Deep Learning Lab
Computer Vision Track
Assignment 3 Report

Kinan Alzouabi
WS 2018/2019

Task

I have trained an Encoder-Decoder network with 4 different configurations of the decoder network. The encoder-decoder network is used for semantic segmentation. Moreover, we are studying the impact of the number of upsamples performed and the use of skip connections on the performance of the network. The configurations are described in the following table.

Table 1- configurations to be implemented

Number of upsamples	Upsampling rates	Configuration
One	16x	1
Two	2x -> 8x	2
Three	2x -> 2x -> 4x	3
Four	2x -> 2x -> 2x -> 2x	4

Network Architecture

Configuration 1

There is no refinement block and we directly upsample the feature map from the encoder to the size of the image. So we have an upsampling layer that give an output feature map of size 120, upsampling rate of 16 and kernel size of 3x3. This layer is followed by a Conv layer with output feature map of a size equal to the Number of classes, upsampling rate of 1, and kernel size of 1x1.

Configuration 2

There is one refinement block with one skip connection. This configuration is better described in table 2, and in figure 1.

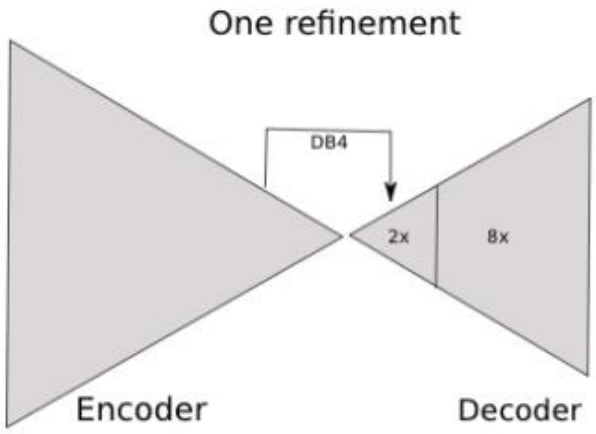

Figure 1- Configuration 2

Table 2- Configuration 2

Layer number	Output feature maps	Upsampling rate	Kernel size
Upsample 1	256	2	3x3
Conv 1	256	1	1x1
Upsample 2	120	8	3x3
Conv 2	Number of classes	1	1x1

Configuration 3

There is two refinement blocks with two skip connections. This configuration is better described in table 3, and in figure 2.

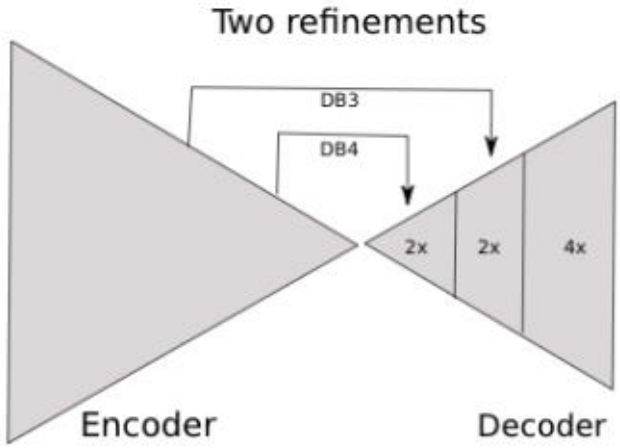

Figure 2- Configuration 3

Table 3- Configuration 3

Layer number	Output feature maps	Upsampling rate	Kernel size
Upsample 1	256	2	3x3
Conv 1	256	1	1x1
Upsample 2	160	2	3x3
Conv 2	160	1	1x1
Upsample 3	120	4	3x3
Conv 3	Number of classes	1	1x1

Configuration 4

There is two refinement blocks with two skip connections. This configuration is better described in table 4, and in figure 3.

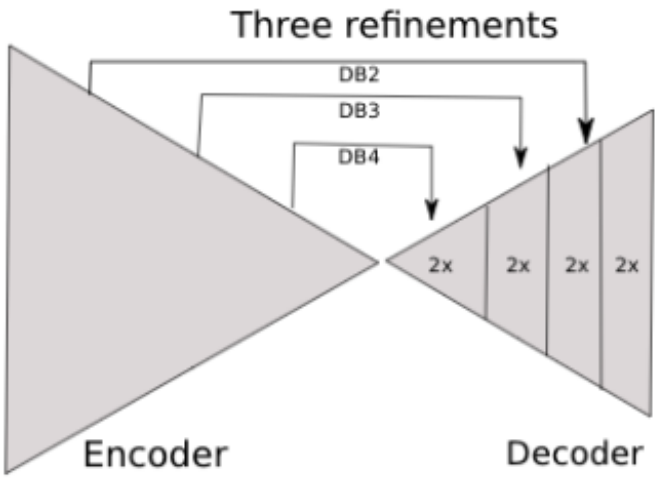

Figure 3- Configuration 4

Table 4- Configuration 4

Layer number	Output feature maps	Upsampling rate	Kernel size
Upsample 1	256	2	3x3
Conv 1	256	1	1x1
Upsample 2	160	2	3x3
Conv 2	160	1	1x1
Upsample 3	96	2	3x3
Conv 3	96	1	1x1
Upsample 4	120	2	3x3
Conv 4	Number of classes	1	1x1

Results

Figure 4 is a plot of Intersection over Union (IoU) vs epochs for each decoder configuration. Moreover, table 5 is a table with maximum IoU value for each configuration.

Table 5- Maximum IoU value for each configuration

Configuration	maximum IoU value
1	0.0364831800209
2	0.0715420922407
3	0.162767471557
4	0.196933913661

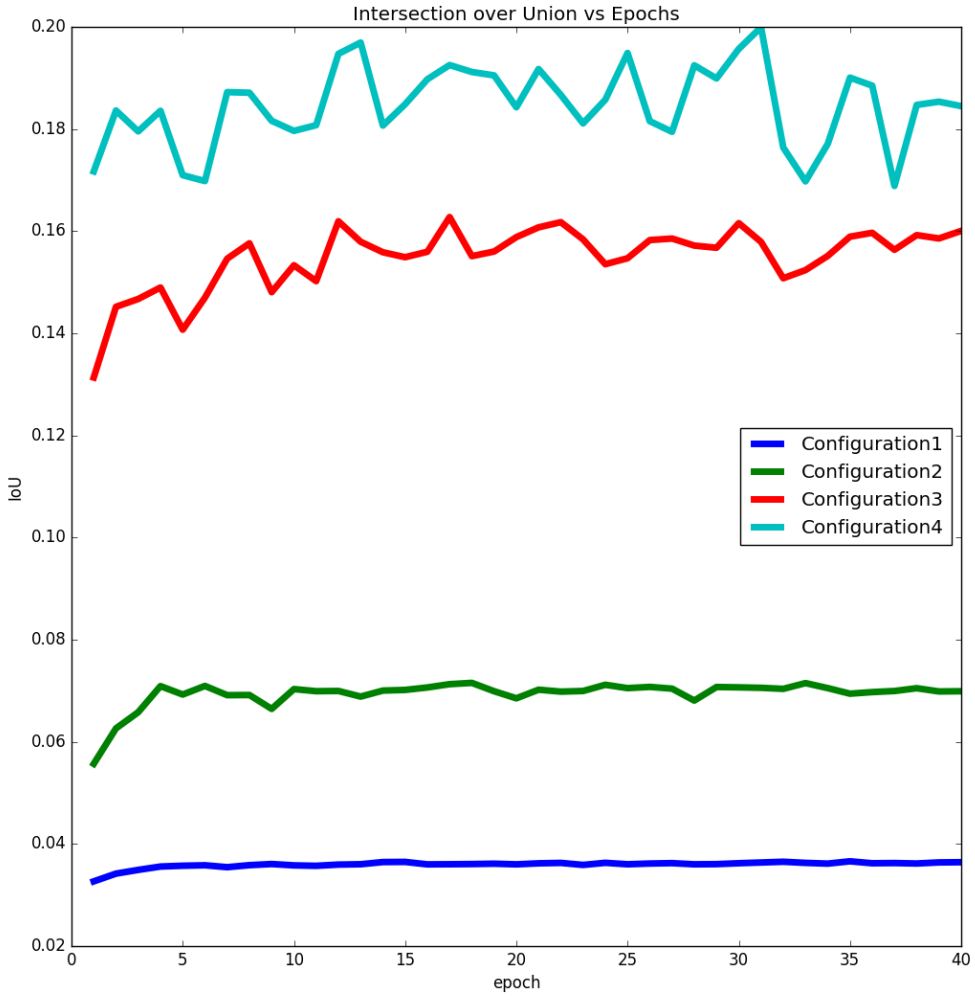

Figure 4- Intersection over Union (IoU) vs epochs for each decoder configuration