CS698O Project: Depth Prediction using single Image

Divyanshu Shende, Rahul Tudu
Indian Institute of Technology Kanpur

In this project we implement a NIPS 2014 paper by David Eigen, Christian Puhrsch, Rob Fergus. The paper is titled Depth Map Prediction from a Single Image using a Multi-Scale Deep Network.

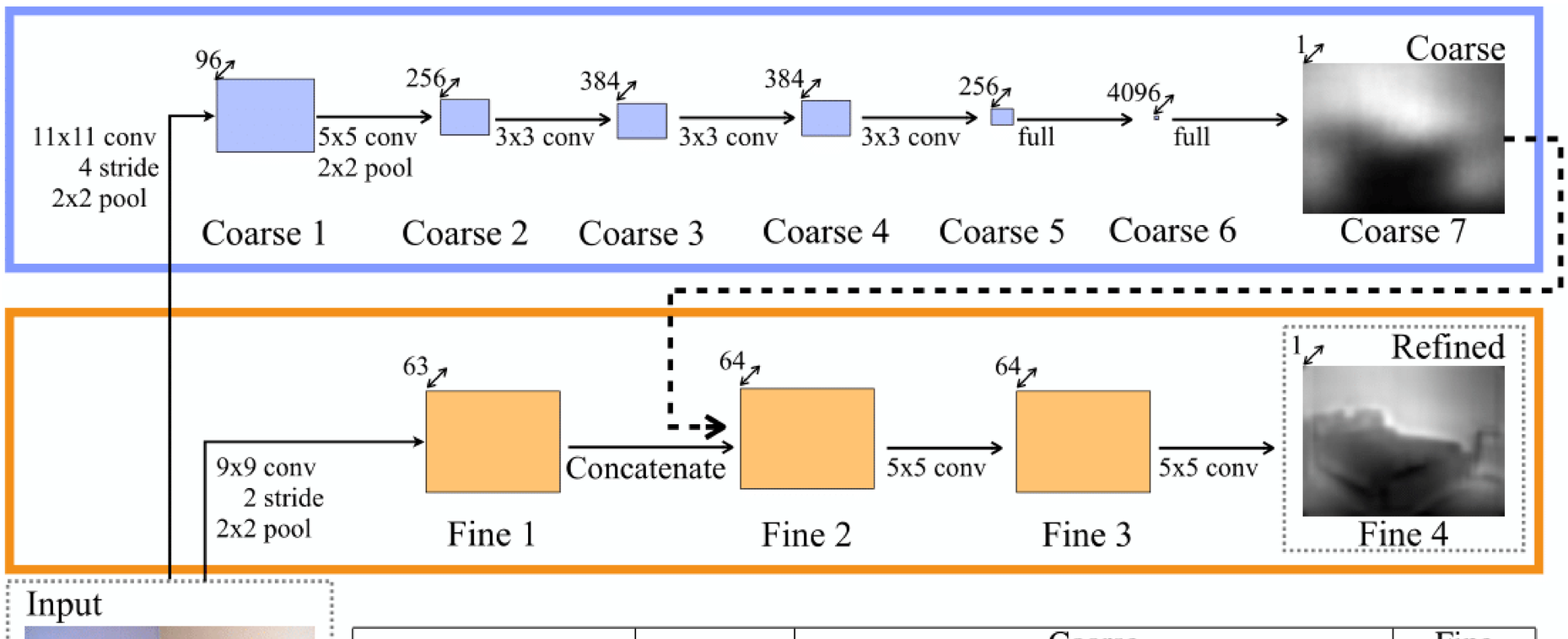
The Dataset

The dataset used by the us is the labelled NYU Depth v2 dataset. This contains the 3 dimensional image matrix and also a depth matrix which gives the depth (in metres) of each pixel from the camera. The dataset we run on is different from the one used in the paper due to memory constraints. The dataset used by the authors is 468 GB in size while the labelled dataset that we plan to use is just 2.8GB. We will split this dataset and use 80% for training and 20% for testing.

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 304x228 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.

The model architecture (from the paper)



		Coarse					Fine
Layer	input	1	2,3,4	5	6	7	1,2,3,4
Size (NYUDepth)	304x228	37x27	18x13	8x6	1x1	74x55	74x55
Size (KITTI)	576x172	71x20	35x9	17x4	1x1	142x27	142x27
Ratio to input	/1	/8	/16	/32	–	/4	/4

Figure 1: Model architecture.

Loss function

The scale of the image affects the depth predictions. For this reason, the scale invariant loss function is used by the authors. This scale invariant loss function is given by $D(y, y^{'}) = \frac{1}{2n}\sum_1^n (log(y_i) - log(y_i^{'}) + \alpha(y, y^{'}))^2$ where y' is the ground truth depth map (from the data). Here $\alpha(y, y')$ is the mean difference in the depth maps. 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 λ is a value in $[0, 1]$. We will use $\lambda = 0.5$ for our project. Note that putting $\lambda = 1$ reconstructs the scale invariant loss.

Implementation details

The implementation of the authors was in Theano. Our implementation uses Keras, which is a library on top of Tensorflow and Theano. In our implementation we have resized the original image to 304x228 as done in the paper.