

ABSTRACT

Object detection and Depth map estimation are prominent subfields of computer vision which achieve state-of-the-art performance using deep learning algorithms. Networks with more layers have shown significant improvement in performance as they go deeper. These networks also require large datasets for training. Collecting large datasets and annotating them manually is a time-consuming task. An alternate approach is to generate datasets synthetically for training these networks.

In this thesis, we addressed the problem of Object detection and Depth map estimation using deep neural networks. In both cases, we observed that a network could be trained entirely on synthetic datasets to achieve promising results on real-world images.

The object detection problem was tackled with the DetectNet architecture. The factors affecting the performance (in mAP), such as the size of the training dataset, finetuning with selected layers, dictionary size of a 3D model, were observed. The result of the object detection network was compared with the available benchmark performance. The network has achieved a mean average precision of 52.59 at 0.5 IoU (Intersection over union).

The depth estimation problem was approached with the three model architecture based on AlexNet. The first stage contains the global context model and the gradient model which outputs the rough depth estimate and the gradient estimate,

respectively. The output of the first stage is then given as input to the second stage refining network to improve the rough depth estimate. The results of the depth estimation network show improvement in performance after finetuning the model.