Assignment-4 Unsupervised Image Segmentation

Abhishek Kumar 18111002
Hemant Parihar 18111021

April 2019

1 Introduction

Image segmentation is the process of partitioning a digital image into multiple segments. The goal of segmentation is to simplify and/or change the representation of an image into something that is more meaningful and easier to analyze. Image segmentation is typically used to locate objects and boundaries (lines, curves, etc.) in images. More precisely, image segmentation is the process of assigning a label to every pixel in an image such that pixels with the same label share certain characteristics.

2 Challenges Faced

Unsupervised Segmentation

The task was unsupervised made it alot harder than the supervised domain. Since there was no notion of semanticness.

Judging Performance

There is no benchmark standard to judge the performance.

Mixed Colors

Shadows in images were creating problems for objects to get mixed with road and other objects.

Object Diffusion

Since there was no object notion given even algorithm like was getting confused between different objects. So segments of different objects were getting mixed.

Too Much learning (Over Fitting)

Since the data was being trained on static camera the neural network was getting over fitted and was predicting single segment for many objects.

3 Methods Tried

3.1 SLIC and Normalized CUT (SNC)

In this approach we divide the image into super pixels using slic algorithm. On this set of super pixels we apply normalized cut merging (i.e merge the superpixles if normalized cut[1] association between them is high).

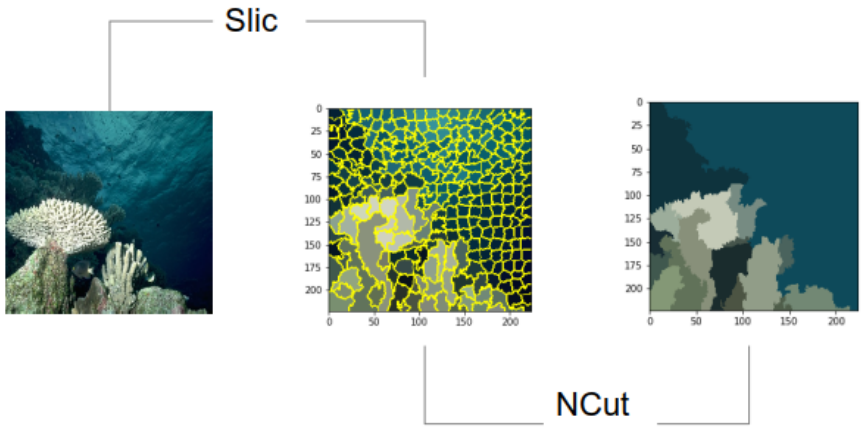

Figure 1:

3.2 SLIC with Back propagation (SBP)

One method we used was to use slic super pixels information to train a neural network to predict classes. We first find the super pixels of image based on the slic algorithm. We pass the image through an same padding convolutional network to classify each pixel into 500 classes. To get the target for these images we have used slic superpixels as follows: We took the model output and took maximum along the channel axis and for each superpixel we calculated the maximum of predicted labels in output and took that value as the target for that pixel to train neural network.

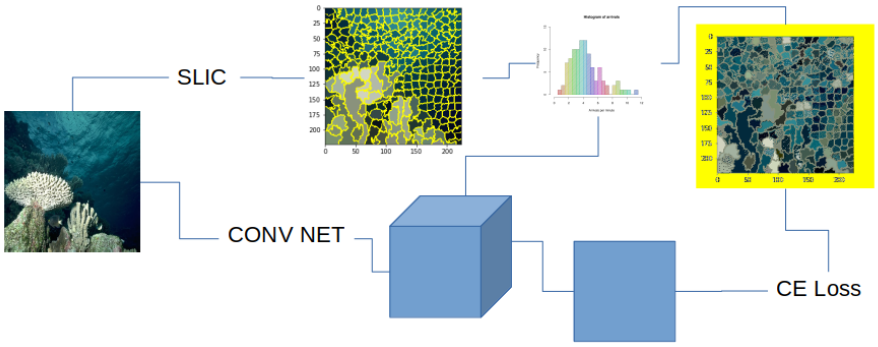

Figure 2: Slic with Backpropagation

3.3

3.4 W-Net: A Deep Model for Fully Unsupervised Image Segmentation

A fully convolutional network encoder produces a segmentation. This segmentation is fed into a fully convolutional network decoder to produce a reconstruction, and training jointly minimizes the normalized cut of the encoded segmentation and the reconstruction of the image. The encoded image is then post-processed to produce the final segmentation.

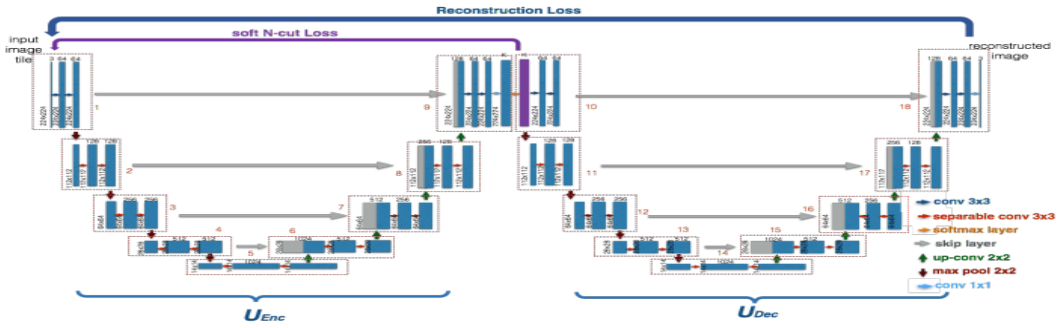

Figure 2. **W-Net architecture.** The W-Net architecture is consist of an U_{Enc} (left side) and a corresponding U_{Dec} (right side). It has 46 convolutional layers which are structured into 18 modules marked with the red rectangles. Each module consists of two 3×3 convolutional layers. The first nine modules form the dense prediction base of the network and the second 9 correspond to the reconstruction decoder.

Figure 3: Source: W-Net Paper

3.5 Using information as Video Optical Segmentation (OS)

We took the leverage of information that the frames are in sequence, we subtracted the background using two consecutive frames and took those output images as target for the neural network where input is second frame . The key thing here to note is that we need to avoid putting loss in place where there is no object detected so that model learns to predict segments without needing two frames. This work due to trade between the NCut segmentation and reconstruction loss between input and decoder output.

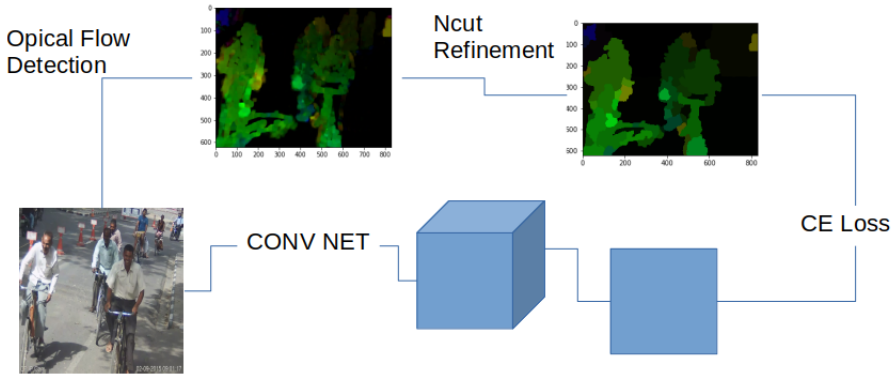

Figure 4: Optical Segmentation

3.6 Modifications Over above models

We did certain modifications over above models some of them were

3.6.1 Optical Flow info with Wnet Ncut loss (OFNC)

In addition to labels obtained using optical flow we also used ncut loss for the output to generate smooth segments.

3.6.2 Optical FLow info with slic as background labels(OFSL)

We instead of placing zero loss black background we instead created labels for background using slic output as described in SNC model.

3.6.3 Combined Everything (OFNC + OFSL)

We used Wnet architecture with Ncut loss and cross entropy loss over optical flow output, and Reconstruction loss euclidean loss.

4 Results

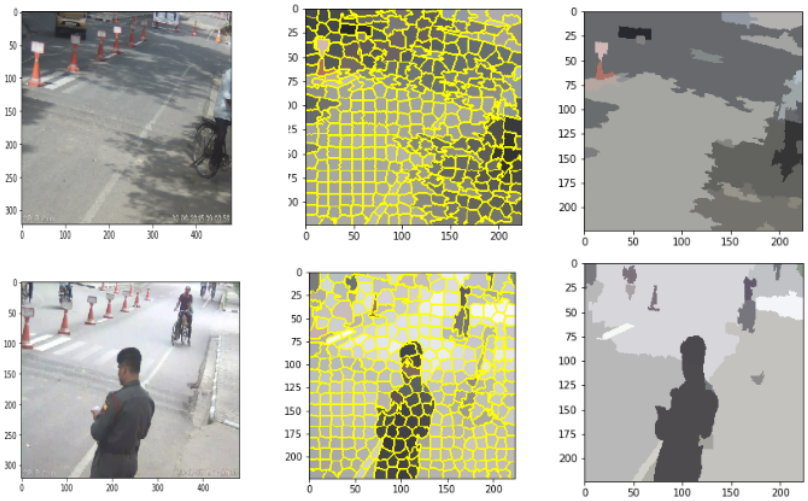

Figure 5: Slic and Ncut Result (SNC)

Figure 6: Slic and BackProp Result (SBP)

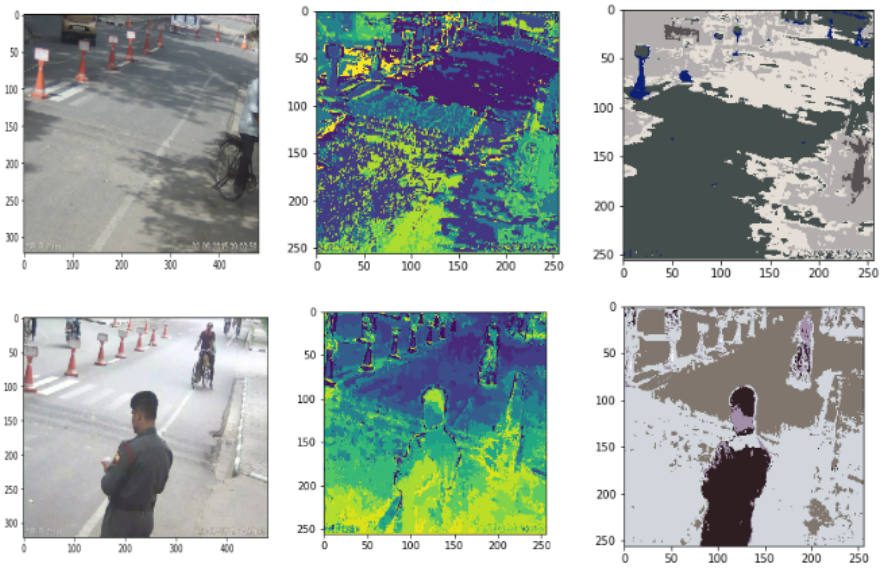

Figure 7: Optical Segmentation (OS)

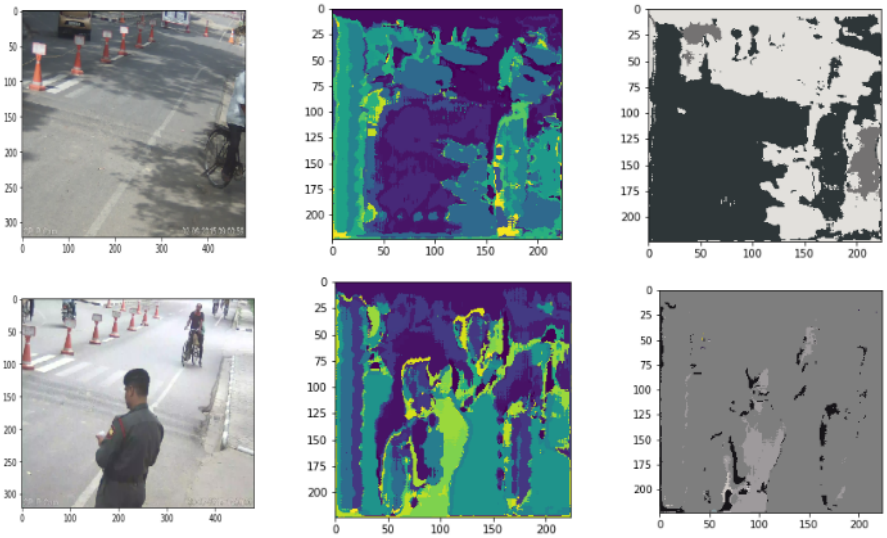

Figure 8: Wnet Segmentation)

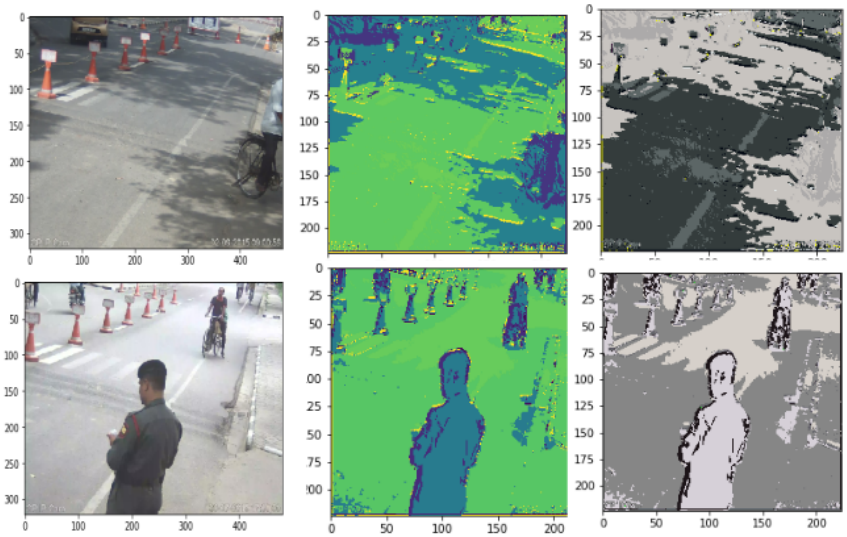

Figure 9: Optical Flow with slic labels and Ncut loss (OFNC + OFSL)

5 Conclusion

Since there are no ground truths its difficult to say how much correct each of this is. But just by manually checking we can say that Slic with Backprop and combined nueral network performs somewhat better than others.

6 References

1. Normalized Cuts and Image Segmentation, Jianbo Shi and Jitendra Malik, Member, IEEE TRANSACTIONS ON PATTERN ANALYSIS AND MACHINE INTELLIGENCE, VOL. 22, NO. 8, AUGUST 2000

2. https://arxiv.org/pdf/1711.08506.pdf
   W-Net: A Deep Model for Fully Unsupervised Image Segmentation