SEMANTIC IMAGE SEGMENTATION USING DEEP-LEARNING MODEL

GUIDED BY: PROF. PANKAJ KUMAR
SUBMITTED BY: TANVI LAKKAD(201811010)
SUESHA GUPTA(201811069)

INTRODUCTION

Semantic Image Segmentation is a process of assigning each pixel of an image to different class labels.

Applications of Semantic Image Segmentation:

- Medical Image Segmentation
- Satellite Image Analysis
- Autonomous Driving
- Industrial Inspection

PROBLEM STATEMENT

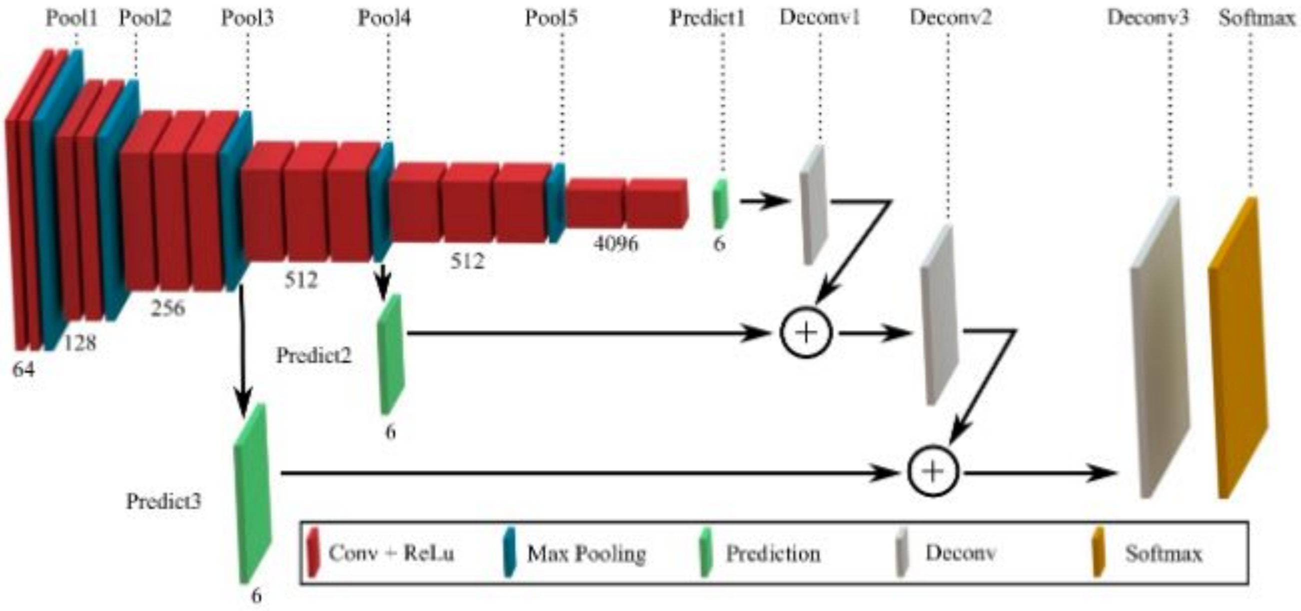
- Study of deep learning based semantic segmentation models.
- Understanding FCN and Deeplabv3 architecture.
- To perform training on deep learning models FCN and Deeplabv3.

Fully Convolutional Networks (FCN)

General semantic segmentation architecture comprises of :

- **Encoder:** It is a pre-trained classification network like VGG/ResNet to extract lower resolution feature mappings.
- **Decoder:** It is used to semantically project the discriminative features (lower resolution) learnt by the encoder onto the pixel space (higher resolution) to get a dense classification.

FCN Architecture



FCN Architecture

- It transfers knowledge from VGG16 to perform semantic segmentation.
- The fully connected layers of VGG16 is converted to fully convolutional layers, using 1x1 convolution. This process produces a class presence heat map in low resolution.
- The upsampling of these low resolution semantic feature maps is done using transposed convolutions.
- At each stage, the upsampling process is further refined by adding features from coarser but higher resolution feature maps from lower layers in VGG16.

FCN Limitation

Fuzzy object boundaries:

- By propagating through several alternated convolutional and pooling layers, the resolution of the output feature maps is down sampled. Therefore, the direct predictions of FCN are typically in low resolution, resulting in relatively fuzzy object boundaries.

DeepLabv3

It focuses on three main components:

- The Resnet architecture
- Atrous Convolutions
- Atrous Spatial Pyramid Pooling (ASPP)

Atrous Convolution

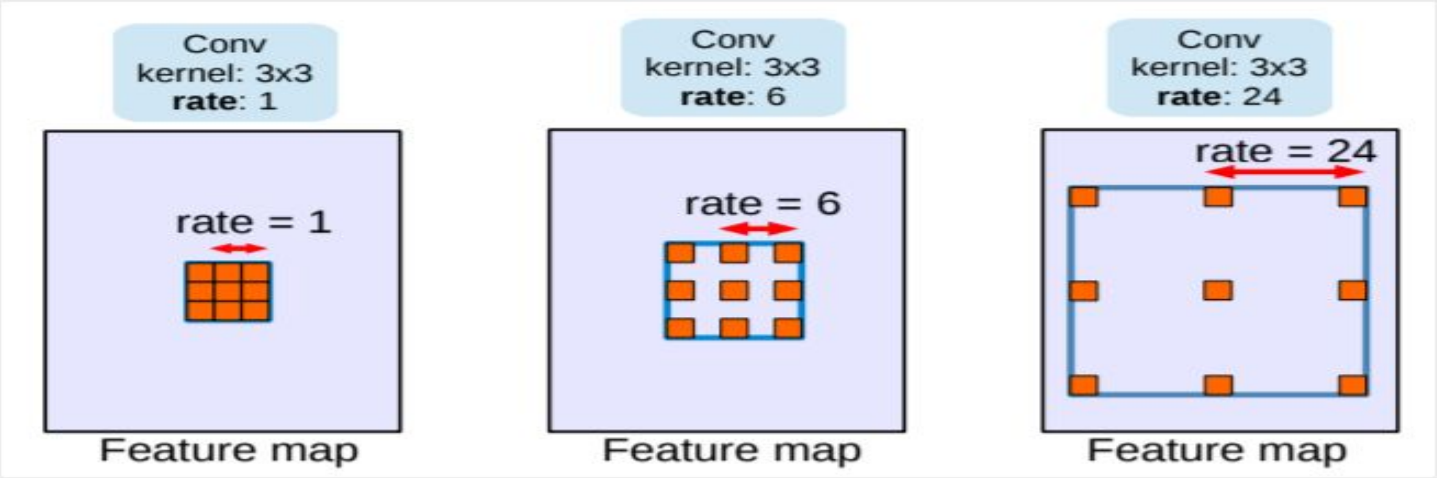


Figure 1. Atrous convolution with kernel size 3 × 3 and different rates. Standard convolution corresponds to atrous convolution with $rate = 1$. Employing large value of atrous rate enlarges the model's field-of-view, enabling object encoding at multiple scales.

Atrous Spatial Pyramid Pooling (ASPP)

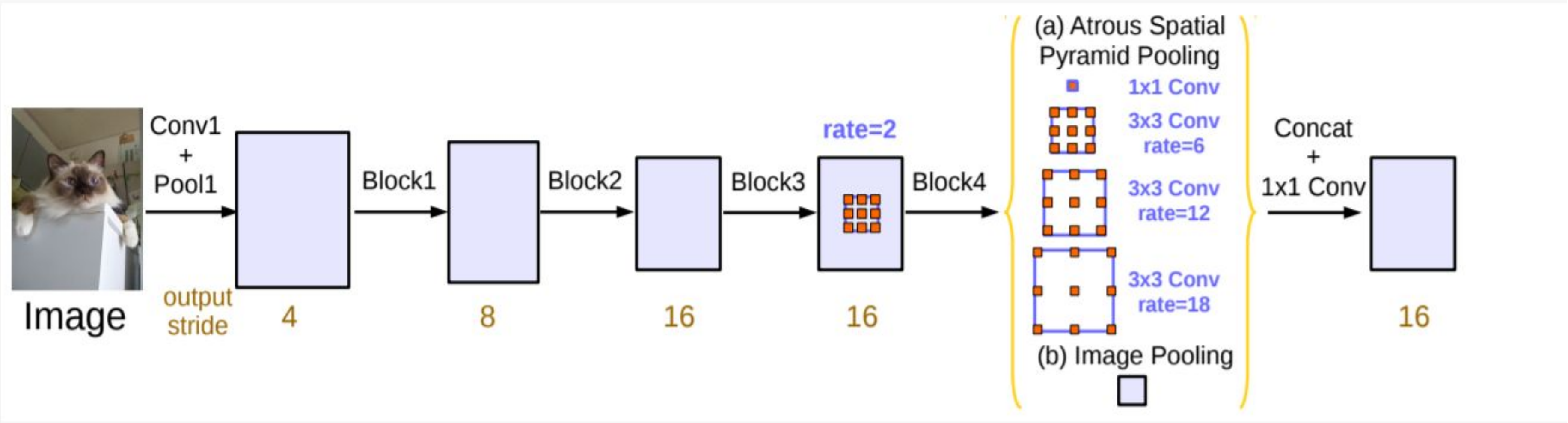


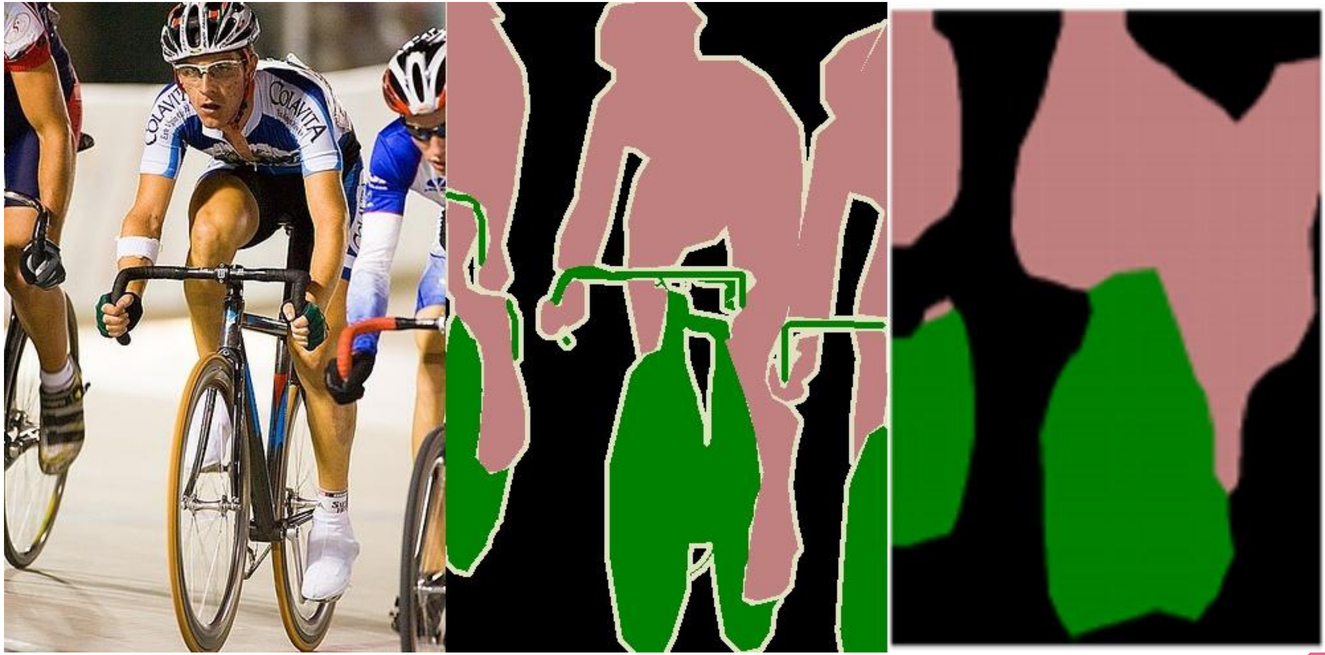
Figure 5. Parallel modules with atrous convolution (ASPP), augmented with image-level features.

Dataset Description - PASCAL VOC 2012

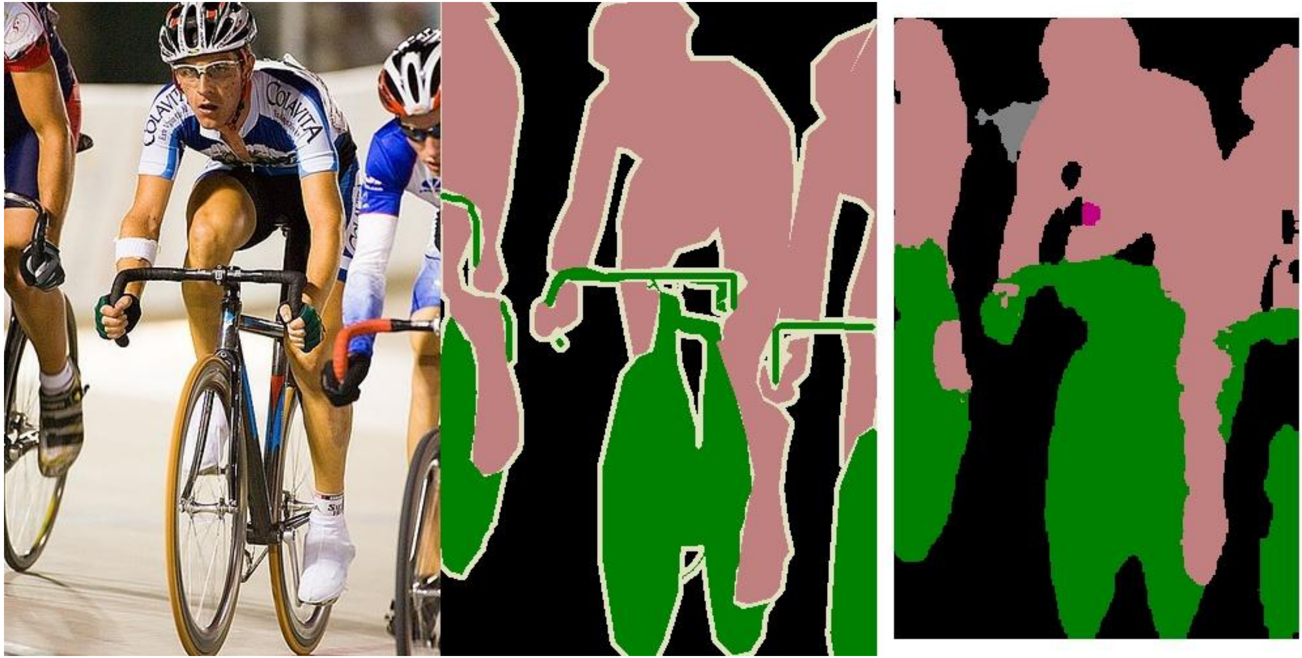
The dataset contains:

- Original images
- 1464 training images
- 1449 validation images
- 1456 test images
- 21 different classes of images eg: bus, car, person, monitor,etc.
- Ground truth 21 segmented images corresponding to each image

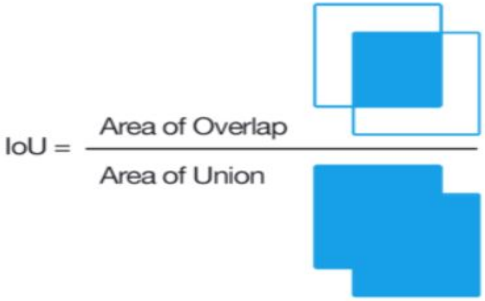
Results (FCN)



Results (DeepLabv3)



Model Evaluation



IoU calculation visualized. Source: Wikipedia

Model	mIoU score
Fully Convolutional Network (FCN)	0.232
DeepLabv3	0.747

References

- Evan Shelhamer, Jonathan Long, Trevor Darrell “Fully Convolutional Networks for Semantic Segmentation” Computer Vision and Pattern Recognition 2015
- Liang-Chieh Chen, George Papandreou, Florian Schroff, Hartwig Adam “Rethinking Atrous Convolution for Semantic Image Segmentation” Computer Vision and Pattern Recognition 2017

Thank You