Video:
https://youtu.be/2XhkMZPAbwc

Classify Subcellular Protein Patterns in Human Cells

Yixiao Song, Dongkai Chen
(syixiao/chedy028)@stanford.edu

Stanford University

Introduction

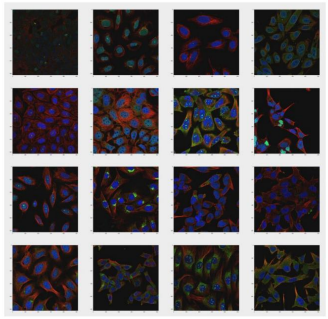
Proteins execute their function at specific times and locations within a cell. Understanding the location of proteins in cell organelles is fundamental to understand biological processes. Researchers previously have been limited to the single protein patterns. In order to understand the intricacy of human cells, we develop deep neural networks to classify the mixed patterns of proteins. The neural network will be able to identify protein types and protein's location from the high throughput images.

Challenge

One of the challenges from the dataset is the significant massive data imbalance, such as "Nucleoplasm" are in majority portion of the dataset, while there are many rare classes, like "Endosomes," "Lysosomes," and "Rods \& rings." It is crucial to augment the rare dataset set or choose a suitable loss function.

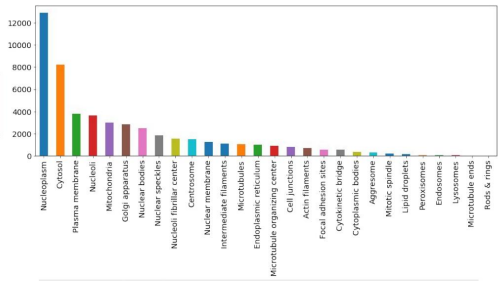
The four-channel images may be different from imagenet, so it is challenging to handle the transfer learning. So we divide the yellow channel data by 3 then combine 1/3 of yellow channel to the RGB images in the milestone stage.

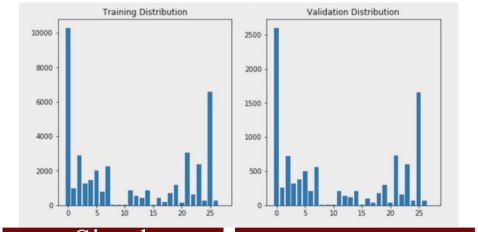
Example of Images(4 channels)



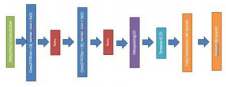
Datasets

The dataset is from Kaggle competition. A scaled set of 512x512 PNG files in train.zip and test.zip is available. There are total 28 different labels in the data set. All image samples are represented by four channels(green, blue, red, yellow).

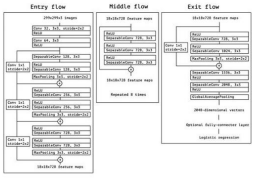




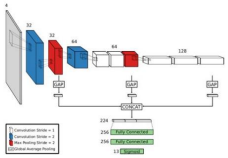
Simple ConvNets



Xception



GapNet



Results&Analysis

The F1 score shows the transfer learning from Xception has significant improvement in classifying protein than the simply convolutional neural network does. GapNet has better performance than the rest of two. The image size is essential for the accuracy and loss.

	Train Acc	Validation Acc	Train Loss	Validation loss
Simply f Convnets(RGB)(299x299x3)	0.122	0.121	0.8	0.78
Simply f Convnets(RGBY)(1024x1024x4)	0.942	0.9	0.22	0.21
Xception(RGB)(299x299x3)	0.99	0.96	0.1	0.02
GapNet(RGBY)(700x700x4)	0.942	0.94	0.0219	0.0218