

Visual Recognition Project : Depth Recognition in images

Divyanshu Shende (13807264), Rahul Tudu (13538)

April 28, 2017

1 The Problem

We took up the problem of estimating depth in an image. The architecture we intend to implement is from the NIPS 2014 paper Depth Map Prediction from a Single Image using a Multi-Scale Deep Network[DEPs] by David Eigen, Christian Puhrsch and Rob Fergus. We aimed at experimenting on their network. Their implementation is in Theano while ours uses Keras + Tensorflow.

2 The Dataset

The dataset is the labelled Kitti dataset. This contains the 3 dimensional image matrix and the ground truth is given as a heat map which can be converted to depth map.

3 The Network

The network used has two component stacks. These are called the coarse and fine layers. The input to both these layers is an 576×172 image with 3 RGB channels. The coarse layer outputs a $1/4$ resolution (of the original) depth. This corresponds to a coarse depth map. This is then combined with one of the outputs of the fine layers and then processed to obtain a finer depth map having resolution $1/4$ -th that of the original image.

4 Loss Function

The scale of the image affects the depth predictions which calls for a specialized loss. The training loss is a generalized version of this scale invariant loss and is given by $L(y, y') = \frac{1}{n} \sum d_i^2 - \frac{\lambda}{n^2} (\sum d_i)^2$. Here $d_i = (\log(y_i) - \log(y'_i))$ and $\lambda \in [0, 1]$.

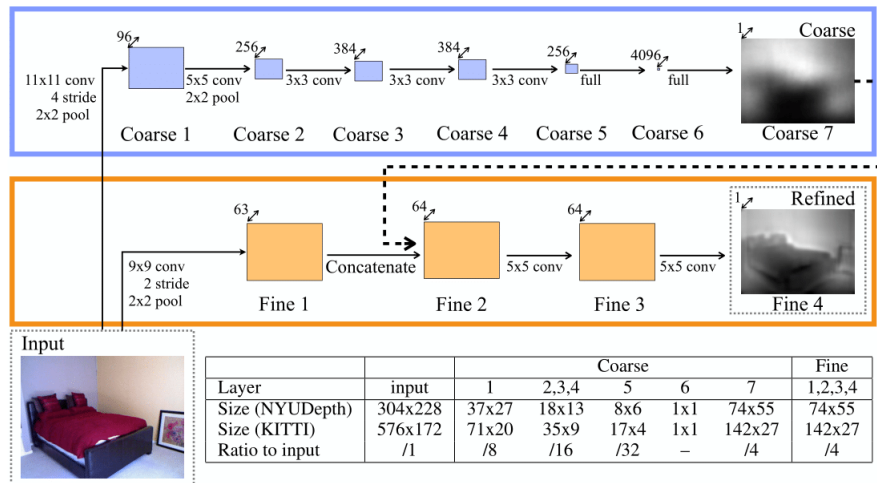


Figure 1: Model architecture.

Figure 1: The network