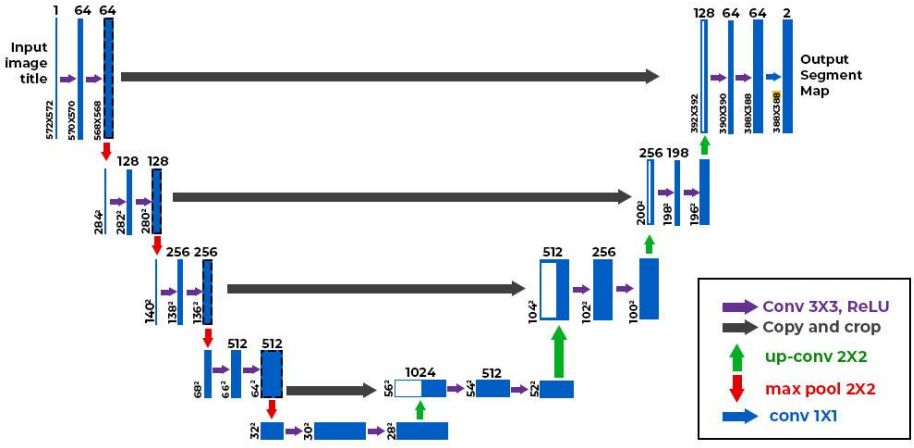
LIVECell Image Segmentation Utilizing the U-Net Architecture

Thomas Yow, Jaime Cardenas, Benjamin Simpson

The LIVECell data set is a large-scale open-source data set for deep learning applications created to address challenges in label-free live-cell segmentation. Given the scope of information we covered in intro to computer vision, we found this data set a welcome challenge to test our understanding of the material. The LIVECell data set consists of over 6000 images of cells from various cell-lines and stages of development with each image representing a unique spatial-arrangement of organisms. The challenge presented by this data set is to be able to identify what part of the image is "cell" and which part is not. In doing so, we can begin to identify boundaries between different organisms. The LIVECell annotations are in COCO Object Detection Format. Each annotation consists of multiple arrays of polygons that outline the individual cells contained in the image. Using these polygons and a RLE technique (run length encoding) we can create a filled binary mask for each image that represents which parts of the image are "cell" with a "1" and which part of the image is "not cell" with a "0".

After heavy consideration, we opted for a U-Net architecture because of its efficiency and relatively low learning curve. The U-Net architecture consists of a series of down samples and up samples that each capture different levels of detail from the original image. During down sampling, the U-Net architecture utilizes max pooling to decrease the spatial resolution and increase the number of channels of the input. This increase in the number of channels allows the model to effectively capture high-level details. During up-sampling, U-Net utilizes up sampling convolutions to decrease the number of channels and increases the spatial resolution, allowing the model to capture more spatial features and produce an image that is the same size as the original input. The resulting output of these operations is a binary image that represents which parts of the image are "cell" and which are "not cell".



***U-Net architecture from Geeks for Geeks*

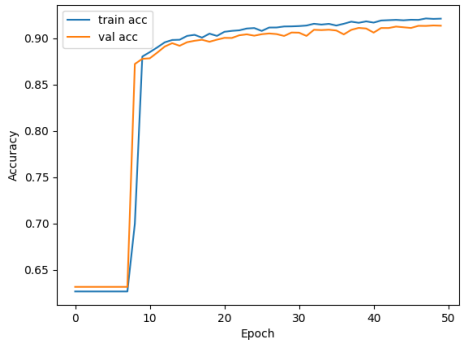
We utilized a binary cross entropy loss function which is of the form

$$BCE = -\frac{1}{N}\sum_{i=1}^{N}\left[y_i \log(p_i) + (1 - y_i)\log(1 - p_i)\right]$$

And applied a sigmoid activation function to categorize the outputs as either one or zero.

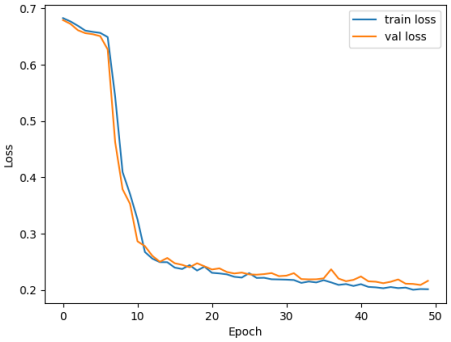
$$\sigma = \frac{1}{1+e^{-x}}$$

Our final U-Net design consisted of three layers of down sampling followed by three layers of up sampling and an initial number of sixteen convolutional filters that doubles with each down sampling layer. We trained our model with 1024 training images each of size 704x520x1. We validated our model with 256 images and tested our model with 512 images all of the same size of the input. These testing, validation, and training sets were formatted as part of the LIVECell data set. We trained our model across 50 epochs, with a batch size of 32, and a learning rate of 0.00005 and found that after around 30 epochs our accuracy began to taper off, achieving a final testing accuracy of 92.19% and loss of 20.43%.



Training and Validation accuracy

Training and validation loss



Overall, we are happy with the performance of our U-Net model. Its reasonably high testing accuracy means that it reliably segments cells from their background environment. Though this model still has room for improvement, it could still be employed by cellular biologists to automate the process of calculating biomass for large amounts of data. We found this project insightful and a good opportunity for applying the techniques we learned in intro to computer vision.