УДК 004.8

Neural Network Model Architecture for 3D Vehicle Detection and Segmentation

Али Аммар (Национальный исследовательский университет ИТМО)
научный руководитель–доцент, к.т.н, Кашевник А.М
(Национальный исследовательский университет ИТМО)

Abstract
In this research, we present a well-known problem in Computer vision (autonumous driving and driver assistant systems). From single monocular RGB image we designed a neural network to detect vehicles location, type and orientation. Our NN also provide an output mask for the bird view of a provided scene.

Introduction.

Driver assitant systems present a necessary development to increase safety factor since autonomous driving is still far from production. The ability to monitor the vehicle outside environment provides information that could help to avoid accidents and to build a communication system between vehicles. One of the most important key factors to build such systems is the ability to detect the vehicles in a current scene in the 3D space. Which is helpful to make 3D mapping, segmentation, modeling and motion prediction.

Main part.

We designed a neural network that is able to provide information about the 3D location and orientation of other vehicles and for each the type of the vechile. We also have 3D models for 79 different vechiles. using our Neural Network we are able to build a 3D map for a particular scene and the bird view for it.
Model architecture:
Efficient Net b3 with noisy student initial weights connected in parrallel with 3 double convolution blocks with upsampling layer the output is connected to 4 double convolution blocks with another upsampling layer and a 2D convolution layers as an output. The loss function used for training is formuled from 3 losses (Cross Entropy loss for the vehicle type, BCE with logits for the 2D mask (bird view) and mean absolute Error for the 3D location (x,y,z) and for the 3D orientation (roll, pitch ,yaw)).

Note: double convolution is a block consists of 2D convolution layers with batch normalization and relu activation for each.

Conclusion.

The proposed architecture and training pipeline showed a good results over the dataset provided from Peking University.