Deep Learning Lab WSE 2018 Exercise 3

Assignment: Implementing a decoder module for FCNs

Objective

The goal of this exercise is to understand the main concepts and components of a decoder module of Fully Convolutional Networks (FCNs) for semantic segmentation.

Implementation

Configuration 1: Implement a simple decoder module to up sample the final features obtained from the encoder. Our FCN encoder consists of four convolution layers with stride 2 which will produce a feature map whose spatial resolution is 16 times smaller than the input image. The first task will be to implement a module to up sample the features back to the original spatial resolution with a single module. To achieve this, transposed convolution with stride similar to the up-sampling rate must be used.

Refinement block to restructure Configuration 1

It consists of the up-sampling module you implemented previously, which is composed of a transposed convolution and an activation function (ELU). After that you must check if the up sampled features and skip connection features are of the same spatial resolution. If not crop the largest one and concatenate both before sending to a convolution layer to fuse the feature maps.

Configuration 2, 3 ,4: Construction of multistage decoder with skip connections

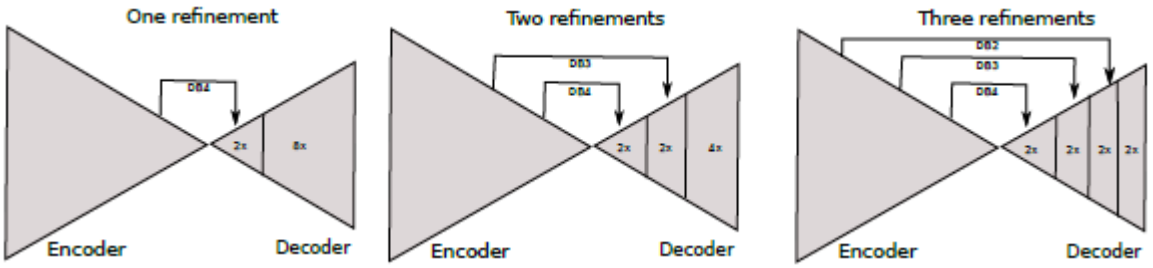

Figure 1. Configurations 2, 3, 4 to be implemented. Each configuration includes a new refinement module with up sampling rate 2x and its corresponding skip-connection.

Results

The following parameters were obtained with the iteration of the parameters listed above.

Variable conditions: ***Number of skip connections and convolutions***

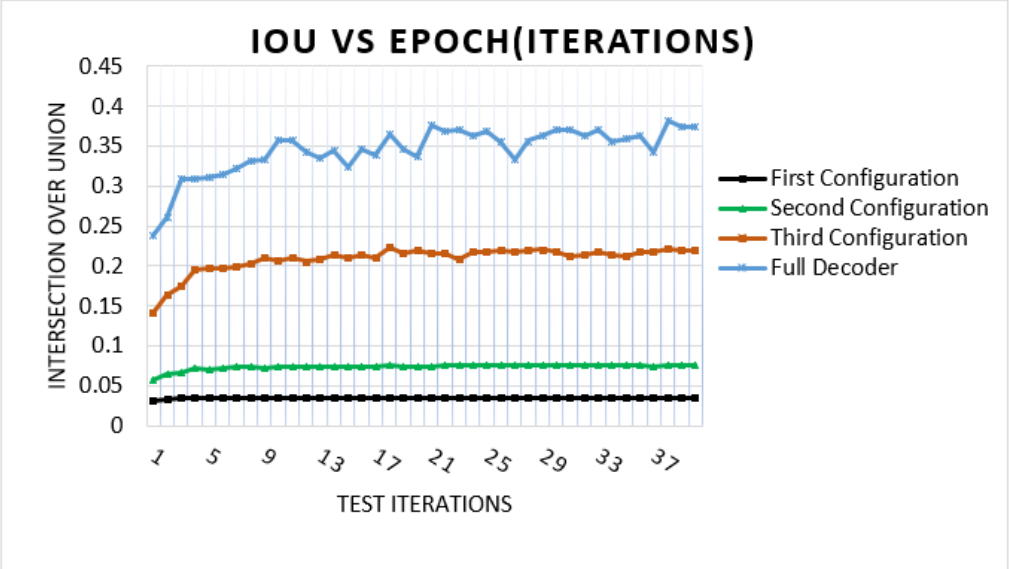

Figure 2. Plot of Intersection Over Union (IoU) vs Epochs (test iterations) for each decoder configuration

Configuration	Maximum IoU
First Configuration	0.035
Second Configuration	0.077
Third Configuration	0.222
Full Decoder	0.382

Table 1. Maximum IoU value for each configuration

Findings:

Code was implemented as an alternate solution following the structure of a "U-NET" from the paper "U-Net: Convolutional Networks for Biomedical Image Segmentation" by Olaf Ronneberger, Philipp Fischer, and Thomas Brox. Results were the same from asked solution and the paper mentioned above structure.

Conclusion:

We can see a big improvement in every stage of the decoder implemented in each configuration. In each stage in comparison with the next we can see an increment of at least 40% in IoU improvement. Full decoder gives the best result obviously since it has greater quantity of refinements, this means it losses less information thanks to the skip connections.

Feedback:

Really interesting and useful topic. Code implementation approximately 10 hours.
Running time and debugging approximately 15 hours.
Implementation/Solution of the code around 25 hours.