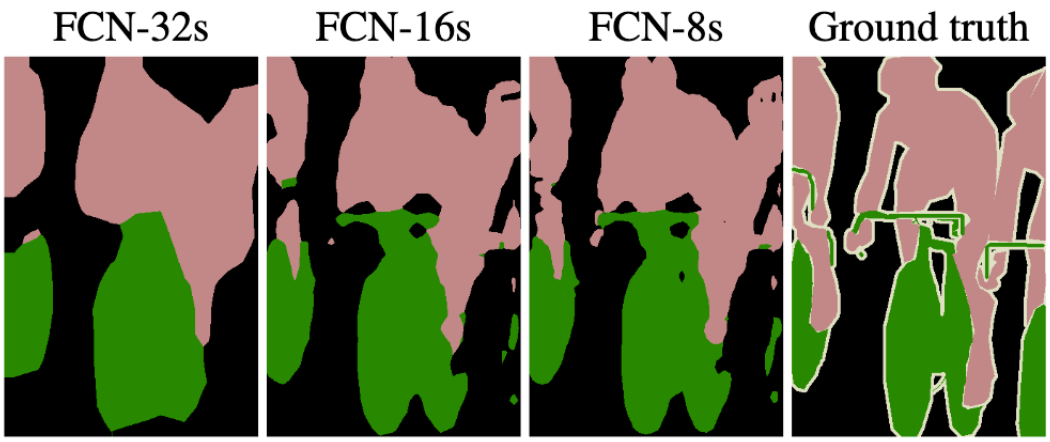
# Project Week1

Zaoyi Chi

Partner: Hannah Caldwell-Meurer



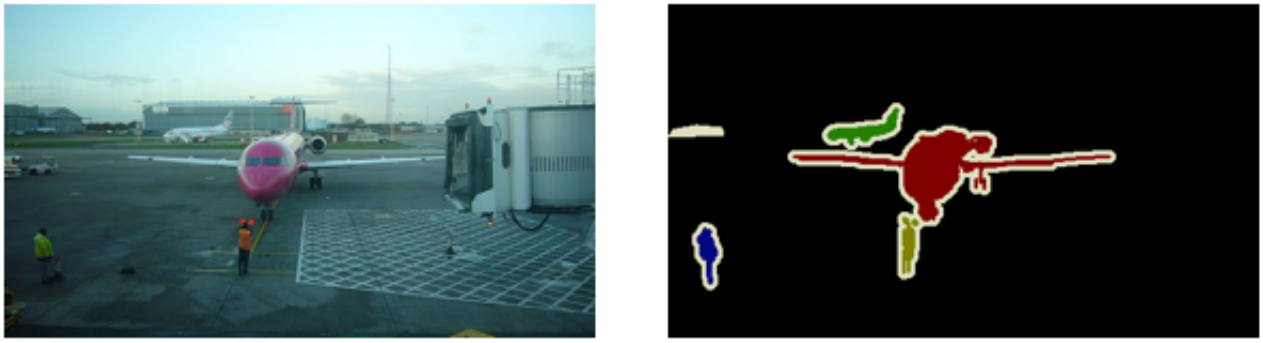

> Figure 1: Long, et al. Fully Convolutional Networks for Semantic Segmentation, Figure 4


> Figure 2: Ren, et al. Faster R-CNN: Towards Real-time Object detection with Region Proposal Networks, Figure 3 Right.

The first figure I want to reproduce is from Fully Convolutional Networks for Semantic Segmentation proposed by Long, et al. The second figure is from Faster R-CNN article proposed by Ren, et al.

In this first figure, the paper is trying to use CNN to do image segmentation. The inputs are the images with the same size; the outputs are the pixels with the labels. The desired learned function is to have a general model that can segment the object from the image. The second figure is a real-time object detection problem. The purpose of that paper is to detect the object on images in real-time and be faster. The inputs are also images in the same size, and output is the frame with the label, based on the prediction probability.

Both two figures are trying to solve real-world problems. They want to segment the image and detect the object given the image, respectively. The further implementation of the first paper could be the medical image, such as the auto segmentation of cancer. The Faster R-CNN paper can be implemented into the real-time video to detect the object or even the base of the auto-car driving system. Both papers could use Pascal VOC dataset; the specific version of this data is VOC 2007. Here is the description of this dataset in detail. For the entire dataset, there are 9,963 annotated images, containing 24,640 annotated objects. And 50 % of them will be split into training and validation set, half to be the test.



> Figure 3: The first traning sample and label

The algorithm or we can say the model is studying the local feature in the shallow layer and the global feature of an object in a deeper layer. The F-CNN is using those learned features to unsample, and obtain the segmentation of the image object. In specific, that is the difference in the performance of FCN-32, FCN-16, and FCN-8. The algorithm I want to code is the model of the learning function. Currently, I am not sure the algorithm of the Faster R-CNN, but I can provide a pseudo-code for F-CNN. In F-CNN, it works similar to CNN, but we unsample the learned features instead of pooling images after a couple of layers. And that is the idea of the F-CNN. Be more specific, the unsampled features make comparison with the label, and backpropagation to improve the CNN learning kernel. The first figure I want to reproduce does not have the number inside; the second figure has the probability of the confidence of that detection, which could be computed based on the Softmax function. In this project, I will have a deeper understanding of the PyTorch package and the structure of the image segmentation and detection model. I believe this model could also be used in the 1-d data or high-dimension image sample.