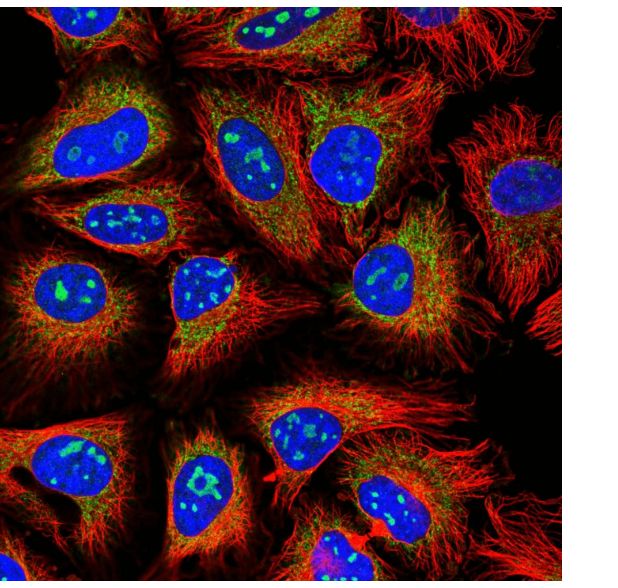


Classifying Subcellular Protein Patterns in Human Cells

Ethan Furstoss, Mohamed El Hhibouri, Chisom Amalunweze

Repository: <https://github.com/chisomama/Kaggle-Protein-Competition>

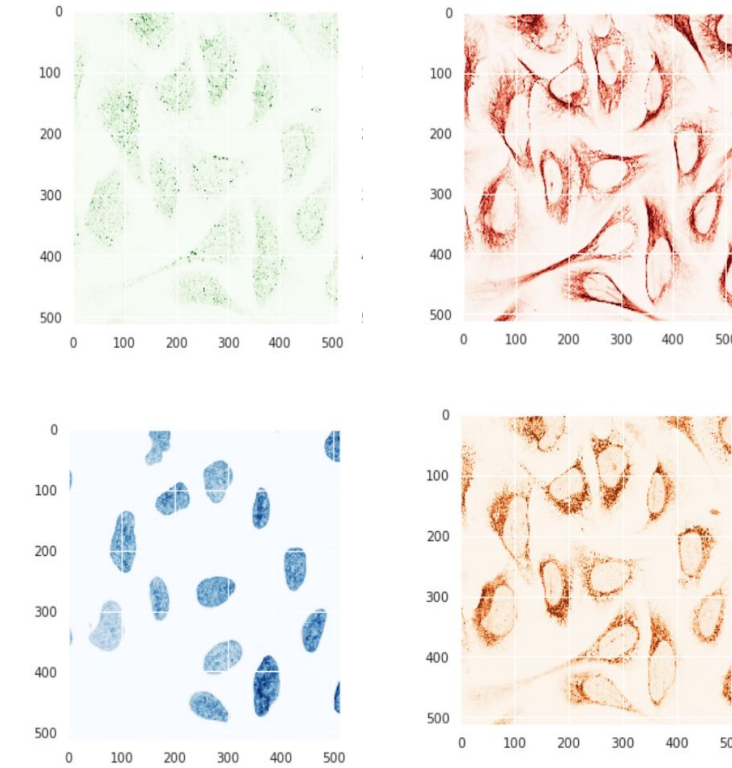


The Problem

The problem at hand is the classification of mixed proteins into their organelle localization labels from microscopic images. There are 28 possible cell types, and each microscopic image can consist of more than one cell classification.

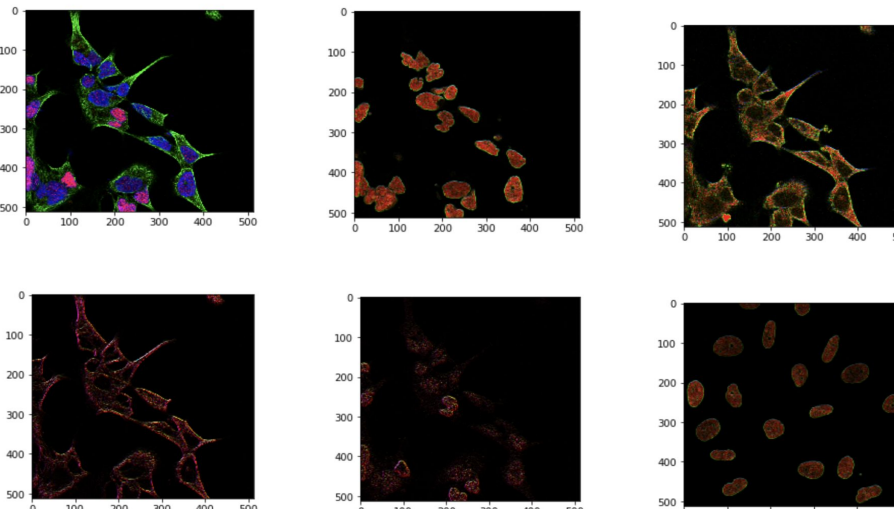
Data Input

4 Different Filters for Each Protein Image.

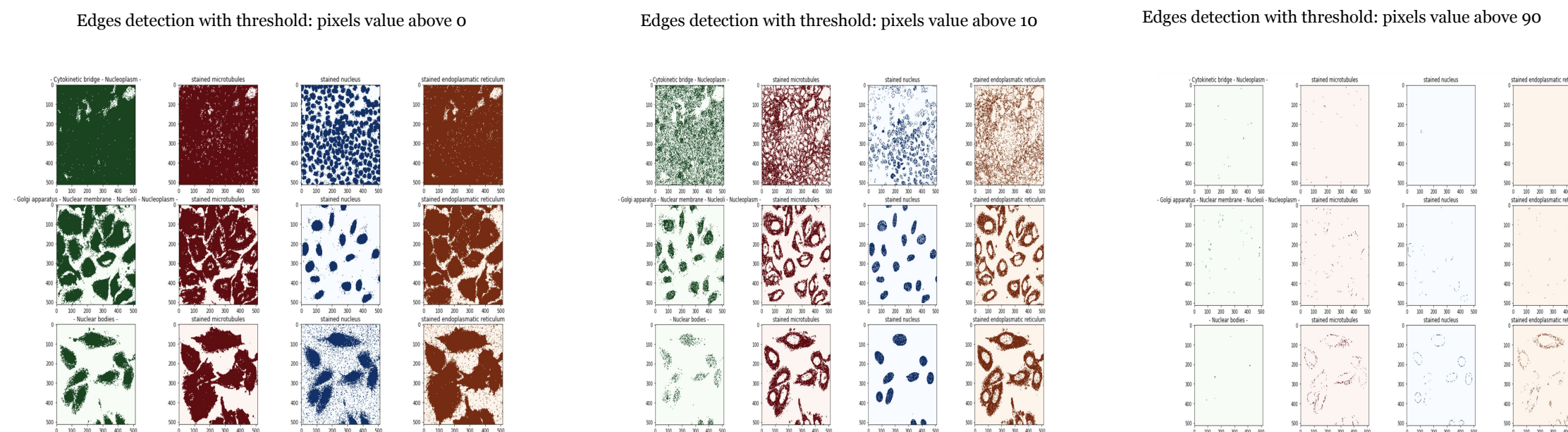


Feature Augmentation

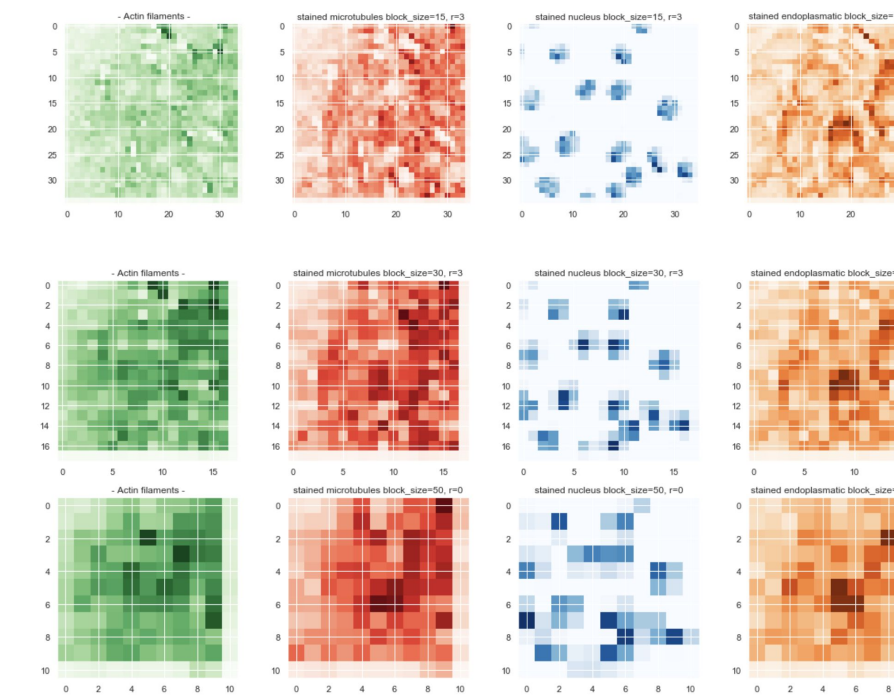
Multiple RGB Channels from Directional Gradient



Normed Gradient Using Sobel Kernel and Threshold



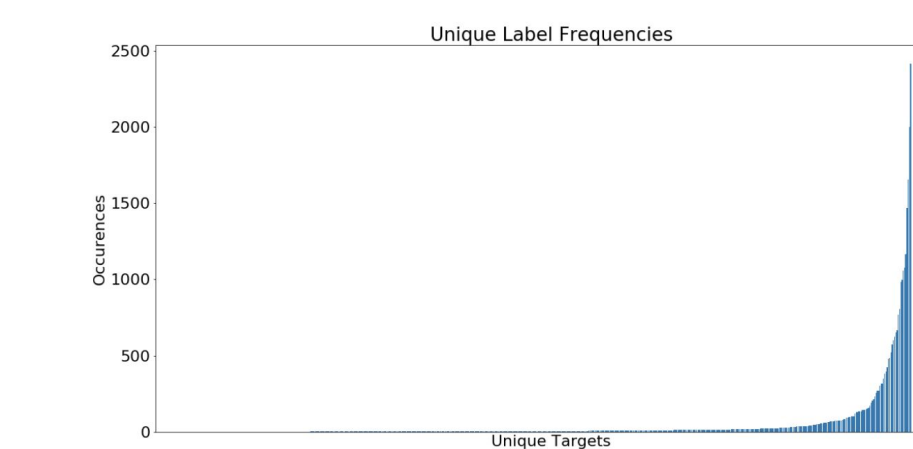
Different size pixels blocks of the mean of the gradient in multiple direction



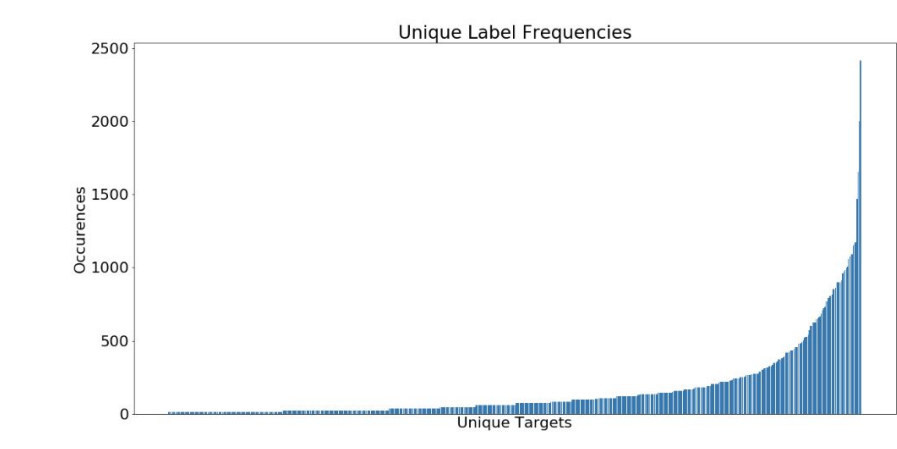
Data Augmentation

To improve the poor distribution of output data and the underrepresentation of minority targets, the following 11 augmentations were performed on the minority portion of the training data: rotations in 3 additional directions, horizontal flip with rotation in 4 directions, and vertical flip with rotations in 4 directions. Increasing our training samples from a size of 31,072 unique entries to 94,960 unique entries.

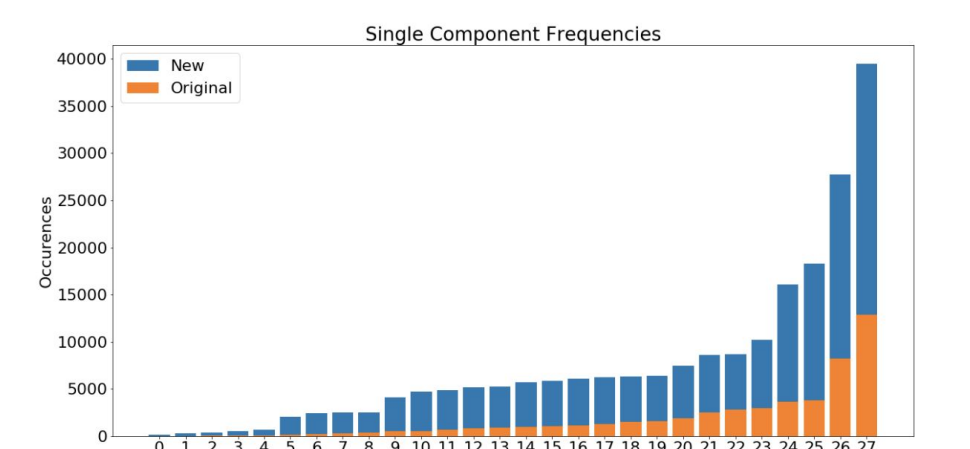
Output Frequency Distribution



Output Frequency Distribution with Augmented Data

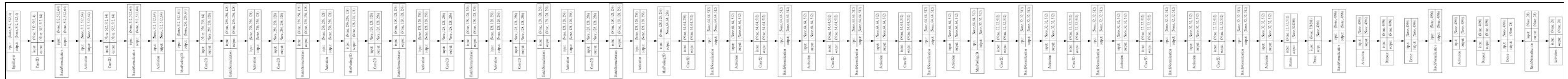


Protein Frequency before/after Augmented Data Addition

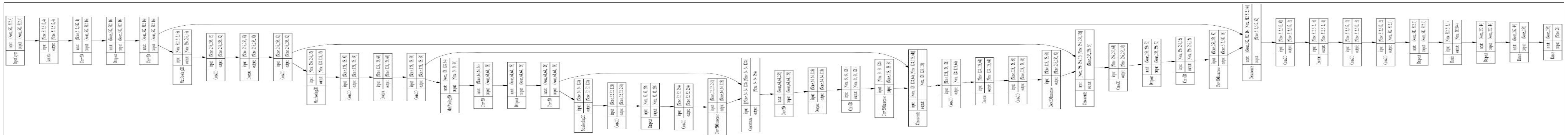


Models

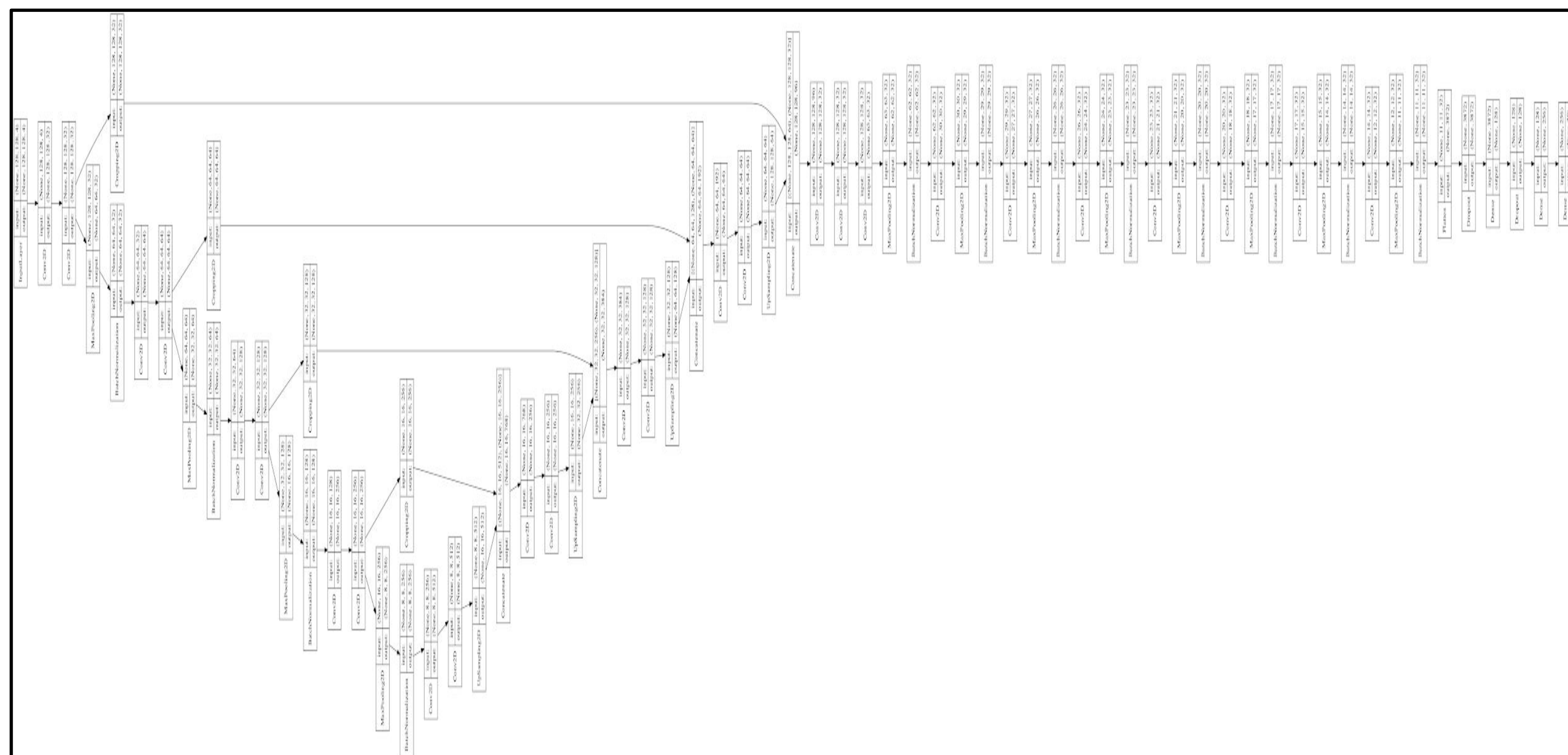
VGG-16



U-Net

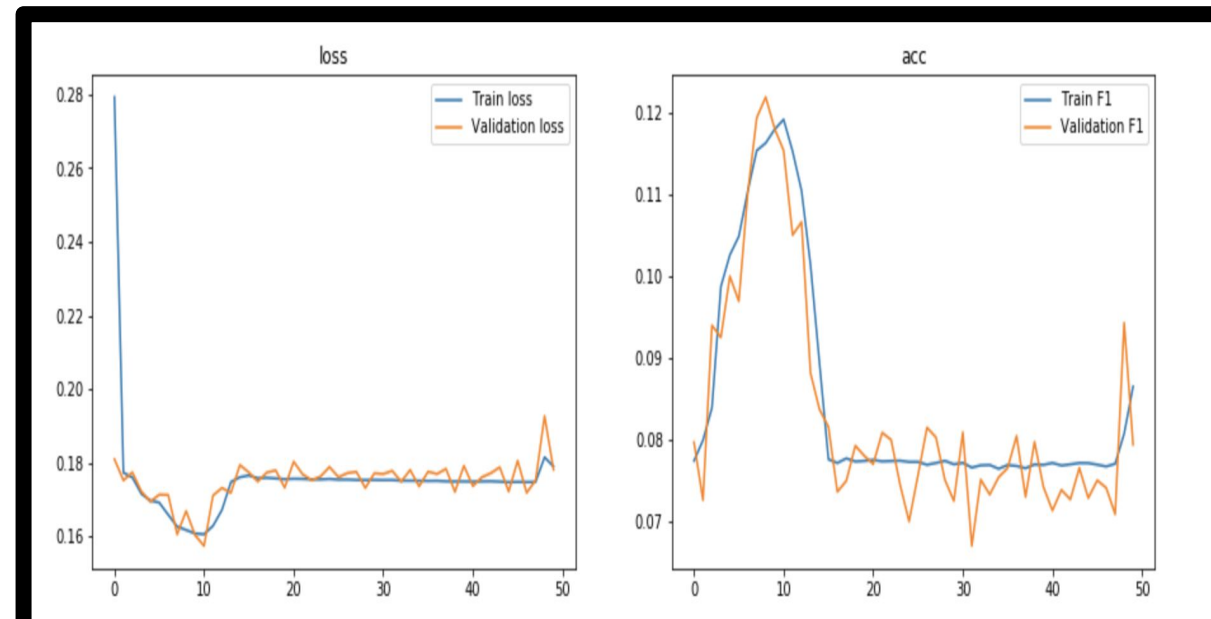


Modified U-Net on Augmented Data (Spoon Net)



Results

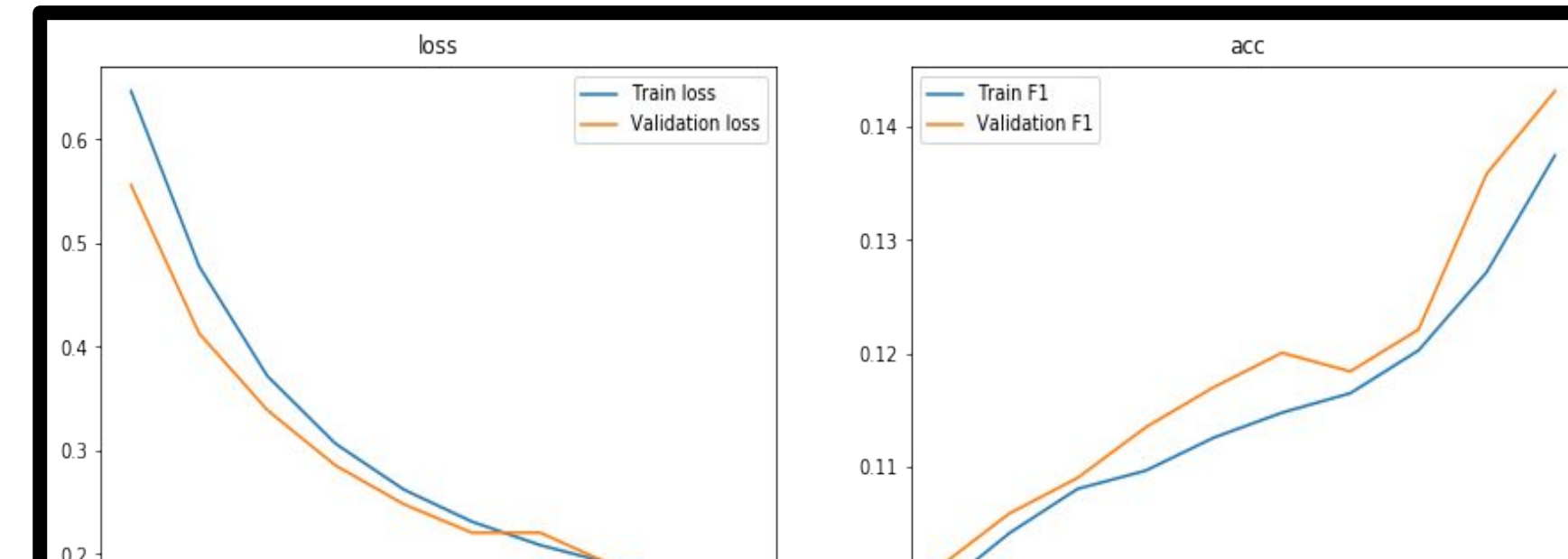
U Net Validation Set Performance



Individual F1-scores for each class:

[0.7313978 0.15753425 0.31086439 0.23152709 0.2533532 0.24495892
0.16166282 0.25434783 0.0173913 0.08529101 0.08501253 0.16692427
0.11715481 0.04180715 0.23024831 0.04878849 0.04854369 0.01529412
0.08224299 0.16030534 0.1 0.41785714 0.10433827 0.34002677
0.1509434 0.56630824 0.09375 0.]
Macro F1-score CV = 0.180659361127381

VGG16 Validation Set Performance



Individual F1-scores for each class:

[0.79559972 0.67768595 0.64370861 0.51369863 0.71666667 0.4990099
0.39506173 0.60037523 0.19354839 0.04347826 0.10526316 0.37313433
0.26315789 0.28301887 0.83076923 0.00243309 0.15151515 0.29411765
0.37735849 0.4 0.10666667 0.58371041 0.40816327 0.59758204
0.32258065 0.62024576 0.16981132 0.]
Macro F1-score CV = 0.3917271801499398

Spoon Net Validation Set Performance

