Identifying Protein Distribution in Bacterial Cells

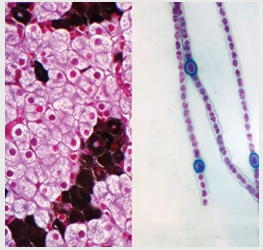
Rose Hudson (Pennsylvania State University), Fangwei Si (Carnegie Mellon University)

Background

Our knowledge of cell structure and composition is severely limited by the size of cells, even more so in bacteria cells due to lack of clear organization.

While a lot is known about which proteins are present within bacteria, the difficulty of accurately imaging a living thing on the scale of micrometers means a lot less is known about *where* and *how* these proteins function. This project seeks a method for learning the former.

Eukaryotic cells (left) vs prokaryotic cells (right)

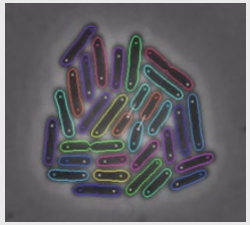


https://www.amazon.com/Prokaryotic-Eukaryotic-Cells-Individual-Microscope/dp/B007VCL0AE

The Machine Learning Approach

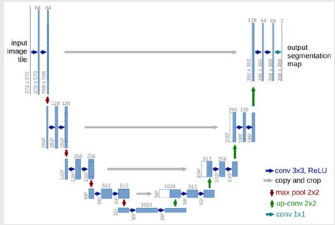
Our code is adapted from an existing project called DeLTA, written by biologists at Boston University. It uses a neural net setup called UNet. UNet uses a "downstream" set of neuron layers to extract computer-parsable information, followed by an upstream set which recreates the shape of the original input in a binary mask.

DeLTA's ability to parse individual cells and the well-documented neural net training code it provides create the basis of our research. However, the end goal will have to be completed with a separate neural net (see 'progress and future goals')



Cell segmentation with DeLTA 2.0
https://gitlab.com/dunloplab/delta

This project aims to train a neural net with the different characteristics of fluorescence from the membrane and the cytoplasm, and have it return what percentage of a protein is in which position.

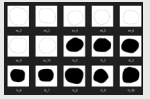


UNet visualization
https://towardsdatascience.com/unet-line-by-line-explanation-9b191c76baf5?gi=2caf539332b9

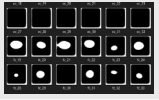
The upwards steps are unique to UNet's structure, making it helpful for DeLTA but not for our project

Progress and Future Goals

Early progress involved testing a UNet with examples more akin to fluorescence than DeLTA's tracking and segmentation code. We attempted to train the neural net to associate empty and full circles with simple black or white arrays.
Since experimentation indicated that the UNet cannot be trained to give an output significantly different than the input, future progress will use a different neural net within the already-adapted training shell.



Example training set (left) and groundtruth (right) – black array for empty circle, white for full. The testing input looks similar to the training input.



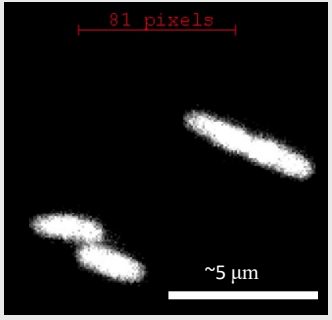
Fluorescence

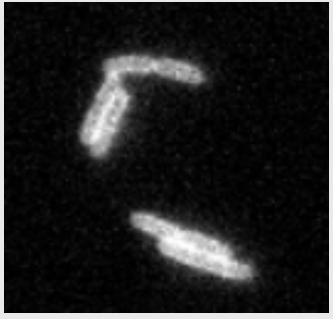
To facilitate protein imaging / location, the genes coding for the proteins of choice have been altered to fluoresce, or glow.

For this project, we aim to categorize protein location into two positions: in the membrane (the cell's 'shell'), and in the cytoplasm (the area within the cell).

Visually we can see that there's a difference – but what if the protein doesn't lie exclusively on the membrane or in the cytoplasm?

We can't accurately determine the distributions by eye, but a neural net could.





100% in cytoplasm (coherent ovals) vs 100% on membrane (accented outlines)

References:
O'Connor, O. M., Alnahhas, R. N., Lugagne, J.-B. & Dunlop, M. J. DeLTA 2.0: A deep learning pipeline for quantifying single-cell spatial and temporal dynamics. Plos Comput Biol 18, e1009797 (2022).

Carnegie Mellon University