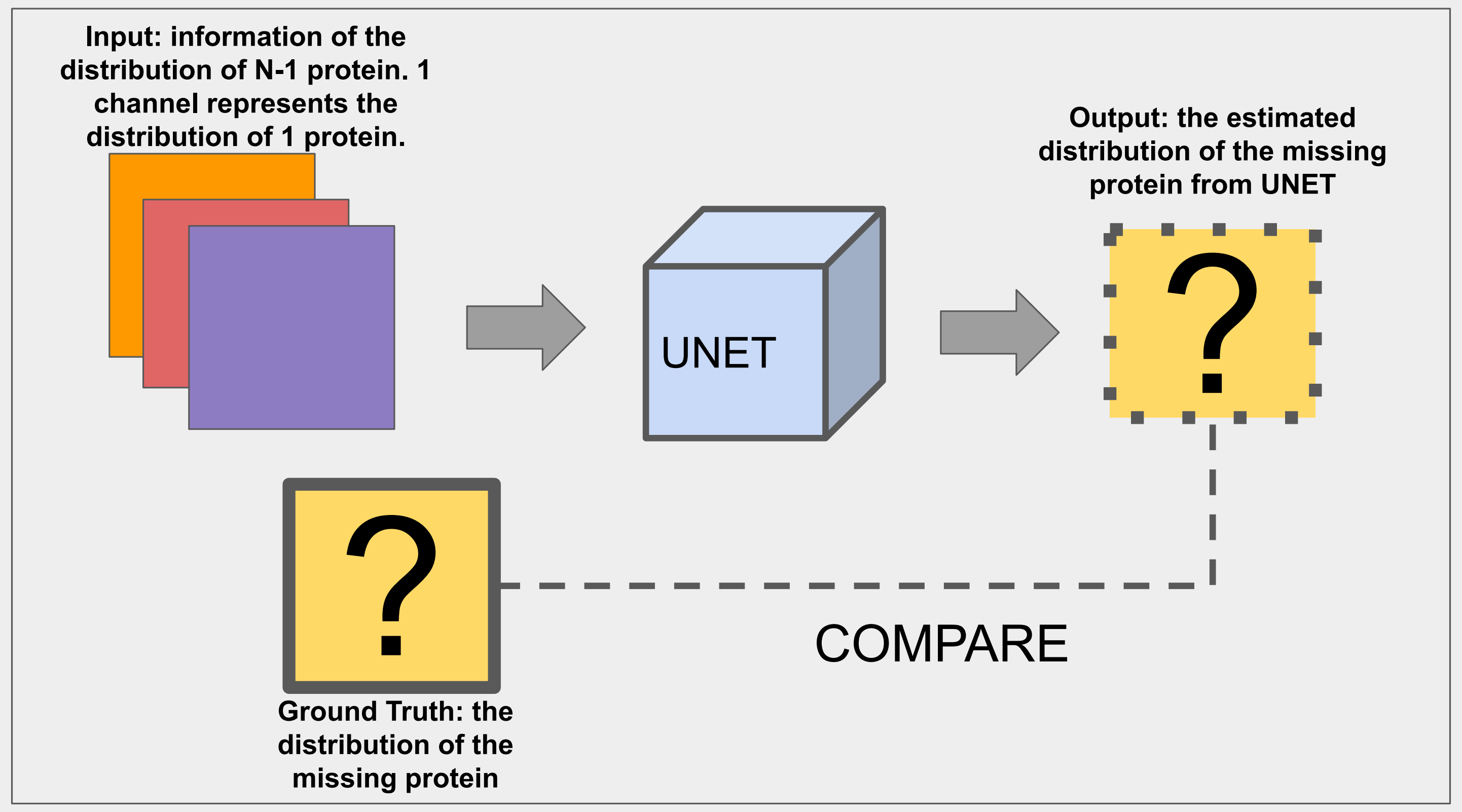
Input: information of the distribution of N-1 protein. 1 channel represents the distribution of 1 protein.
UNET
Output: the estimated distribution of the missing protein from UNET
?
?
COMPARE
Ground Truth: the distribution of the missing protein

Figure 3. General overview of the model, which takes in N-1 channels to represent N-1 protein distribution and outputs a reconstructed channel representing the missing protein distribution. The reconstructed output channel generated by our model would be compared with the ground truth, which is the actual distribution of the missing protein.

Result & Impact

- Goal: Reconstruct all 56 proteins
- Example
  - Input: 3 proteins - CD44, CD3, beta-catenin
  - Output: 1 protein - hoechst (nuclei)

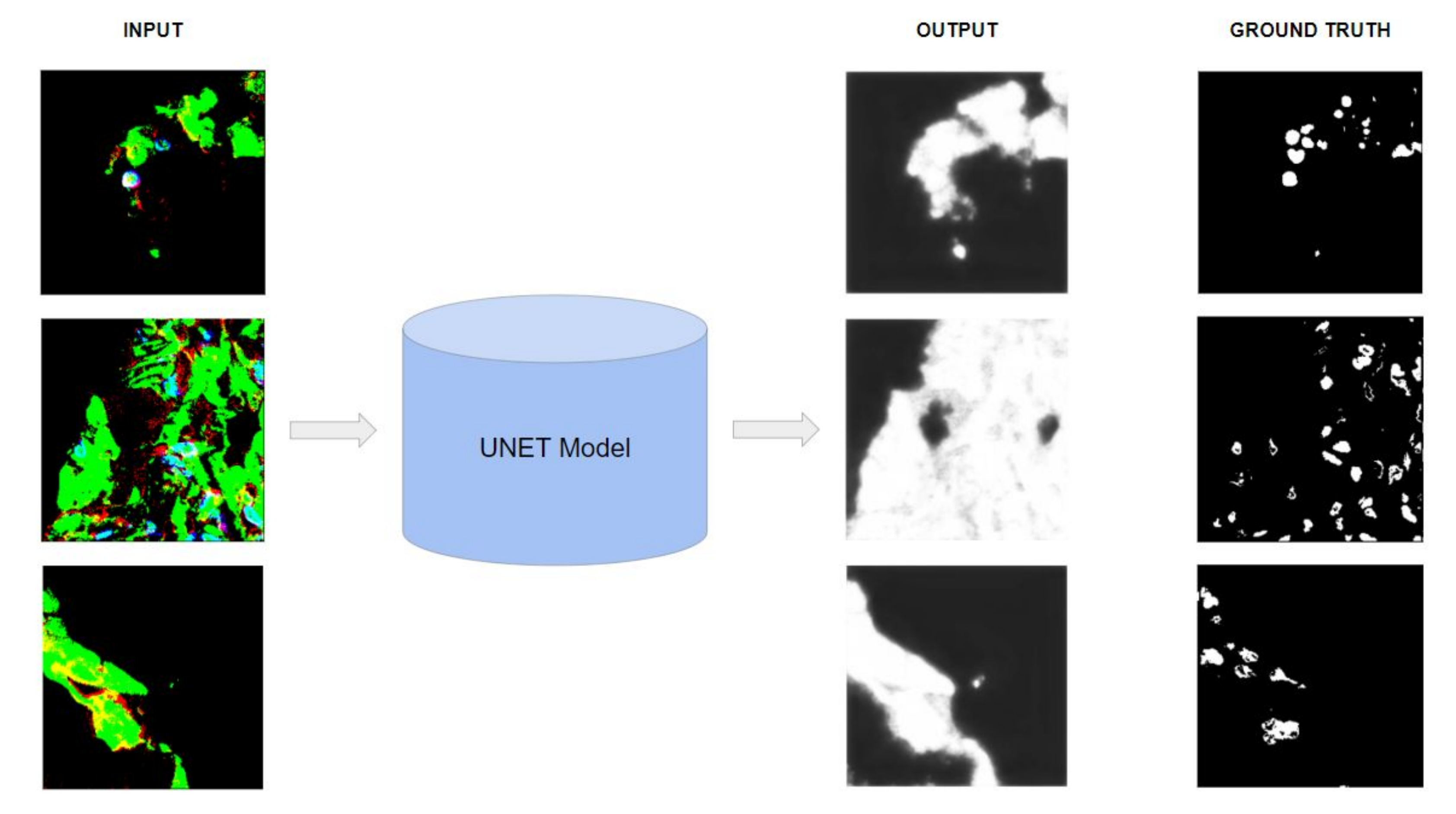

Figure 4. The input of our model is a 256x256 3-channel image, where each channel represents the distribution of CD44, CD3, and beta-catenin, respectively. The output is a 256x256 grayscale image predicting the distribution of a nucleic staining. The ground truth of the nuclei distribution is a 256x256 binary image shown on the far right.

Impact:

Researcher can save time and money by generating synthetic reconstructions of protein distributions.

UNET Structure

Encoder
- Image contraction
- Feature extraction
- Convolutions

Skip connections
- One at each level
- Copy and crop

Decoder
- Spatial expansion
- Mask generation
- Up convolutions

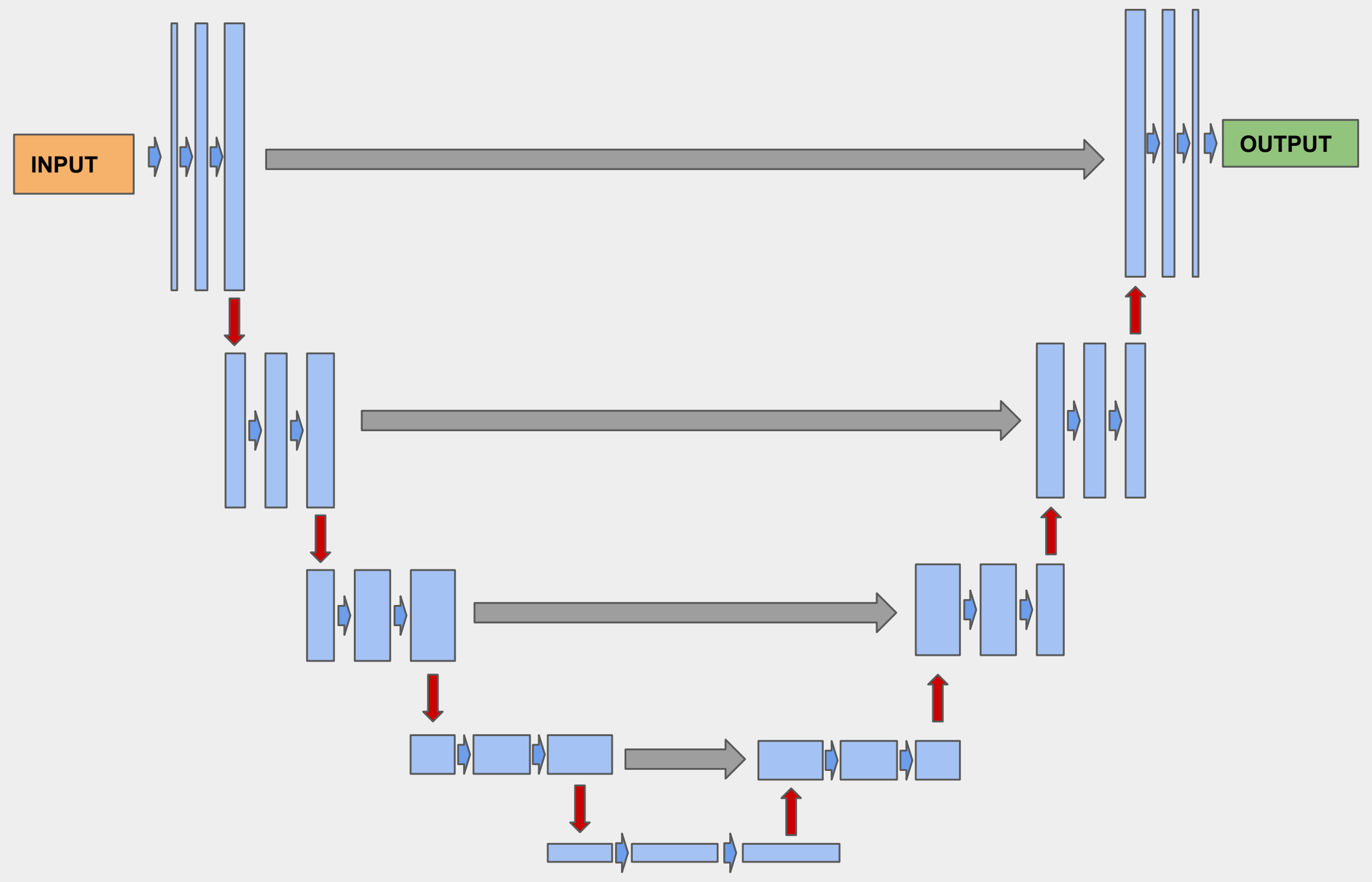

Figure 5. General UNET structure with the left half being an encoder that extracts features and the right half being a decoder that generates binary mask and decodes information. Skip connections at each level carry some information directly to the right side.

Methods

Our model can be defined as: $f(X_{!c}) = \widehat{X_c}$

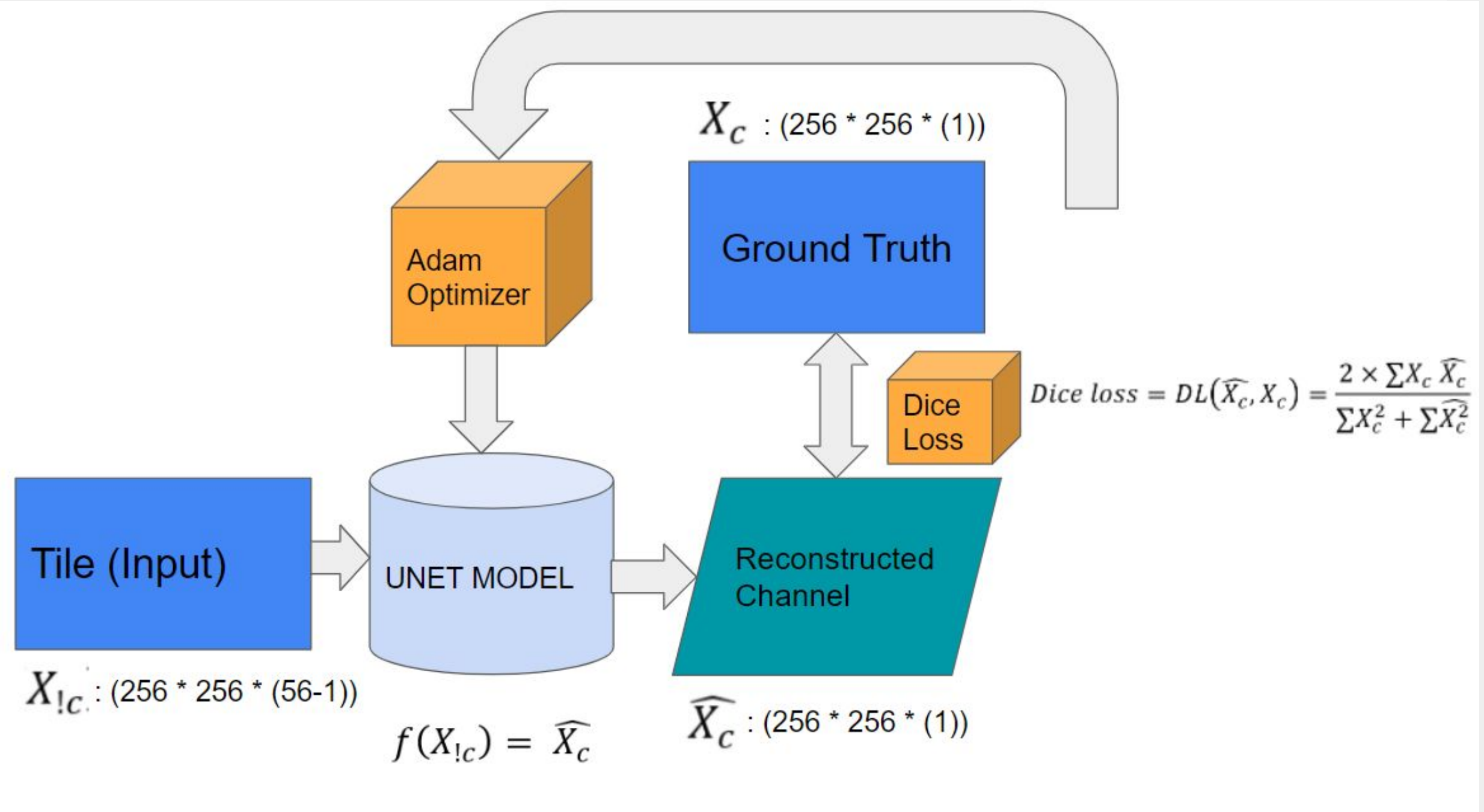
X_c : (256 * 256 * (1))
Adam Optimizer
Ground Truth
Dice Loss
$Dice\ loss = DL(\widehat{X_c}, X_c) = \frac{2 \times \sum X_c \widehat{X_c}}{\sum X_c^2 + \sum \widehat{X_c}^2}$
Tile (Input)
UNET MODEL
Reconstructed Channel
$X_{!c}$: (256 * 256 * (56-1))
$f(X_{!c}) = \widehat{X_c}$
$\widehat{X_c}$: (256 * 256 * (1))

Figure 6. The model pipeline describes the size of the input tile, the ground truth, and the outputted reconstructed channel. Dice loss is used to calculate the difference between the ground truth and the output channel. Adam optimizer is used to optimize the model weights.

Discussion

Current Work:
- Estimate the location of nuclei based on the distribution of CD44, CD3, and beta-catenin.

Future Work:
- Tune model hyperparameters to improve model performance
- Estimate the distribution of all 56 proteins based on the distribution of other proteins

Acknowledgements

We would like to acknowledge that the 2022 Spring Undergraduate Symposium is held by the Undergraduate Research Opportunities Program.

References

[1] Schürch, C. M., Bhate, S. S., Barlow, G. L., Phillips, D. J., Noti, L., Zlobec, I., Chu, P., Black, S., Demeter, J., McIlwain, D. R., Kinoshita, S., Samusik, N., Goltsev, Y., & Nolan, G. P. (2020). Coordinated Cellular Neighborhoods Orchestrate Antitumoral Immunity at the Colorectal Cancer Invasive Front. Cell, 182(5), 1341-1359.e19. https://doi.org/10.1016/j.cell.2020.07.005
[2] PhenoCycler - Akoya. (2021, December 8). Akoya - the Spatial Biology Company. https://www.akoyabio.com/phenocycler/
[3] milesial. (2022, February 19). Pytorch-UNet/unet at master · milesial/Pytorch-UNet. GitHub. https://github.com/milesial/Pytorch-UNet/tree/master/unet