Fully Convolutional Networks (FCN-GoogLeNet) for Road Segmentation

ELEN 239B_Fall 2018 Final Project Report

Yucheng Chen, Justin Huang

1. Abstract

We implement a FCN model with tensorflow and Python for road segmentation. We modify the official GoogLeNet model (InceptionV3) to FCN and trained with Kitti Road Dataset and Berkely Deep Road Dataset. The results are better than expected with 10 Epochs and 289 training data.

2. Introduction

Road segmentation is a fundamental task for self-driving cars as it can provide the drivable area for path planning. GoogleNet is an outstanding CNN architecture designed by researchers at Google. Fully convolutional neural network is an existing method to do the segmentation; however, the majority of FCN models we found are based on VGG16, we would like to leverage the advantage on the GoogLeNet and see if we can get a better performance. GoogLeNet is a lighter, wider, and deeper CNN models, so we expected we can have better performance with shorter training when integrating with FCN.

3. Methods

The traditional CNN includes convolution layer, pooling layers flatten layers and fully connected layer. The input pictures are convolved with the filters in the convolution layer and feature maps are generated as the output of convolution layer. The pooling layer can reduce the size of the feature map. The flatten layer and fully connected layer will generate the predicted class.

We have to modify the CNN to do the segmentation, the fully connection layer should be removed and an upsampling layer have to be connected to generate the pixel-wised classification. As Fig.2, upsampling layer do the transpose convolution on the pooling layer and generate the heat map having the same size as input picture. This kind of network called "Fully Convolutional Networks (FCN)."

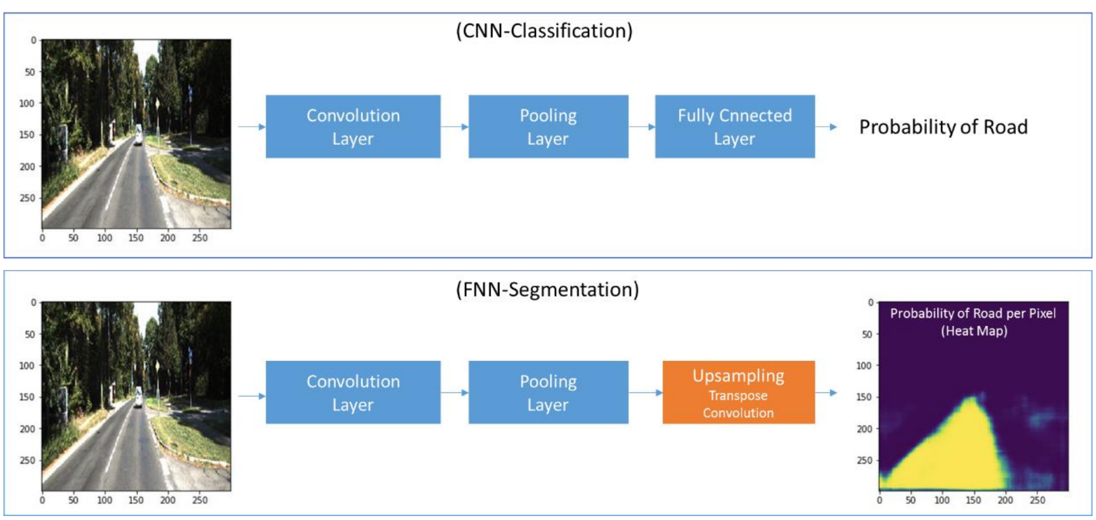

Fig.1 CNN-Classification and FCN-Segmentation

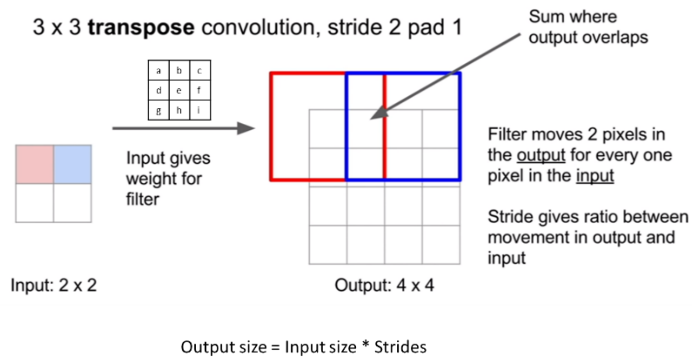

Fig.2 Transpose Convolution

In this project, we chose GoogLeNet as the core of FCN. As Fig.3, GoogLeNet has 22 layer, and almost 12x less parameters then Alexnet. GoogLeNet also have a better accuracy as it has some innovative “inception modules” in the network, which include different sizes of filters (1*1, 3*3, 5*5)to detect different size of features.

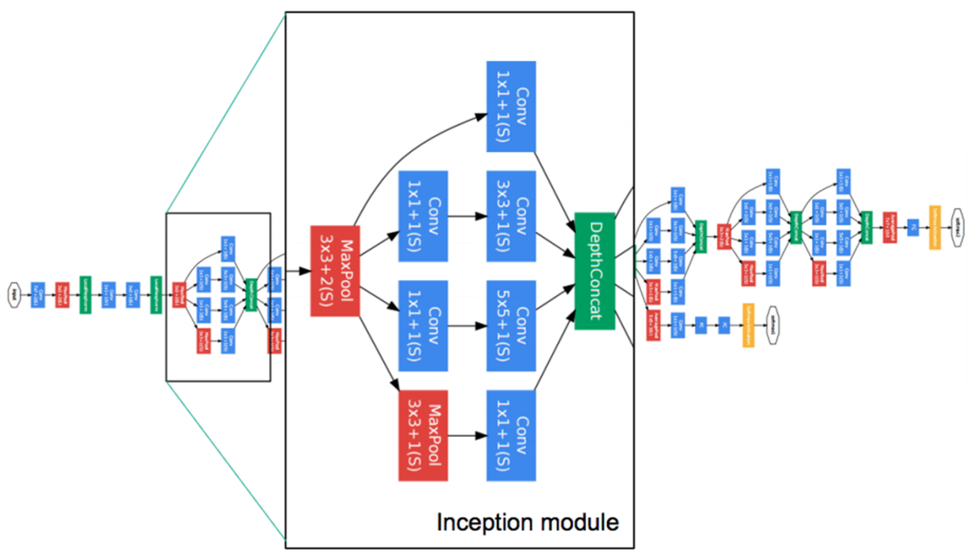

Fig.3 GoogLeNet Architecture (InceptionV1)

We implement the FCN-GoogLeNet architecture by Python and Tensorflow framework. As Fig.4, we get the tensor pool_3, mixed_7, and mixed_2 from the official InceptiopnV3 models. We upsample pool_3(1*1) to 17*17 and add with mixed_7(17*17). Then, we upsample and add a padding to size 35*35 and add to mixed_2. Finally, we use the same method to upsample to 299*299 and generate the heat map.

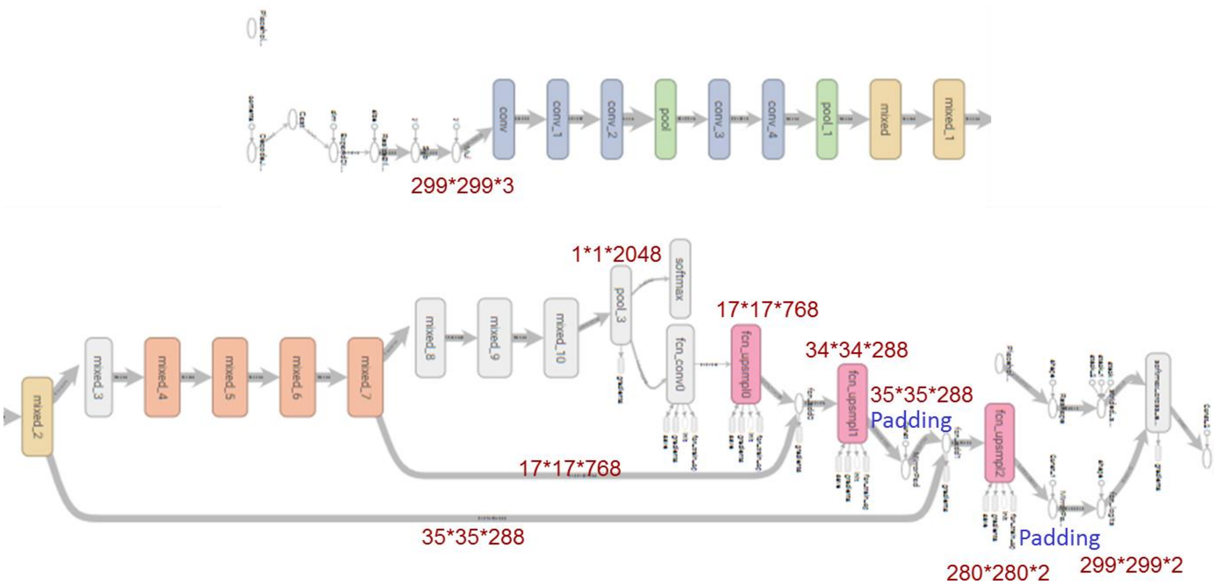

Fig4. Our model

We selected the Kitti from Honda Research Institute road dataset to train out model. This dataset is 289 training (with Label) and 290 test images (Fig.5). We also use Berkely Deep Road Dataset, about 7000 labeled training pictures, to enhance our model. (Fig.6)

Fig5. Kitti Road Dataset

Fig6. Berkely Deep Road Dataset

We need to do some image preprocessing as GoogLeNet only supports images with size 299*299. Therefore, we resize the picture to 299*299 before put into the model. Then we labeled the road on the resized pictures based on the probability on heat map. All pixels have P>0.5 will be marked. Finally, we resize the image to their original size. (Fig.7)

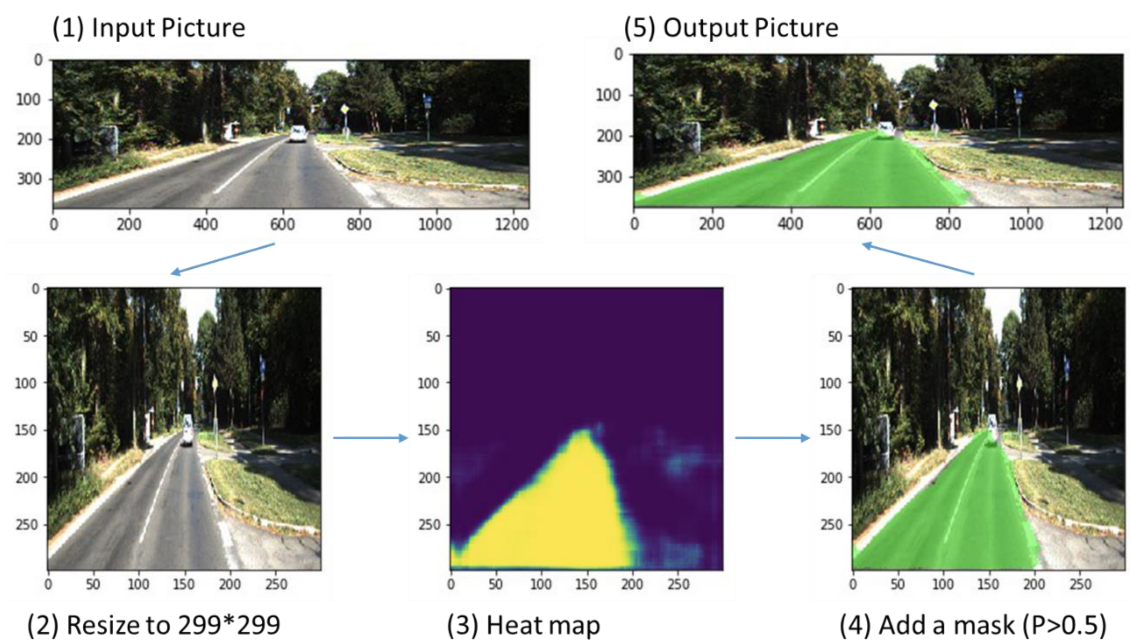

Fig.6 Preprocessing

4. Results
   We trained our model by 289 training data, with 0.0001 learning rate and 10 epochs. It takes about 3 hous on my laptop to train.

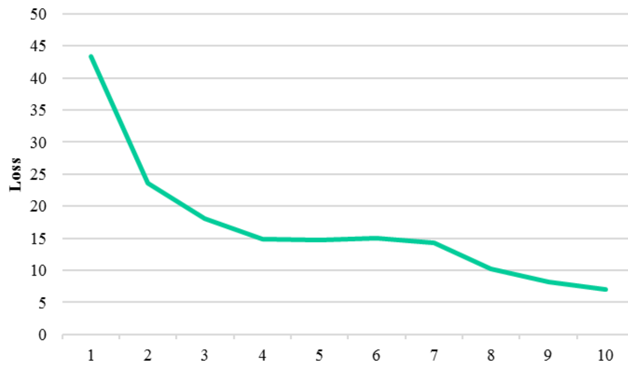

Fig.7 Epochs-loss diagram

Fig.8 Results

5. Conclusion

Our original FCN-GoogLeNet model trained by Kitti Dataset can correctly predict most of the scenarios in the testing dataset, especially the simple country road. However, it has poor accuracy on the prediction of pictures created by me. To resolve this issue, we applied a better dataset with about 7000 training data to our network and it works as Fig.9.

Train by Kitti Dataset
(289 Image; 10 Epochs)

Train by Berkeley Dataset
(~2000 Image; 2 Epochs)

Fig.9 Issue resolved by training with Berkeley Dataset

6. References

- Long, Jonathan, Evan Shelhamer, and Trevor Darrell. "Fully convolutional networks for semantic segmentation." *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2015.
- Fritsch, Jannik, Tobias Kuehnl, and Andreas Geiger. "A new performance measure and evaluation benchmark for road detection algorithms." 16th International IEEE Conference on Intelligent Transportation Systems (ITSC 2013). IEEE, 2013.
- Szegedy, Christian, et al. "Inception-v4, inception-resnet and the impact of residual connections on learning." AAAI. Vol. 4. 2017.
- https://github.com/darienmt/CarND-Semantic-Segmentation-P2
- http://bdd-data.berkeley.edu/
- http://www.cvlibs.net/datasets/kitti/eval_road.php