Deep Learning Course 2018 *Vanditha Chandrashekar Rao*
Autonomous Intelligent Systems
University Freiburg *Date: 04.12.2018*

Deep Learning Lab Course 2018 - Exercise 3

Introduction

Semantic segmentation is a natural step in the progression from coarse to fine inference. A general semantic segmentation architecture can be broadly thought of as an **encoder** network followed by a **decoder** network:

- The **encoder** is usually is a pre-trained classification network followed by a decoder network.
- The task of the **decoder** is to semantically project the discriminative features (lower resolution) learnt by the encoder onto the pixel space (higher resolution i.e. upsample) to get a dense classification.

The upsampling is done using transposed convolution and Exponential Linear Unit (ELU) activation function. The upsampling process is further refined by adding features from coarser but higher resolution feature maps from lower layers. Slowly upsampling or stage upsampling adds 'skip connections' from encoder layers to the decoder by fusing features which pass through less downsampling operations. Skip connection is introduced after convolution block to enable the subsequent block to extract more abstract, class-salient features from the previously pooled features.

In this exercise we do Semantic Segmentation with Fully Convolution network.

Implementation

In this exercise we have to implement a simple single stage and the multi-stage decoder module to upsample the encoder features back to the image resolution for semantic segmentation..

Here, we implement one, two and three refinement block which are configuration 2,3 and 4 respectively. Each configuration include a new refinement module with upsampling rate of 2× and its corresponding skip-connections. Configuration 2 has one skip connection (DB4_skip_connection), Configuration 3 has two skip connections (DB4_skip_connection, DB3_skip_connection), and Configuration 4 has three skip connections (DB4_skip_connection, DB3_skip_connection, DB2_skip_connection).

Intersection Over Union is also computed to measure the accuracy of an object detector on a particular dataset. Intersection over Union is an evaluation metric

Dataset

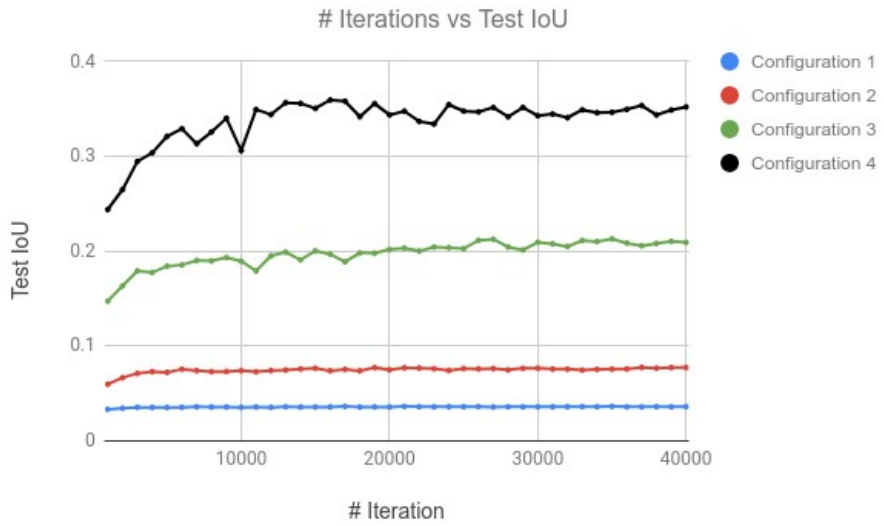
The used dataset is the CamVid semantic segmentation dataset. The dataset is constituted by 468 training images and 233 testing images. The dataset is divided in 11 classes: Sky(0), Building(1), Pole(2), Road(3), Sidewalk(4), Tree(5), Sign(6), Fence(7), Car(8), Pedestrian(9), Cyclist(10).

Results

Intersection Over Union is computed and the table below gives the maximum IoU value for each configurations:

Configuration	Maximum IoU
1	0.03636062814
2	0.07744444213
3	0.2322438777
4	0.3593828471

A plot of Intersection Over Union(IoU) vs epochs for each decoder configuration is as shown below:



It can be observed that configuration 1 with no skip connection has the lowest testIoU whereas configuration 4 with 3 skip connection has highest testIoU.

Reference

1. http://ais.informatik.uni-freiburg.de/teaching/ws18/deep_learning_lab/Segmentation.pdf
2. https://medium.com/nanonets/how-to-do-image-segmentation-using-deep-learning-c673cc5862ef