Computer Vision IT524 Project

(Suesha Gupta-201811069 Tanvi Lakkad-201811010)

Semantic Image Segmentation using deep learning on the PASCAL VOC2012

Abstract

In this project, we have performed semantic image segmentation using DeepLab v3 model and Fully Convolutional Networks (FCN). The dataset used for training and testing and analytical comparison is PASCAL VOC2012 consists of 1464 training images, 1449 validation images and 1456 test images. Transfer learning is used for training FCN and DeepLabv3 model and each network has been trained for NN epochs before getting the results for analysis. Pre-trained Resnet 101 and VGG-16 weights have been used to initialise the training process in DeepLabv3 and FCN model respectively. The pre-processing, input and output layers have been modified to match the data organisation. The results of training, validation, and testing of the two deep learning CNN's have compared in the form of mIoU. We found the performance of DeepLabv3 is better than that of FCN based on their mIoU score.

Keywords: Deeplab; Fully Convolutional Networks; Mean Intersection over Union (mIoU)

1. Introduction

Semantic image segmentation is a process of assigning each pixel of an image to different class labels. It has various applications in image processing and computer vision domain such as medical area, autonomous driving, industrial inspection and many more. Over the years there were many techniques proposed for this task such as semantic segmentation using thresholding, regions and parts [1], graph based semantic segmentation [2], etc.

2. Dataset Description

The PASCAL VOC2012 dataset consists of 1464 training images, 1449 validation images and 1456 test images. It has total 20 object classes and 6929 objects instances. For each test image pixel, predict the class of the object containing that pixel or 'background' if the pixel does not belong to one of the twenty specified classes. The output from system should be an indexed image with each pixel index indicating the number of the inferred class (1-20) or zero, indicating background.

3. Choice of Deep Learning Networks

There are few state-of-the-art models on image semantic segmentation challenges such as FCN [3], segnet [4], Deeplab [5], Deeplabv3 [6]. One of the main issues between all the architectures is to take into account the global visual context of the input to improve the

prediction of the segmentation. The state-of-the-art models use architectures trying to link different part of the image in order to understand the relations between the objects.

For our project, we have chosen FCN [3] and Deeplabv3 [6] model.

4. Data Preprocessing

Assign each class a unique ID. In the segmentation images, the pixel value should denote the class ID of the corresponding pixel. We have generated segmentation maps from Pascal VOC dataset for all the images. The size of the input image and the segmentation image should be the same.

5. Methodology

Fully Convolution Network (FCN)

Fully convolutional indicates that the neural network is composed of convolutional layers without any fully-connected layers usually found at the end of the network. For the segmentation task, spatial information should be stored to make a pixel-wise classification. FCN allows this by making all the layers of VGG to convolutional layers. FCN motivates the use of fully convolutional networks by "convolutionalizing" popular CNN architectures e.g. VGG can also be viewed as FCN.

The model FCN8 deplicates VGG16 net by discarding the final classifier layer and convert all fully connected layers to convolutions. It appends a 1 x 1 convolution with channel dimension the same as the number of segmentation classes (in our case, this is 21) to predict scores at each of the coarse output locations, followed by upsampling deconvolution layers which brings back low resolution image to the output image size. In our example, output image size is (output_height, output_width) = (224,224).

DeepLabv3

The DeepLabv3 employs atrous convolution with upsampled filters to extract dense feature maps and to capture long range context. Specifically, to encode multi-scale information, the cascaded module gradually doubles the atrous rates while the atrous spatial pyramid pooling module augmented with image-level features probes the features with filters at multiple sampling rates and effective field-of-views.

DeepLab V3 uses ImageNet's pretrained Resnet-101 with atrous convolutions as its main feature extractor. In the modified ResNet model, the last ResNet block uses atrous convolutions with different dilation rates. It uses Atrous Spatial Pyramid Pooling and bilinear upsampling for the decoder module on top of the modified ResNet block.

We have used 700 images for training the model, 200 images for validation and 200 images for testing the models. The optimizer used is SGD (Stochastic Gradient Descent) with 32 batch size and learning rate is 0.01. The model has been trained for 600 epochs.

6. Results

Fig 1. Results obtained from FCN v3 a) Original image b) ground truth c) segmented output image

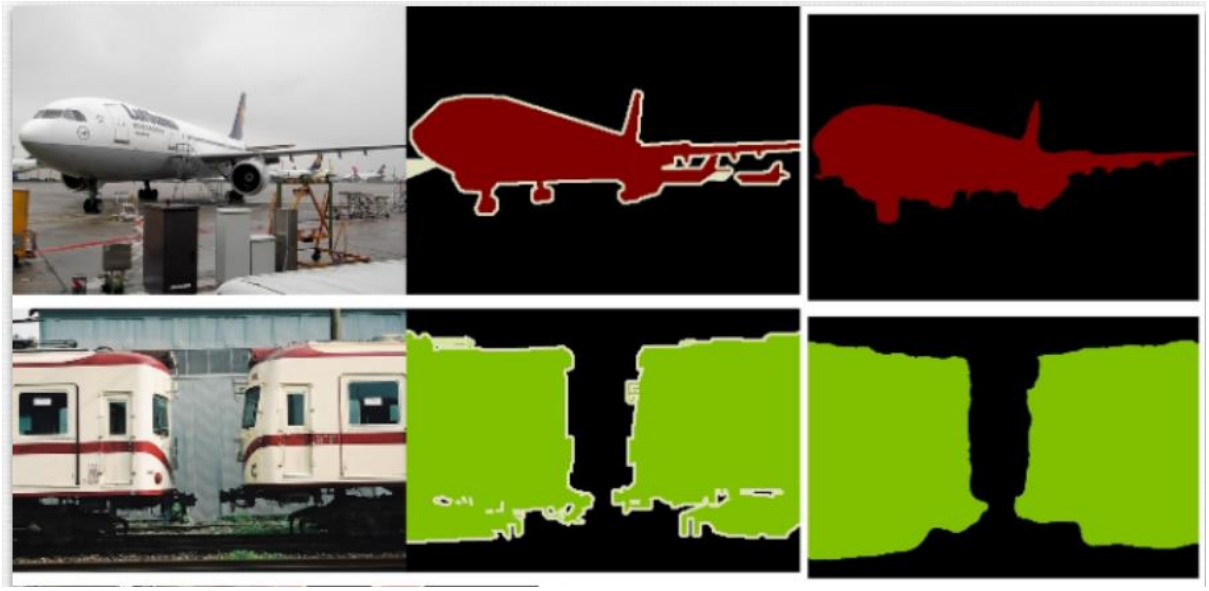

Fig 2. Results obtained from DeepLabv3 v3 a) Original image b) ground truth c) segmented output

Model Evaluation

Model	mIoU score
FCN	0.232
DeepLabv3	0.747

References

1. Pablo Arbel´aez, Bharath Hariharan, Chunhui Gu, Saurabh Gupta, Lubomir Bourdev and Jitendra Malik "Semantic Segmentation using Regions and Parts" Computer Vision and Pattern Recognition, IEEE Computer Society Washington, DC, USA (2012), pp. 3378-3385
2. Cevahir Çiğla; A. Aydın Alatan "Efficient graph-based image segmentation via speeded-up turbo pixels" 2010 IEEE International Conference on Image Processing, Sept. 2010
3. Evan Shelhamer, Jonathan Long, Trevor Darrell "Fully Convolutional Networks for Semantic Segmentation" Computer Vision and Pattern Recognition 2015
4. Vijay Badrinarayanan, Alex Kendall, Roberto Cipolla "SegNet: A Deep Convolutional Encoder-Decoder Architecture for Image Segmentation" IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 39, Issue: 12, Dec. 1 2017)
5. Liang-Chieh Chen, George Papandreou, Iasonas Kokkinos, Member and Alan L. Yuille "DeepLab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs Kevin Murphy" IEEE Transactions on Pattern Analysis and Machine Intelligence (Volume: 40, Issue: 4, April 1 2018).
6. Liang-Chieh Chen, George Papandreou, Florian Schroff, Hartwig Adam "Rethinking Atrous Convolution for Semantic Image Segmentation" Computer Vision and Pattern Recognition 2017.