Project Report - Object Segmentation

Kavu Maithri Rao
Matriculation No: 7002954

Prakash Kondibhau Naikade
Matriculation No: 7000433

Abstract

In this project we tried to design, implement, train and evaluate U-Net, Residual Convolutional Neural Network (RCNN), Recurrent Residual Convolutional Neural Network (RRCNN) based on U-Net architecture using PASCAL-VOC and cityscape dataset.

1 Introduction

Computer vision is an interdisciplinary scientific field that deals with how computers can gain high-level understanding from digital images or videos. From the perspective of engineering, it seeks to understand and automate tasks that the human visual system can do.(Wikipedia)

There are many problems defined in computer vision like image classification, classification with localization, object detection, semantic segmentation and many more.

In this project, we tried to understand semantic segmentation with various deep learning techniques based on U-Net. The aim of the semantic segmentation task is to label every pixel to respective class; this classes may represents different objects and background.

The output of the semantic segmentation task is high resolution image of same size of input image in which every pixel represents particular class. So it is a pixel level classification. This problem is also referred as dense prediction as it classifies each pixel for corresponding class.

There are numerous applications of semantic segmentation like in autonomous driving, precision agriculture, Geo sensing, bio medical image diagnosis, etc.

2 Methodology

2.1 Task 1

In task 1, we used U-Net architecture for segmentation task on PASCAl-VOC dataset [1]. U-Net architecture for segmentation task is proposed by Ronneberger et al. for Bio Medical Image Segmentation. It is made up of a contracting path and an expansive path. The contracting path follows the convolutional network architecture. It consists of the repeated application of two 3x3 convolutions which are unpadded convolutions, each followed by a rectified linear unit (ReLU) and a 2x2 max pooling operation with stride 2 for downsampling. For each downsampling step we double the number of feature channels. The expansive path consists of an upsampling of the feature map followed by a 2x2 convolution ("up-convolution") that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3x3 convolutions, each followed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution. At the final layer a 1x1 convolution is used to map each 64-component feature vector to the desired number of classes. The network has 23 convolutional layers. To model our architecture we referred this code on GitHub:https://github.com/Dhr11/Semantic$_{s}egmentation/blob/master/Unet_{m}odel.py$.

34th Conference on Neural Information Processing Systems (NeurIPS 2020), Vancouver, Canada.

2.2 Task 2 and Task 3

2.2.1 Architecture

In task 2, we used Res-Net and R2U-Net for segmentation task on cityscape dataset [2]. The R2U-Net architecture consists of convolutional encoding and decoding units same as U-Net. But instead of regular forward convolutional layers in both the encoding and decoding units, RCLs with residual units are used.

The architecture proposed in this paper differs from the UNet in following ways: First,The recurrent residual block is used instead of the traditional conv + relu layer in the encoding and decoding process, which can effectively increase the network depth. Second, Adopt feature summation at different time steps to get more expressive features, which also helps to extract lower-level features. Third, When skip connections are made, the original UNet cropping method is not adopted, but only concatenation operation is used, which can improve network complexity and accuracy.

Therefore, this architecture has the following advantages over UNet: First, Same parameters as UNet but better segmentation performance. Second, more efficient feature accumulation which ensures stronger features representation. Third, the proposed residual and recurrent methods are versatile and can be transplanted to other networks such as SegNet, 3D-UNet, V-Net, etc.

For ResNet we referred this code on GitHub:https://github.com/SatyamGaba/semantic$_{s}egmentation_{c}ityscape$
For R2U-Net we referred this code on GitHub:https://github.com/bigmb/Unet-Segmentation-Pytorch-Nest-of-Unets/blob/master/Models.py

2.2.2 Optimizer

The paper suggests to use ADAM as a optimizer for segmentation task based on U-Net architecture. Also, ADAM uses the power of adaptive learning rates methods to find individual learning rates for each parameter. ADAM combines the advantages of Adagrad, which works well with sparse gradients, and RMSprop, which works well in on-line settings as ADAM works as combination of RMSprop and SGD with momentum. In Adam, update rule of step size is independent of the magnitude of the gradient, which helps a lot when going through areas with saddle points. In these areas SGD sometimes fails to navigate through them.

2.2.3 Loss Function

Cross-entropy is defined as a measure of the difference between two probability distributions for a given random variable or set of events. It is widely used for classification objective, and as segmentation is pixel level classification it works well. So we decided use binary Cross-Entropy for task 2 as it works best in equal data distribution among classes scenarios [3].

2.2.4 Learning Rate

After many attempts we settled on ADAM optimizer with learning rate of 2*10^{-}4.

3 Results

3.1 Metric Graphs

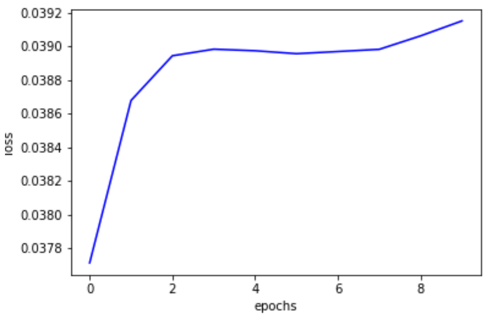

Figure 1: Dice coefficient for U-Net

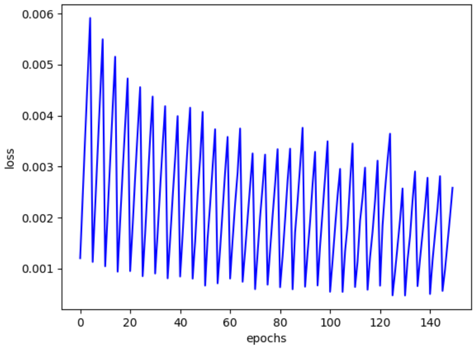

Figure 2: Training loss for Res-Net

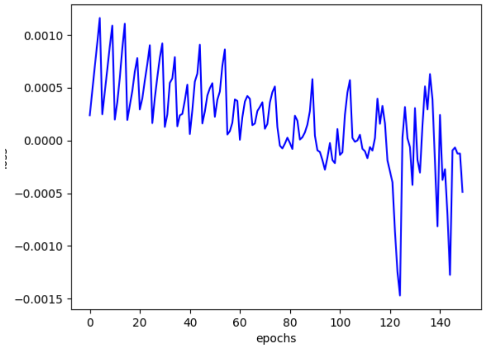

Figure 3: Training loss for R2U-Net

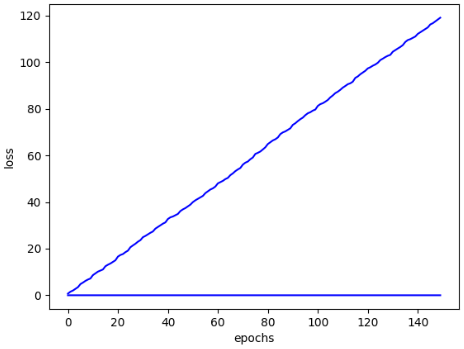

Figure 4: Validation loss for R2U-Net

3.2 Metrics

Table 1: EXPERIMENTAL RESULTS FOR DIFFERENT ARCHITECTURE WE USED FOR SEGMENTATION TASK ON TEST DATASET

Model	DC	JSC	F1-Score	Accuracy
ResNet	0.7803	0.3699	0.4555	0.7780
R2U-Net	0.0000	0.1351	0.1754	0.5404

Table 2: EXPERIMENTAL RESULTS FOR DIFFERENT ARCHITECTURE WE USED FOR SEGMENTATION TASK ON VALIDATION DATASET

Model	DC	JSC	F1-Score	Accuracy
U-Net	0.0024	0.0345	0.0400	
ResNet	0.8256	0.2878	0.3373	0.7960
R2U-Net	0.0000	0.1510	0.1824	0.5744

IoU/Jaccard Index is the area of overlap between the predicted segmentation and the ground truth divided by the area of union between the predicted segmentation and the ground truth, as shown on the image to the left. This metric ranges from 0–1 with 0 signifying no overlap and 1 signifying perfectly overlapping segmentation. By looking at our results it seems our model is inefficient to perform segmentation task.

Dice Coefficient is 2 * the Area of Overlap divided by the total number of pixels in both images. By looking at our results it seems our model is inefficient to perform segmentation task.

4 Conclusion

In this project, we tried to implement and evaluate “U-Net”, “RU-Net” and “R2U-Net”. These models were evaluated using two different datasets and on four different metrics. Experimental results shows we are not able to achieve already set standard benchmarks. In future, we will continue improving results using the models by increasing our understanding of architectures, python coding and creating different variations of architectures.

References

[1] Olaf Ronneberger & Philipp Fischer & Thomas Brox (2015) U-Net: Convolutional Networks for Biomedical Image Segmentation.

[2] Md Zahangir Alom & Mahmudul Hasan & Chris Yakopcic &Tarek M. Taha &Vijayan K. Asari(2018) Recurrent Residual Convolutional Neural Network based on U-Net (R2U-Net) for Medical Image Segmentation

[3] Shruti Jadon,(2020) A survey of loss functions for semantic segmentation