Intro Deep Learning: Homework 4

Daniel Novikov

March 27th, 2022

Problem 1

Object detection and Object Segmentation are two very important tasks in computer vision. There are many real world applications of these tasks. Take one of these tasks and explain how you would implement it. For example, Image segmentation is used in medical imaging.

Answer:

Given a dataset of segmented images that were manually segmented by experts, as well as their non-segmented original versions, I would train an autoencoder with a small bottleneck to force the network to learn only the most significant features of the image prior to segmentation. The autoencoder will train by minimizing reconstruction error on the original (non-segmented) images. After the autoencoder I will add a fully connected layer to classify each pixel in the autoencoded image. My loss function will compare the pixel class predictions to those given by the manually-segmented images. Cross Entropy loss will work.

Architecture:

x $->$ Autoencoder $->$ fully connected classifier $->$ Loss

Problem 2

how does YOLO (You only Look once) works? What is the the difference between YOLO and Faster-RCNN?

Answer:

YOLO and Faster-RCNN are object detection algorithms. The major difference between them is that YOLO is not a recurrent network. It finds the bounding boxes and classifies the objects simultaneously, in a single forward pass through a convolutional neural network. Faster-RCNN is a recurrent convolutional network that must make many recurrent passes to iteratively refine its bounding

box predictions, and then once all the objects have been detected, then they are classified in a seperate stage.

YOLO works by segmenting the image into subregions and finding candidate objects and their classes. The deeper convolutional layers see more surrounding subregions and learn context, which will eventually allow it to decide the final bounding boxes and their class.

As it is not a recurrent network, YOLO is very fast, and usable for realtime applications requiring object detection at 45 frames per second.