27_March_Update

1. As reported yesterday, the time taken to complete 1 iteration for the Decoder.Upsample.ConvA layer was ~250 seconds. This amounted to ~30hrs for inference over a single image.

   So I am exploring with a smaller encoder network - **MobileNetv2** for the same Depth Estimation problem to be able to complete inference over layers on a practical time period. The implementation that I derived from is present here: https://github.com/alinstein/Depth_estimation.git

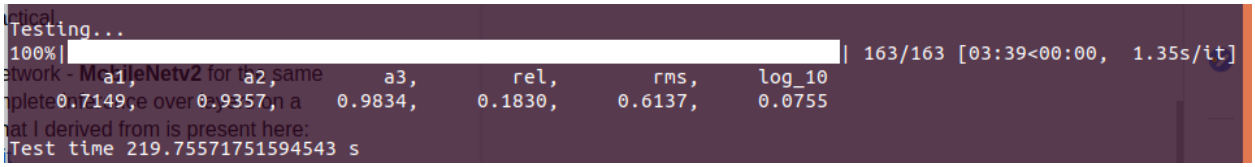
2. Trained the model (with MobileNetv2 as encoder) on 10,000 images from the NYU_Depthv2 dataset; 5 epochs; batch_size=16.

   The scripts for training and saving of model is available here: https://drive.google.com/drive/folders/1AJ8_IT8nGVJtBbLdTkpXp0IGHyInKpGj?usp=sharing

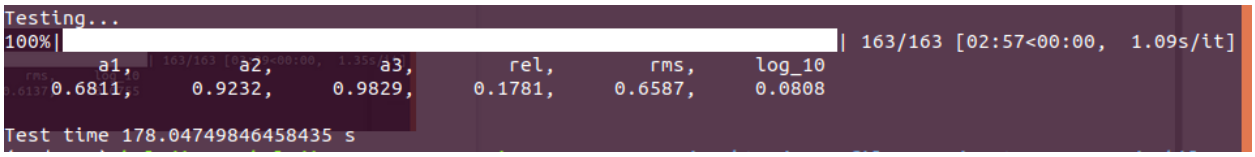
3. Profiled the time taken for 1 iteration in all the layers of decoder and proposed an estimated time for inference over 1 image. The readings are noted here: https://docs.google.com/spreadsheets/d/1tmXCuR8P1yGrYK8_bC07wBkrz_x6DhPl-a2OtFICZnw/edit?usp=sharing

4. The original implementation had decoder_width = 0.6, which I reduced to decoder_width = 0.4. This was done simply to try and reduce the inference time for 1 image. The overall error metrics of both the models has similar values.

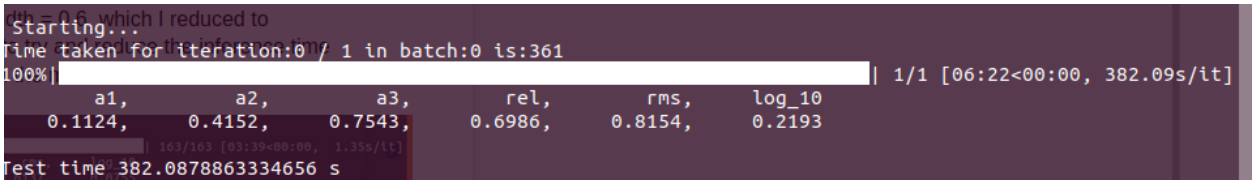
   (a) MobileNetv2 - decoder_width = 0.6

   (b) MobileNetv2 - decoder_width = 0.4

5. I ran inference over 1 image for the decoder.conv3 layer and got the following error metric values.



The error values are falling short behind the expected results.

6. The possible reasons for this could be the following:
    a. **Quantization**: I am exploring the possibility of converting the trained model.h5 file to an in8 quantized model.
    b. **Im2col - Custom conv func**: Checking if the use of im2col function in c++ implementation would improve the performance in any manner.
    c. **Decoder width and training set**: I have trained the model on 10K images for 5 epochs. Try out the possibility of training with the entire 50K NYUDepthv2 dataset to observe impact on performance.

While I am exploring these possibilities, it would be great if you can confirm if I'm heading in the right direction. If not, I request you to kindly give guidance for the same.