UNIVERSITY OF TEHRAN

COLLEGE OF ENGINEERING

DEPARTMENT OF ELECTRICAL AND COMPUTER ENGINEERING

NEURAL NETWORK & DEEP LEARNING

SEGMENTATION

SIAVASH SHAMS

MOHAMMAD HEYDARI

SCHOOL OF ELECTRICAL AND COMPUTER ENGINEERING

UNIVERSITY OF TEHRAN

Apr. 2022

1 CONTENTS

SEGMENTATION

In this part we intend to implement one of the most common application of neural network which is image segmentation.

Image segmentation is a method in which a digital image is broken down into various subgroups called Image segments which helps in reducing the complexity of the image to make further processing or analysis of the image simpler. Segmentation in easy words is assigning labels to pixels. All picture elements or pixels belonging to the same category have a common label assigned to them. For example: Let's take a problem where the picture has to be provided as input for object detection. Rather than processing the whole image, the detector can be inputted with a region selected by a segmentation algorithm. This will prevent the detector from processing the whole image thereby reducing inference time.

Further, we are going to analyse two of the most important models in case of image segmentation.

DEEPLAB

The first model I have implemented called DEEPLAB which is a state-of-the-art semantic segmentation model having encoder-decoder architecture. The encoder consisting of pretrained CNN model is used to get encoded feature maps of the input image, and the decoder reconstructs output, from the essential information extracted by encoder, using up sampling. Further you can find the architecture of DEEPLAB V3:

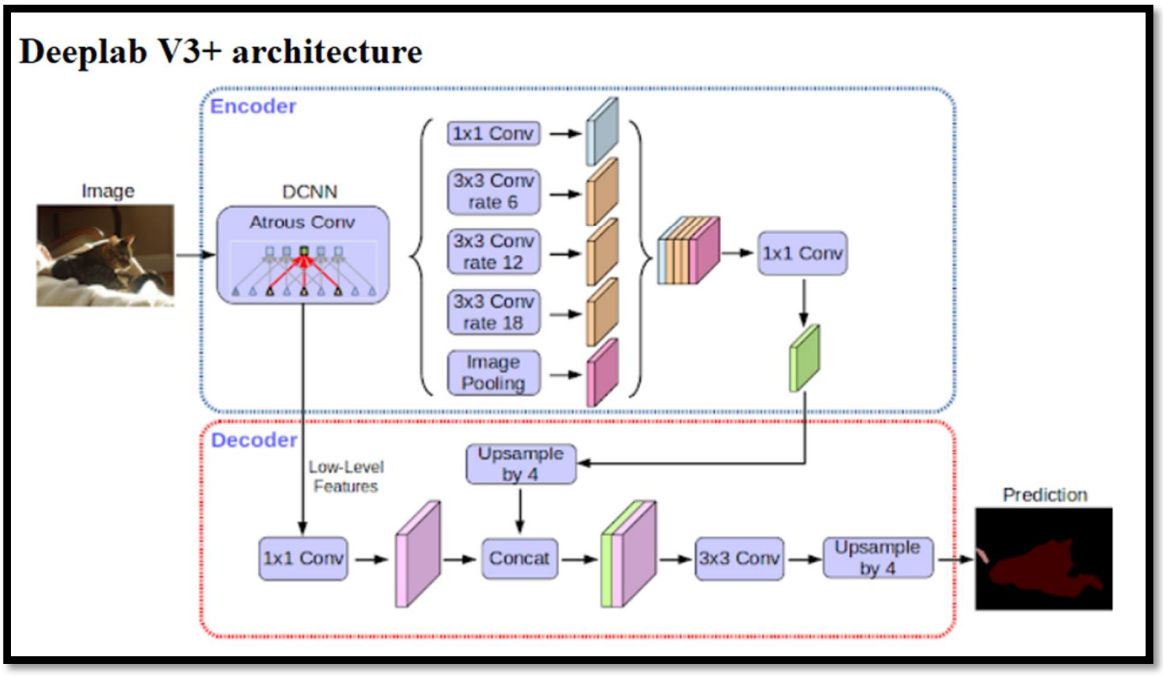

Figure21. Architecture *of Deeplab V3*

FCN

The second model I have implemented called FCN.

One of the ways to do semantic segmentation is to use a Fully Convolutional Network (FCN) i.e., we stack a bunch of convolutional layers in an encoder-decoder fashion. The encoder down samples the image using strided convolution giving a compressed feature representation of the image, and the decoder up-samples the image using methods like transpose convolution to give the segmented output.

Further you can find the architecture of FCN:

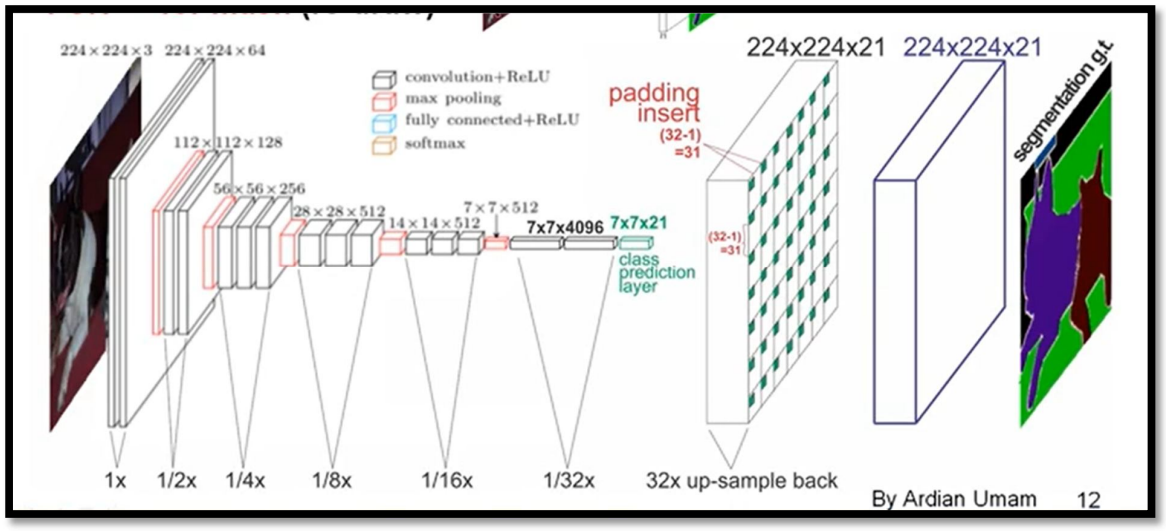

Figure22. Architecture *of FCN*

QUICK REVIEW ON PERFORMANCE & ARCHITECTURE OF BOTH MODELS

DeepLab V3 (2017):

First introduce the following four methods to solve multi-scale problems:

- Image pyramid: Scale the image to different scales, share models and parameters, input different sizes, and then merge and output;
- Encoder-decoder: Encoding and decoding network, the decoding part gradually merges the encoded information;

- Context module: add additional modules at the end, such as CRF
- Spatial pyramid pooling: Add spatial pyramid pooling to the last layer

Now we express some of important notes about the architecture:

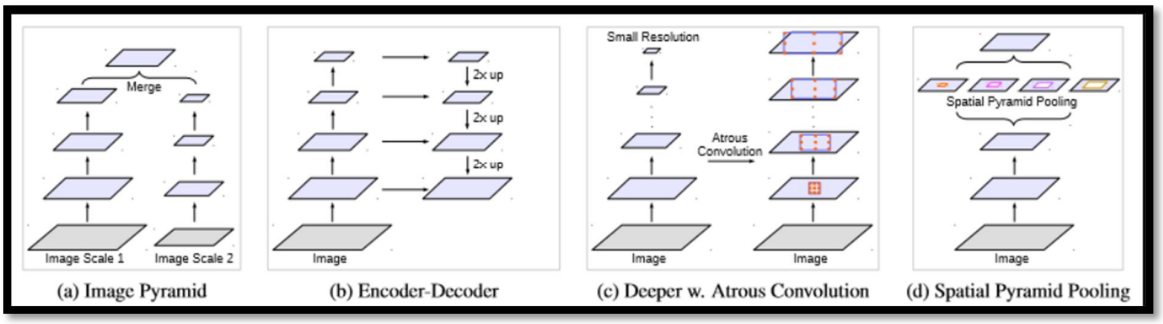

Figure23. rate value representation

As the rate value increases, the effective weight decreases. When the rate value is large, the hole convolution is equivalent to a 1x1 convolution, because only the middle weight of the convolution kernel is valid, so that global information cannot be obtained. Therefore, certain modifications have been made on the basis of V2: (a) ASPP includes 1x1 convolution and three hole convolutions with rates of (6, 12, 18); (b) using global pooling on the last feature , And connect a 256-channel 1x1 convolution and BN, and then bilinearly up-sampling. After concat, pass a 256-channel 1x1 convolution and BN layer.

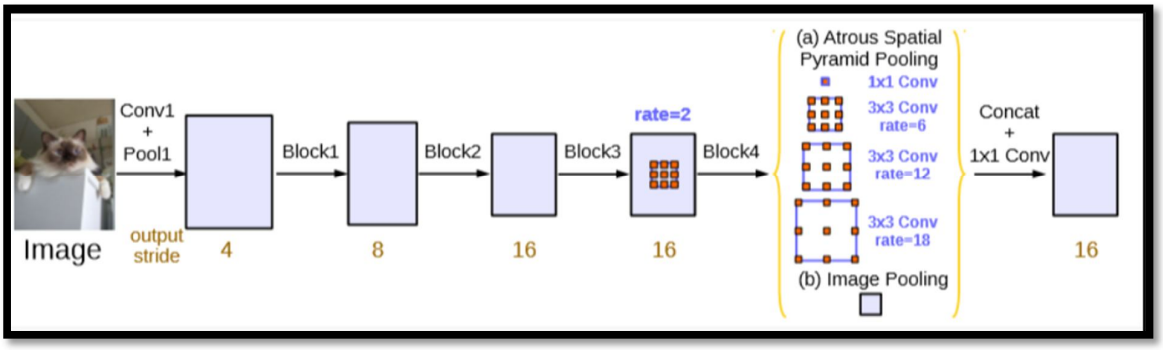

Figure24. Atrous Spatial Pyramid Pooling Module

Brief summary:

- Using ResNet's 4 blocks to extract features

- The brown numbers at the bottom of the figure represent the ratio of the original input image spatial resolution to the output resolution
- The original text says that if output_stride=8, the rate should be doubled, which means that if the ASPP structure is used one block in advance, the rate value should be changed accordingly (Original: Note that the rates are doubled when output stride = 8.)

FCN (2014):

FCN is the basis of semantic segmentation, and many subsequent segmentation methods are developed on the basis of FCN.

Brief summary:

FCN turns the fully connected layer of some classification networks (such as VGG) into a fully convolutional layer, obtains a low-resolution feature map and then uses deconvolution to up sample to the original image size

Pooling can obtain advanced features and also cause resolution reduction and loss of spatial information, so FCN combines features with different roughness (mentioned in the notes) to refine the segmentation results

RESULTS OF IMPLEMENTATION

I have used three separated images as of our test-input, one of them from the one-class category, another one from two-classes category and the final from the multi-class category.

First of all, I have used **transform command** and also **preprocess command** to do some the important preprocessing steps such as **suitable-scaling** for the input image , **normalizing** and **resizing the input images**.

Afterwards, I call the both models to import the correspond weights for doing appropriately segmentation task and at the final I have passed the input images to the models and get the results **which are abbreviated below for both Deeplab and FCN approaches:**

DeepLab V3:

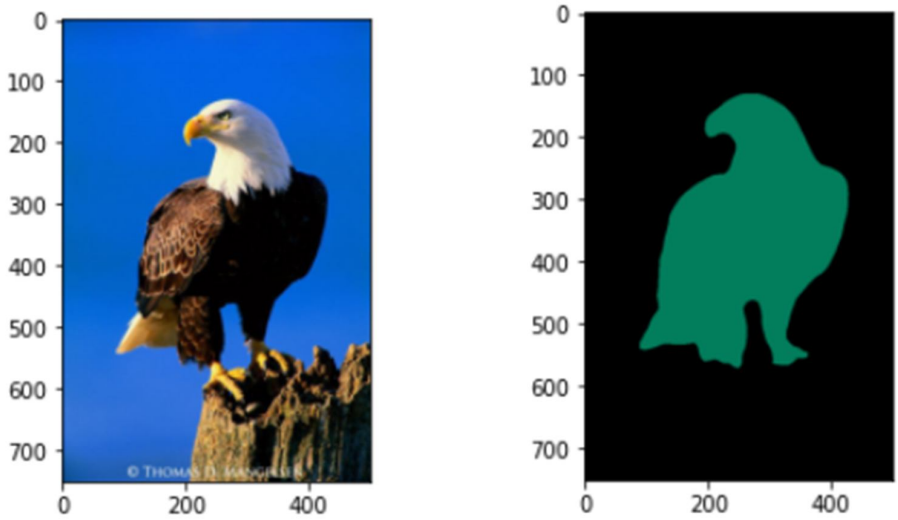

Figure25. One-Class Segmentation

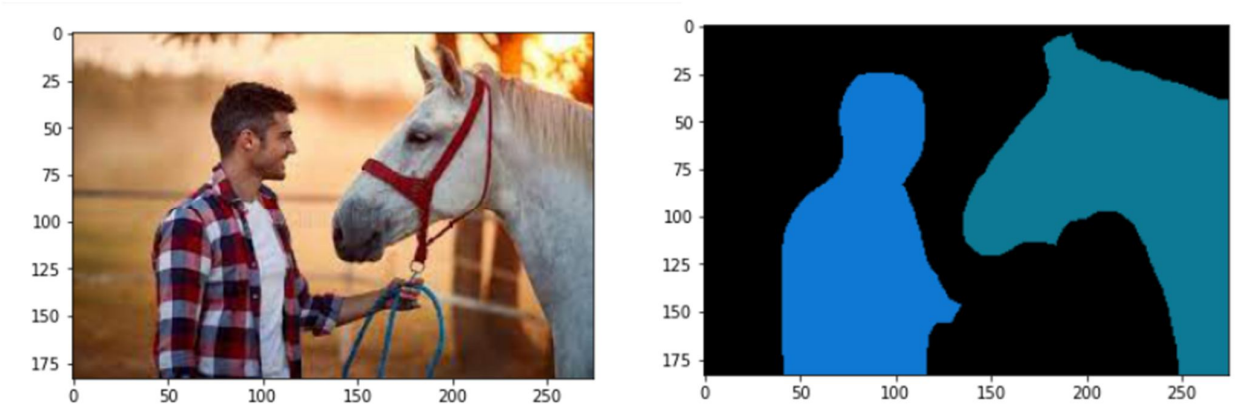

Figure 26. Two-Classes Segmentation

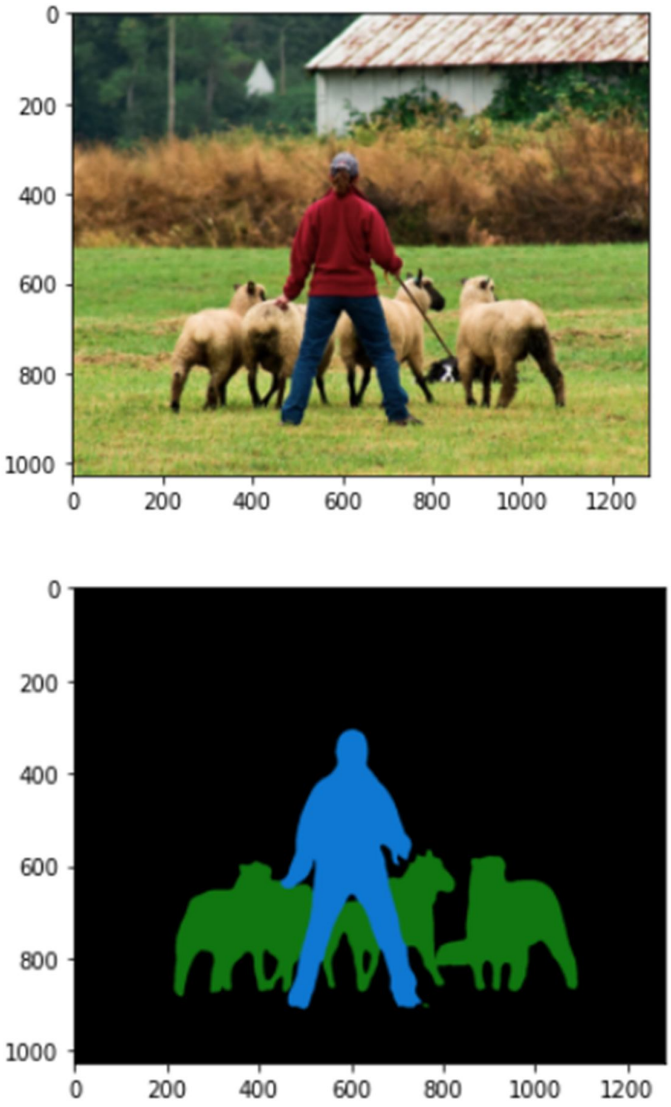

Figure27. Multi-Classes Segmentation

FCN:

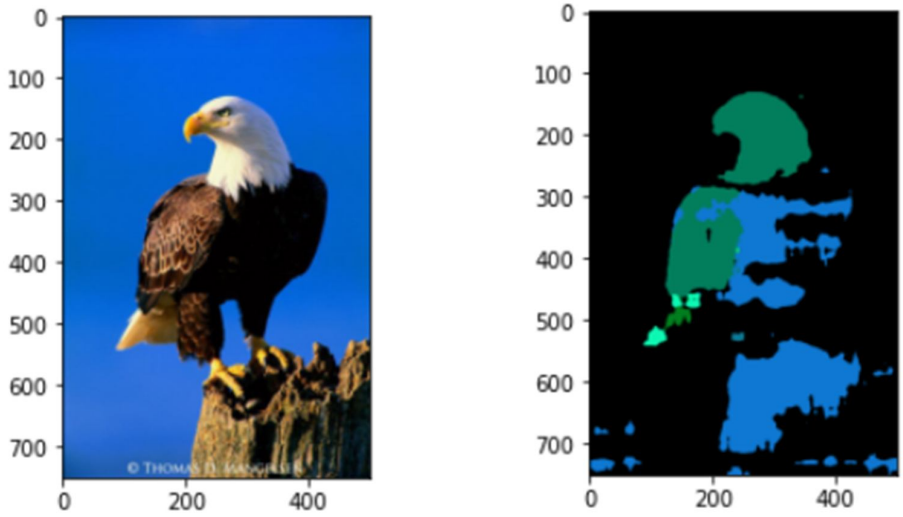

Figure28. One-Class Segmentation

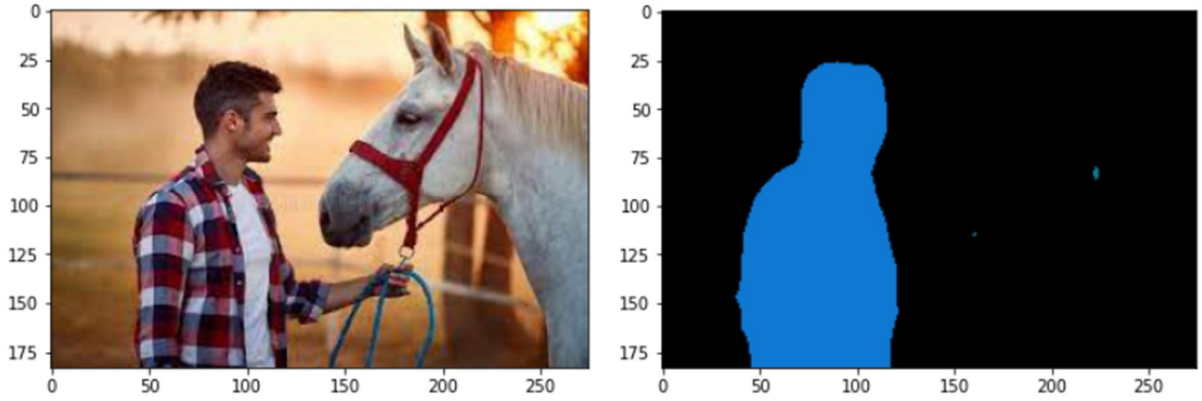

Figure29: Two-Classes Segmentation

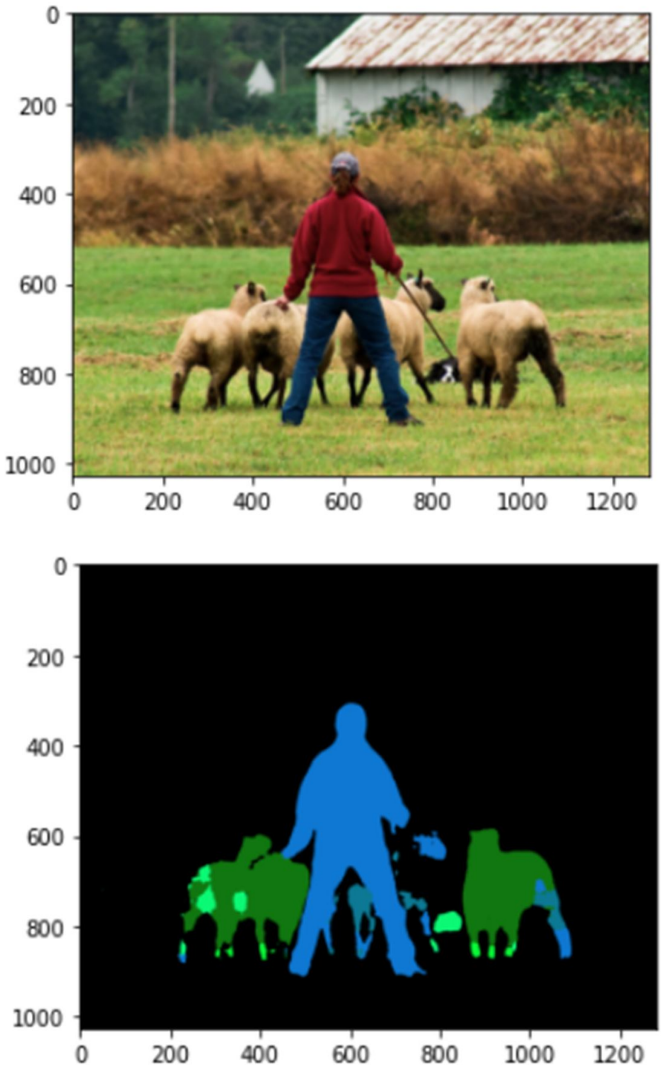

Figure30: Multi-Classes Segmentation

Analyze:

As we expected, **the DeepLab V3** acts **pretty much better** in comparison to the **FCN models.**

Because according to the model's architecture, DeepLab uses **more down-sampling module** which helps the model to extract more detailed features from the input image and then It leads to have a better classification accuracy and consequently have a **better power to recognize** every separated classes.

One more thing I want to mention here is about encoding image information module in **DeepLab V3 architecture** which let us to get a **better rate** in **task of encoding information** such as **edges** and **floors** detection which leads to have a better strength for **classifying different classes.**