Course: IST 718

Group: Sharat Sripada, Srihari Busam

Date: 2-27-2021

Human Protein Atlas – Single Cell Classification

# Project Status:

1. What we trying to solve:
   1. We are trying to solve [Kaggle competition](https://www.kaggle.com/c/hpa-single-cell-image-classification) for single cell classification. The input data was images containing multiple proteins. Each image labelled weak (labels only tell what type of proteins are present but do not provide object localization)
2. Observations so-far
   1. The images are huge went 2048 \* 2048 dimensions and in-house GPU is not able to handle the data. We have about 20k training samples.
   2. Resizing the images to 512\*512 seems to be helping with the GPU processing.
   3. We are planning to find a way to separate each cell in the image and run classification on the single cell. We are planning to leverage [HPA-Cell-Segmentation](https://github.com/CellProfiling/HPA-Cell-Segmentation) to extract and use the single images to perform the classification. This may provide 10X of training images assuming there are atheist 10 proteins in each image given.
3. Analysis
   1. Currently working on extracting single cells from the image with multiple cells.
   2. There are about 10k images which seems to have singe protein, we are thinking of using these images for cell extraction and convert weak label to strong label. If we could extract a cell, we can clearly label it as the label of the image itself. We think this we we may get about 100k training images to work with.
   3. Setup a modeling using resnet50 and vgg16 with ImageNet weights to build the model
4. Recommendation
   1. Work to do
      1. Try running the deep learning model on the input images
      2. Extract images into single cell images and run classification modeling using vgg or resent
      3. Image cleanup as necessary
         1. Probably remove samples that are low
         2. Remove tiny or large image extractions
      4. Evaluation of the results

# References:

Kaggle project: https://www.kaggle.com/c/hpa-single-cell-image-classification