CS 553 Project II Proposal
Arturo Gourentchik, Jiovanny Alejos, Alan Vu

U-Net Convolutional Network for Image Segmentation

We will be exploring the U-Net architecture introduced in the research paper *U-Net: Convolutional Networks for Biomedical Image Segmentation*- Ronneberger et al. (2015). Previously, the CNN architecture was mainly used for classification tasks and was limited when it came to finding the position of a subject or in identifying different instances of a subject in an image, which are known as localization and segmentation tasks respectively. The motivation for the U-Net architecture was to extend the use of the CNN to accomplish these tasks.

In the paper the authors acknowledge the strong performance of traditional CNN networks on visual recognition tasks, with the strongest results achieved on image classification tasks. However they identify some issues in that the current models require very large labeled training samples (which is not feasible on many biomedical tasks) and that they had poor performance on image segmentation. They introduce a U-shaped architecture that has a contracting and expanding path that utilize convolutional and pooling layers seen in CNNs. In the expanding path, feature maps from corresponding downsampling steps are used to help determine where the activations showing the subject's presence were found at each step. This architecture was found to give very precise image segmentation results while requiring very few training samples to train. Through this project, we'll explore methods of using convolution layers to downsample an image into its most pertinent information, and upsampling using concatenations from previous downsampling steps to generate a mask showing where and which pixels belong to each individual instance of a class.

For our project, we plan to use a subset of the Common Objects in Context (COCO) dataset. The entire dataset includes over 300,000 images and 80 different object categories for classifying and segmenting objects in images. Our goal will be to train a U-Net model on a small subset of these images to show that this architecture can be applied to other segmentation tasks. We will also train a CNN model on the same subset and compare these two models to show that U-Net outperforms the basic CNN architecture on these tasks. We plan to use COCO's Cocoeval from their API and compare the average precision (AP) metric between the two models to evaluate their performances.

For role assignments, Alan will handle the data preprocessing and subsetting to get it ready to be used for model inputs. Arturo will work on implementing the model architecture with pytorch, and train and adjust it to get reasonable preliminary results. Jiovanny will work on evaluating and optimizing the model using different optimization techniques to fine tune it until it gets significantly lower test losses and better accuracy. We will all contribute equally in writing the final report by taking on individual sections.

References:

https://github.com/cocodataset/cocoapi
https://cocodataset.org/#home