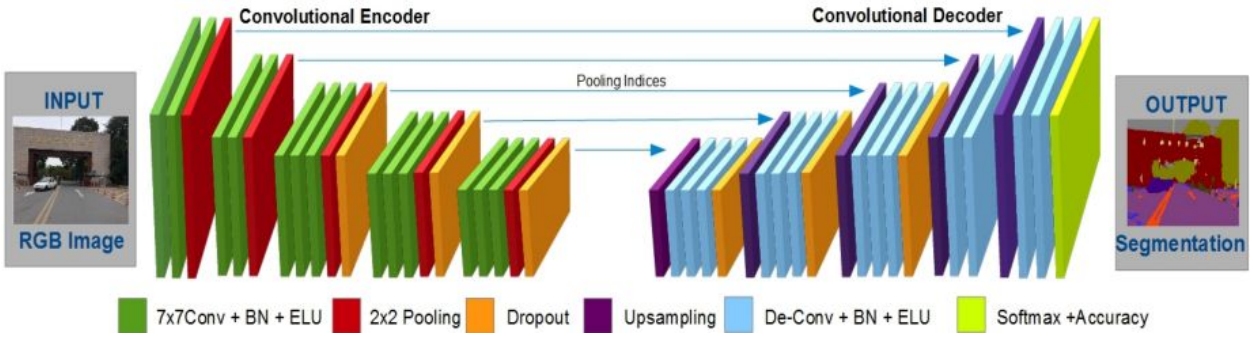
Deep Learning Lab 2018
Exercise 3 : Encoder Decoder for Semantic Segmentation

Name: Jatin Dhawan
Matriculation No. : 452322

Introduction: The aim of this exercise was to implement a decoder module of Fully Connected Network (FCNs) for semantic segmentation. The encoder usually takes the input and outputs a downsampled feature vector whereas the decoder again a network, takes the feature vector from encoder as an input and outputs a segmented image of same size as that of input image. Below is a figure which shows whole structure of encoder and decoder. Skip connections are used to provide necessary details in order to reconstruct accurate shapes for segmentation boundaries.

Figure 1: Yasrab et al. 2017



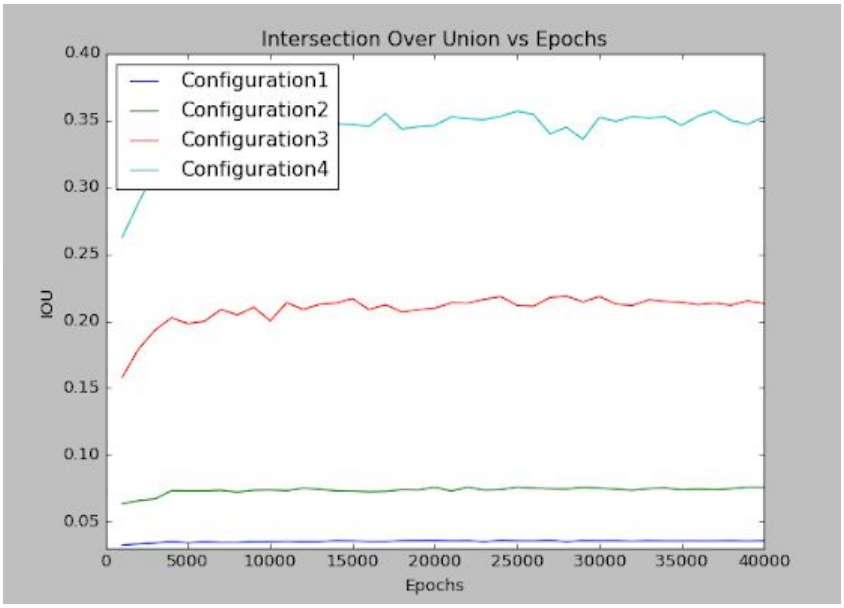
Implementation: In this exercise, we were given with the FCN encoder module which consists of 4 convolution blocks. The downsampled feature map, the output of encoder is 16 times smaller than the input image. A decoder module was implemented to upsample the features using transpose convolutions. Skip connections are also used to provide refinement. Crop function is also used in case the shape of skip connections and upsample layer is not same. After doing the transpose convolution, convolution is done to achieve refinement, so the number of features are equal to number of targets at the end.

4 configurations, with different stride and skip connections, are done to examine different accuracies.

- Configuration1: In configuration 1 Single upsampling layer is produced using transposed convolution with stride 16.

- Configuration2: An upsampling layer with a stride of 2 is added with a skip connection and then again transpose convolution is done with stride of 8 to achieve the size of the input image.
- Configuration3: An upsampling layer with a stride of 2 after transpose convolution is added with a last skip connection from encoder. After that transpose convolution and skip connection after 3rd convolution block from encoder is added again. Finally a stride of 4 with transposed convolution is done to achieve the shape of an input image.
- Configuration4: Configuration 3 is done again which also includes the 2nd skip connection from encoder and a final transpose convolution with a stride of 2.

Results: Figure below shows the different IOUs with epochs. Configuration 4 shows the maximum value because of many skip connections which helps us to provide more refinement.



Maximum IOUs vs Iterations

Config1	0.0358540568460363
Config2	0.0757008859713841
Config3	0.2187961498283557
Config4	0.3570832742927067