Analysis of Semantic Segmentation Techniques for Autonomous Vehicles

Anish Madan , 2016223
Apoorv Khattar , 2016016
Group 8

INDRAPRASTHA INSTITUTE *of*
INFORMATION TECHNOLOGY
DELHI

Introduction

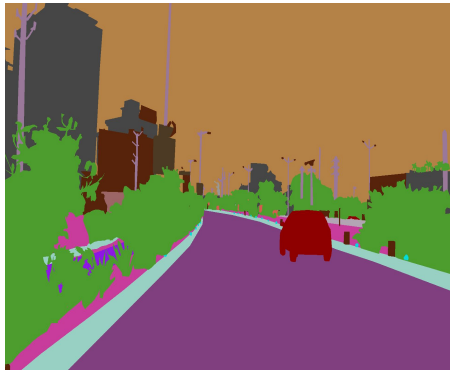
- **Semantic Segmentation** is useful for scene understanding, forming coherent clusters in image with meaningful attributes and has applications in vision-based tasks of autonomous vehicles which include obstacle detection, pedestrian detection, etc
- **Autonomous vehicles** require an understanding of the environment which is provided by the images from the camera.
- Segmentation is an important task in understanding the surroundings more robustly when compared to object detection, by carefully segmenting out pedestrians , obstacles, etc and uses this knowledge to take informed decisions about the state of the vehicle.

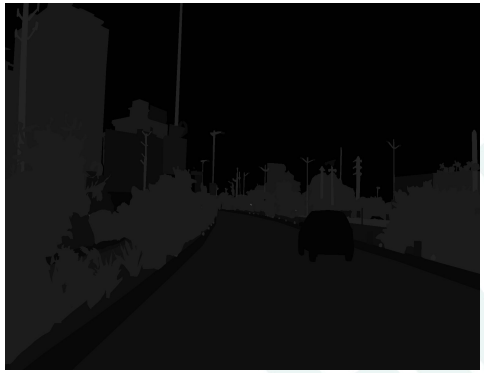
Dataset

- The dataset chosen for our task is the **Indian Driving Dataset**, curated by IIIT-Hyderabad.
- We chose this dataset since all the previous segmentation models available are trained on international datasets. But the type of classes required for Indian streets vary from other datasets like BDD or CamVid.
- This dataset consists of 39 classes and contains about ~7900 images with polygon annotations for different classes.

Data Preprocessing, Analysis and Pixel Level Annotations

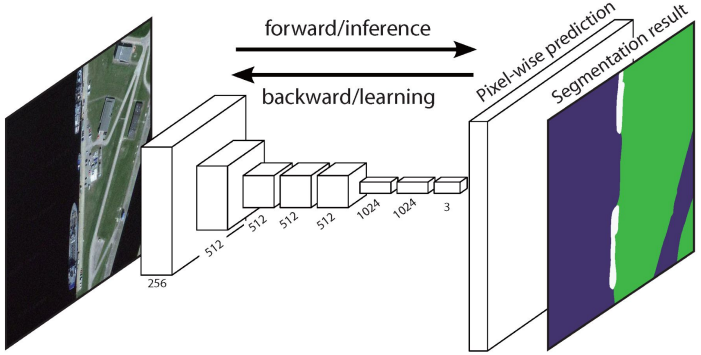
- There was discrepancy in some of the annotations as they did not have the polygons corresponding to the classes. Such cases were removed.
- Also, since we had polygon level annotations, it implied that there were overlapping polygons of different classes. This was solved by coloring the images according to classes, and then pixels were assigned class ids according to their color value.





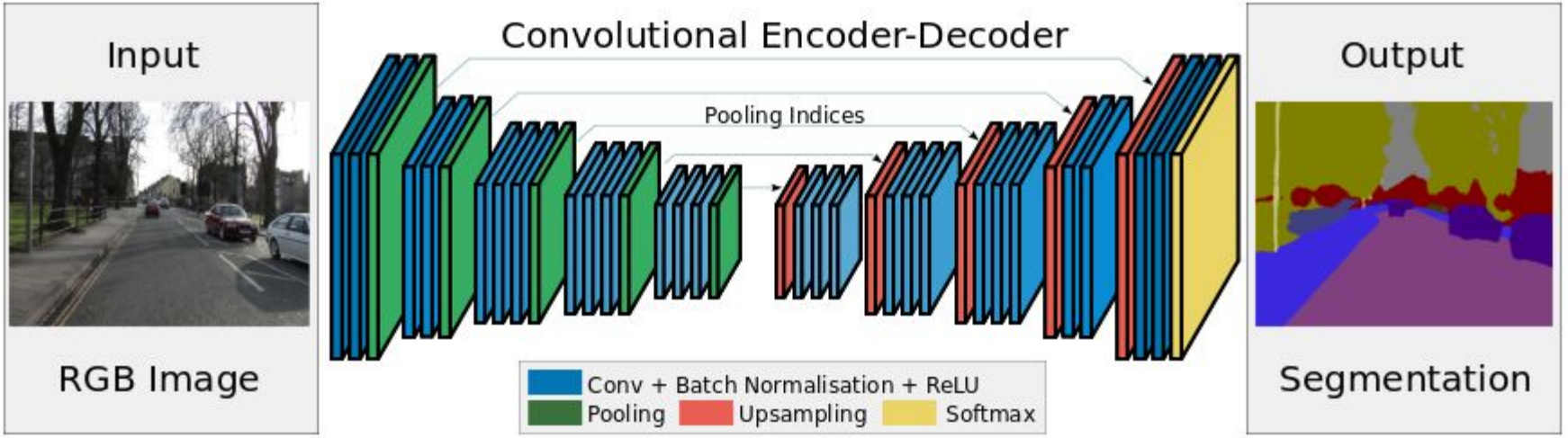
Approaches used

- **FCN(Fully Convolutional Network)**
  - It uses a VGG-16 network in which the final layer is converted into a 1X1 convolutional layer.
  - VGG-16 is used as a feature extractor then these features are upsampled to calculate the pixel wise label map.



Approaches Used – II

- **SegNet (Encoder-Decoder Architecture)**
  - ***Encoder -*** Topologically identical to VGG11 (removed dense layers)
  - ***Decoder -*** Consists of a hierarchy of decoders one corresponding to each encoder



Results

Model	mIOU	Avg Pixel Accuracy	Inference Time
FCN	0.4671	0.4795	0.238 s
FCN(weighted Loss)	0.3748	0.2874	0.238 s
SegNet	0.4680	0.4842	0.22 s
SegNet(weighted Loss)	0.3640	0.3234	0.22 s

Results – II

Ground Truth

SegNet

Both models trained for 12 epochs each

Weighted SegNet

Analysis

1. ***Class Imbalance***: The dataset consists of 39 classes which are quite imbalanced with the majority including classes like roads, sky, etc.
2. This problem results in mainly the segmentation of dominant classes and (maybe) gives a higher mIOU, but leaves out less dominant classes.
3. We try to tackle this problem by using a weighted loss function , i.e a weighted cross entropy loss. We assign small weights to dominant classes and vice versa. This is done by taking the count of the pixels belonging to a particular class in a mini batch and then taking its reciprocal.
4. We can observe the above behavior in the previous slide.

Analysis – II

FCN for Semantic Segmentation:

Since FCN has alternating convolutional and pooling layer, the output feature maps are downsampled. Therefore, the direct predictions of FCN are typically in low resolution, resulting in relatively fuzzy object boundaries.

One way to correct this, is to add activations of previous layers, this is called skip connections.

SegNet:

Eliminates the need for upsampling by using its decoder network to upsample its lower resolution input and by using max pooling indices from the encoder network.

We get better delineation of boundaries as compared to FCN.

Analysis – III & Contributions

- We see that both SegNet and FCN have comparable mIOU values and average pixel accuracy values. We incline towards the reason that this might be due to the imbalanced nature of classes.
- In the case of SegNet and FCN with weighted losses, we see that the mIOU and Avg pixel accuracy values go down, although we see more of an inclusion of other classes in the segmented regions. We believe that these results could be improved with more training.

Contributions

- Apoorv Khattar - FCN
- Anish Madan - SegNet