



SEMANTIC SEGMENTATION FOR SELF DRIVING CARS

M.Vaishnavi

Department of Computer Science and Engineering
Madanapalle Institute of Technology and Science,
Madanapalle, Andhra Pradesh, India

K.Varshitha

Department of Computer Science and Engineering
Madanapalle Institute of Technology and Science,
Madanapalle, Andhra Pradesh, India

G.Usha

Department of Computer Science and Engineering
Madanapalle Institute of Technology and Science,
Madanapalle, Andhra Pradesh, India

C.Mounika

Department of Computer Science and Engineering
Madanapalle Institute of Technology and Science,
Madanapalle, Andhra Pradesh, India

C. Narasimha^(ph.d)

Associate Professor

Department of Computer Science and Engineering
Madanapalle Institute of Technology and Science,
Madanapalle, Andhra Pradesh, India

Abstract— This paper proposes a novel approach for Semantic segmentation which is one of the biggest challenge increasing in an order and have been making humans hold keen active interest to result in fast and accurate semantic segmentation. Whereas At present, we are trying to solve this problem of semantic segmentation using the segnet which makes its more accurate interms of accuracy, computational time, and inference time. and here we are using segnet model to take this to the next level which includes max-pooling, Batch normalization techniques to map low-resolution features to input resolution for pixel-wise classification and the architecture here consists of an encoder which takes the input image and is identical to 13 convolutional layers and a decoder that uses segnet followed by pixel-wise classification layer. and also when compared with other architectures segnet provides good performance with competitive inference time and most efficient memory. So, therefore here we are presenting deep fully convolutional neural network architecture for semantic pixel-wise segmentation termed SegNet.

Keywords— Semantic segmentation, Fully Convolutional Layers, SegNet, Conv 2D, Max pooling, Batch Normalization

I. INTRODUCTION

Artificial Intelligence has revolutionized the area of autonomous vehicles like self-driving cars by incorporating complex models and algorithms. Self-driving cars require a deep understanding of their surroundings. To support this, we are using camera frames to recognize the road, pedestrians, cars, and sidewalks at pixel-level accuracy. A Self driving car is a vehicle capable of sensing its environment and operating with little or without human involvement. And with this, the popularity of these self-driving cars is also increasing tremendously. And it is also the one that involves Advanced concepts of Artificial Intelligence

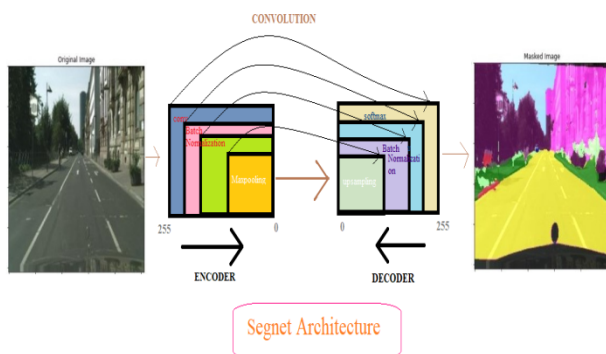
that have one of the subdomains i.e., Deep Learning.

Self driving cars work by using the semantic segmentation approaches which are one of the very important perceptions for a self-driving car. Here in this project, we use the segnet model for semantic segmentation and here we employ deep learning for semantic segmentation as after recognizing it realizes the persons or vehicles in the image at higher levels of neural network. The goal of semantic segmentation is to label the similar images as a particular semantic class. It performs pixel-wise classification of images. In a typical urban scene, the classes could be Pedestrians, Trees, Cars, and all other images. And we develop a deeplearning neural network model that involves a convolutional neural network model and does not require manual feature extraction and optimize it to perform semantic segmentation using Segnet. Semantic segmentation is the task of assigning the meaning of part of an object. this can be done at the pixel level where we assign each pixel to a target class such as road, car, pedestrian, sign, or any number of other classes. The architecture of this model is designed in such a way that it consists of an encoder, decoder, and pixel convolutional layer. The encoder uses a pre-trained network and preprocess the input image which is of low resolution and then comes the role of Decoder that includes Segnet network which is one of the fully convolutional networks using Conv2D, max-pooling which consumes low memory and is used to reduce the feature size and Batch Normalization produces the high-resolution image as output. And these neural networks are trained with powerful GPU's that GPU systems are used for inference on new data. And also here we include highly robust algorithms that facilitate the functioning of these self-driving cars by reducing problems like drunken driver problems and violation of traffic rules problems etc.

II. LITERATURE REVIEW

There are a lot of practices made to bring some variations among regular cars and self-driving cars. So far, in all the data gathered by analyzing various aspects in this particular field, it might appear that self-driving cars are much safer than human-driven cars, but a lot of research needs to be done to prove this. Self-driving cars are made safe, as long as their sensors and response times are all electronic and mechanical, making them by nature faster than humans. Automation can help much better to reduce the number of crashes on our roads. Automated cars had been implemented already in many countries with various architectures using deep learning techniques. The prevailing models lack localization accuracy and recognition performance and also, Many models are power consuming and therefore the accuracy was low. Government data identifies driver behavior or error as an element in 94 percent of crashes, and self-driving vehicles can help reduce driver error. Higher levels of autonomy have the potential to scaleback risky and dangerous driver behaviours. Here we're solving the problem of a self-driving car by using deep learning for the complete autonomous integrated driving stack from perception, to motion planning, to controls. We're using an integrated architecture to make a more seamless approach. The deep learning approach instead learn by itself what to do by fundamentally understanding the data. Deep learning is far more analogous to the way humans learns. we show an algorithm good and bad examples, and so that it learns to generalize. The proposed model attempts to design an efficient and low power consumption by utilizing deep learning techniques like CNN as they reduce vanishing gradient problems, encourage feature reuse and strengthen feature propagation. And therefore the goal is to improve the accuracy of the system. For a dynamic environment that's extremely complex, we believe this is often the best way to solve the problem.

III. SEGNET ARCHITECTURE



The Architecture we use is segnet architecture which is one of the effective architecture. It consists of an encoder-decoder pair.

Encoder :

The first layer consists of 64 filters.

The second layer consists of 128 filters.

The third layer consists of 256 filters.

The fourth layer consists of 512 filters.

The fifth layer consists of 1024 filters.

Decoder:

The first layer consists of 1024 filters.

The second layer consists of 512 filters.

The third layer consists of 256 filters.

The fourth layer consists of 128 filters.

The fifth layer consists of 64 filters.

And the very last layer consists of 13 filters.

The encoder consists of VGG-16 which is pretrained. and decoder followed by softmax unit. softmax is said to be a pixel-wise classification layer. and the segnet consists of all 13 convolutional layers. And here, the images of the dataset pixel values range from 0 to 255.the maximum pixel value here is 255.and also most importantly, the segnet encoder part consists of max-pooling indices and batch normalization to make the training process easier and within very efficient computational time.

IV. WORK INVOLVED IN SEMANTIC SEGMENTATION

In this section, we are going to discuss in brief the major work involved in semantic segmentation.

A. SEMANTIC SEGMENTATION WITH FCN USING SEGNET

Semantic segmentation can be described as a recognizing and understanding of an image at a pixel level. and working on semantic segmentation is a big task. To achieve this task we have come up with using FCN. FCN can be known as fully convolutional networks. These are also the networks that yield hierarchies of features. And FCN can be trained by themselves end to end, pixels to pixels, and also are capable of improving themselves based on the best previous result that occurs on using semantic segmentation. Also, this could help result in efficient inference and learning time.

The model of semantic segmentation we are using here is Segnet. Segnet can be said to be a core trainable architecture. This segnet architecture consists of a deep convolution Encoder-Decoder Network following a pixel-wise classification layer. next comes the section that discusses the implementation of it.

Now the Segnet model is implemented. The boundaries of the images are then extracted which can form as a vector matrix and sent to the encoder.

The image is generally of the dimensions 256*256 pixels, during which all the objects like vehicles, trees, pedestrians, etc., are present which can be classified.

For each pixel, there exists a particular color so there'll be plenty more colors during a single image. To cluster, all the closest colors joined we use the K-means model. By using the function color(), Reshape() these are done.

The clusters are then grouped into classes and are given class labels. The classes include Roads, sky, trees and all other things.

The segnet model has the attributes of:

Convolutions – To detect the boundaries and extract features.

Maximum pooling – To decrease the size of the matrix.

It uses Max Pooling which is used to reduce the feature size and also it uses low memory. The max-pooling reduces the dimensions of the vector after processing which leads to a decrease within the resolution. Then the training images are

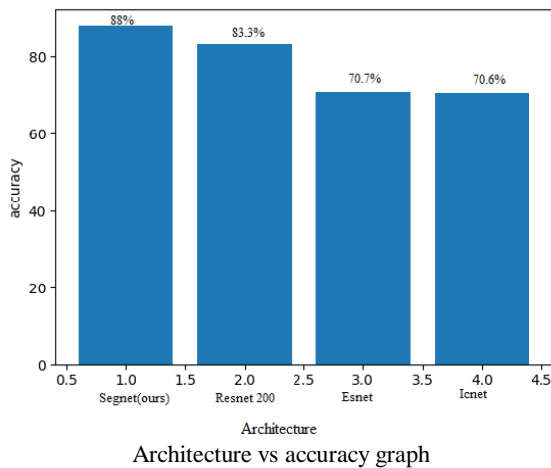


going to be taken by using a data generator. Because the training images are large in number, we use the data generator within which the number of images that are sent is divided into batches. For a batch, we will take the 'n' number of images and repeat it till we train all the images. To increase the resolution, the up sampling is going to be done. Then the predicted image is going to be generated. The Prediction gives the original image, masked image, and the resulted image.

V. COMPARISON OF SEGNET WITH OTHER NETWORKS

The Segnet possesses honest results for semantic segmentation tasks in self-driving cars as compared with other networks. And it's one among the foremost efficient significant architecture for pixel-wise classification. Segnet model achieves great segmentation performance with competitive inference memory and efficient computational time.

Graph1:



The above comparison table shows the accuracies occurred by segnet and other networks on using cityscapes dataset for semantic image segmentation of self-driving cars.

VI. EXPERIMENT AND RESULT

On Analysis, we observed that the training process of the segnet model is done in 10 epochs and for each epoch of the full training dataset the accuracy improves much better.

```

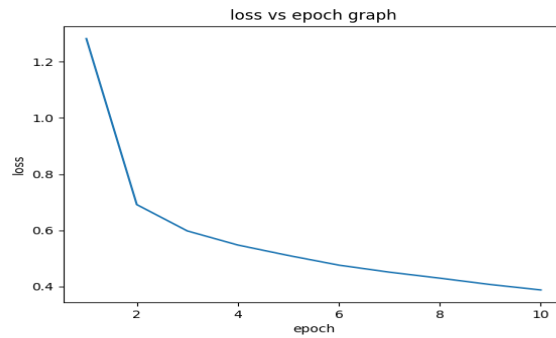
segmentation
File Edit View Run Add-ons Help
Python 3.7.10 Shell
Train params: 71,854,477
Trainable params: 73,396,125
Non-trainable params: 29,352

history = model.fit_generator(train_gen, epochs=10, steps_per_epoch=num_of_training_samples//batch_size,
                             validation_data=val_gen, validation_steps=num_of_testing_samples//batch_size)

App/conda/lib/python3.7/site-packages/tensorflow/python/keras/engine/training.py:1844: UserWarning: Model.fit_generator() is deprecated and will be removed in a future version. Please use Model.fit(), which supports generators.
warnings.warn('Model.fit_generator() is deprecated and will be removed in a future version. Please use Model.fit(), which supports generators.', UserWarning)
WARNING:tensorflow: Model.fit_generator() is deprecated and will be removed in a future version. Please use Model.fit(), which supports generators.
epoch 1/10 [-----] - loss: 1.2088 - accuracy: 0.6373 - val_loss: 0.8208 - val_accuracy: 0.7634
epoch 2/10 [-----] - loss: 0.6933 - accuracy: 0.8062 - val_loss: 0.6839 - val_accuracy: 0.8028
epoch 3/10 [-----] - loss: 0.5938 - accuracy: 0.8257 - val_loss: 0.6368 - val_accuracy: 0.8287
epoch 4/10 [-----] - loss: 0.5456 - accuracy: 0.8352 - val_loss: 0.6382 - val_accuracy: 0.8222
epoch 5/10 [-----] - loss: 0.5633 - accuracy: 0.8403 - val_loss: 0.6609 - val_accuracy: 0.8228
epoch 6/10 [-----] - loss: 0.4742 - accuracy: 0.8568 - val_loss: 0.5928 - val_accuracy: 0.8276
epoch 7/10 [-----] - loss: 0.4647 - accuracy: 0.8645 - val_loss: 0.5984 - val_accuracy: 0.8222
epoch 8/10 [-----] - loss: 0.4224 - accuracy: 0.8728 - val_loss: 0.6287 - val_accuracy: 0.8206
epoch 9/10 [-----] - loss: 0.4622 - accuracy: 0.8762 - val_loss: 0.6339 - val_accuracy: 0.8288
epoch 10/10 [-----] - loss: 0.3778 - accuracy: 0.8839 - val_loss: 0.5982 - val_accuracy: 0.8245
    
```

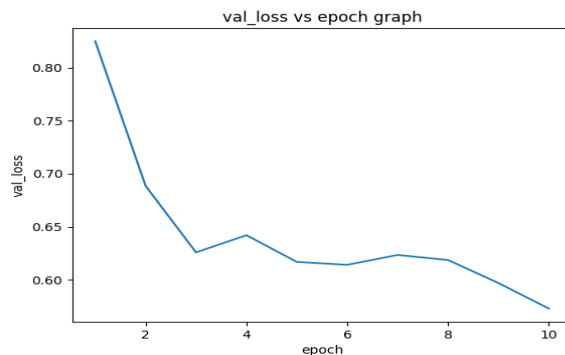
Eventhough the Segnet model is significantly small but also, it achieves high accuracy on massive and well-known datasets like here we used the cityscapes dataset. And here we achieved the highest accuracy of 88%.

Graph2:



Graph3:

Network	Accuracy
Segnet (network we used)	88%
ResNet200	83.3%
Esnet	70.7%
Icnet	70.6%

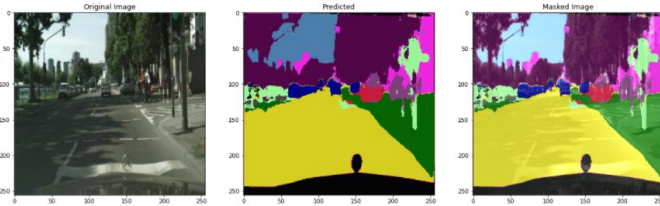


On comparing the above two graphs namely loss vs epoch and val_loss vs epoch graph. We can say that in graph1 the loss accuracy is decreasing and whereas in graph2 the val_loss accuracy first decreases up to some epoch and again partially increases and finally decreases.

Output images obtained for different input sample training images of cityscapes image-pair dataset are:

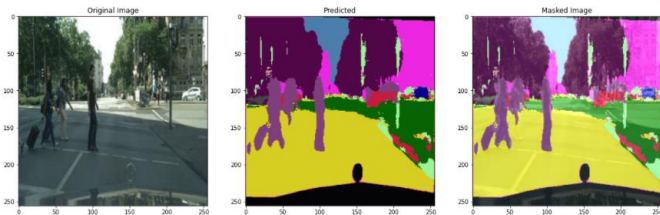
Case(i)

The below is the scenario where we have given the input image number 102 from the cityscapes-image-pairs dataset. And we can see below the original image, the predicted image, and the masked image that is also said to be the output image.



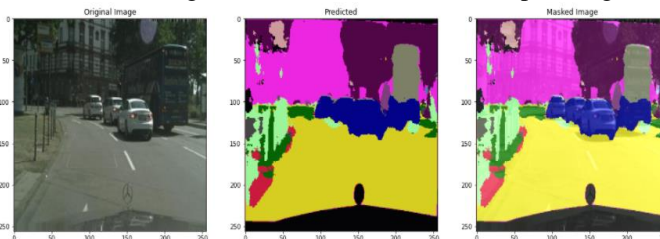
Case(ii)

The below is the scenario where we have given the input image number 116 from the cityscapes-image-pairs dataset. And we can see below the original image, the predicted image, and the masked image that is also said to be the output image.



Case(iii)

The below is the scenario where we have given the input image number 119 from the cityscapes-image-pairs dataset. And we can see below the original image, the predicted image, and the masked image that is also said to be the output image.



VII. CONCLUSION

In this paper, we focused on one of the foremost actively researched areas that are pixel-wise classification of the image in self-driving within the automotive industry. The dataset we used is from Kaggle. And One key ingredient of achieving it here is SegNet that is the utilization of max-pooling indices within the decoders to perform upsampling of low-resolution feature maps. This could result in Greater road safety, Reduction in deaths, More free time, Less Traffic, and mostly Environmental Benefits. and eventually, SegNet was the necessity to style efficient architecture in terms of memory, training time, and computational time. and finally, we have been able to solve and provide a good solution for one of the dense problems that are semantic image segmentation for self-driving cars.

VIII. REFERENCE

- [1] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation, CVPR, 2015.[4]
- [2] V. Badrinarayanan, A. Kendall, and R. Cipolla. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. arXiv preprint arXiv:1511.00561, 2015.[3,4,5]
- [3] A. Paszke, A. Chaurasia, S. Kim, and E. Culurciello. Enet: A deep neural network architecture for real-time semantic

- segmentation. arXiv preprint arXiv:1606.02147, 2016. [3,4]
- [4] M. Cordts, M. Omran, S. Ramos, T. Rehfeld, M. Enzweiler, R. Benenson, U. Franke, S. Roth, and B. Schiele. The cityscapes dataset for semantic urban scene understanding. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016. [6,7]
- [5] https://www.researchgate.net/publication/336638898_A_Survey_of_Deep_Learning_Techniques_for_Autonomous_Driving. [1,2,7]
- [6] The dataset: <https://www.kaggle.com/dansbecker/cityscapes-image-pairs>