IST-718 Final Project Proposal

Topic: Human Protein Atlas Image Classification (Classify subcellular protein patterns in human cells)

Team: Srihari Busam, Sharat Sripada, Alexandra Rosser

Introduction

Historically, classification of proteins has been limited to single patterns in one or a few cell types, but in order to fully understand the complexity of the human cell, models must classify mixed patterns across a range of different human cells. Advances in high-resolution microscopy has made it possible to generate images of cells comprising proteins at greater pace and volume than those generated manually.

As such, the need is greater than ever to automate biomedical image analysis and classification.

Problem statement / Objective

Develop models capable of classifying mixed patterns of proteins in microscope images.

Dataset

This is an active kaggle competition by the Human Protein Atlas (https://www.kaggle.com/c/human-protein-atlas-image-classification/overview). The dataset will comprise the following structure:

- test-directory - 512x512 png images (46.8k)
- train-directory - 512x512 png images (124k)
- train.csv - Filenames and labels of images from test-set (31,073 images)
- Sample_subsmission.csv - Test set with filenames (11,703 images)