Interim Report

1. Details

Name: Davit Buniatyan

Initial Project title: Depth Reconstruction from Image Sequences using Recurrent Neural Networks

Current project title: Depth Reconstruction using Recurrent Neural Networks from single monocular images

Supervisor Name: Dr. Lourdes Agapito

2. Progress

2.1 The choice of framework

In order effectively experiment and train models in Deep Learning, the choice of framework is of absolute importance because: developing from scratch will require enormous amount of time on reinventing 'wheels' such as optimization methods, modular structure of neural networks that enhance the speed of experimentation, certainly the code will not be optimized and will most probably not applicable for GPU. In addition to that as currently my project includes a model (discussed in next chapter) that is considered to be on the edge of research and no current complete implementation has been found available. Hence, limitation of frameworks should have been thoroughly considered.

Most appropriate frameworks for our scope are Lasagne, Keras and Torch. I have started implementing the model in Lasagne and Keras. Both frameworks are based on Theano python library. As our model includes iteration over sequences of more than one dimension and Theano computional graph has only 1 dimensional iterator that is well optimized under GPU processing, extension of iterator scan() function has been attempted, but no successful results reached because of input, output and weights storing and accessing at each spatial time step. On the other hand, during implementation of the model, Torch's modular approach has been found most appropriate with only one limitation that unfolding recurrent structure for long sequences requires long memory processing (e.g. 1 hour for only building a model that supports 160x120 pixel image input without training). However several techniques has been found that will resolve the limitation such as segmenting the input image into superpixels, running training on a supercomputer given access and integrating Torch framework RNN.

2.2 Model & Implementation

Recurrent Neural Networks (RNN) haven't been widely used in computer vision tasks for sequence to sequence labeling because of their one-dimensional usage. Several models have been proposed including multi-dimensional RNNs by Alex Grave and Grid-LSTM by Nal Kalchbrenner, however as priory noted no current open source implementation has been found. I have examined several code bases that similarly or partially implemented Grid-LSTM based solely on 1D iterator and

extended it to support spatial domain. Currently I am testing the model for its effectiveness and efficiency and further on will expand to support n-dimensional input. The code and current progress could be found online in Github[1].

2.3 Literature review & Theoretical framework

Concurrent to my implementation I have completed literature review of state-of-the-art models for depth estimation from single or multiple images including any statistical/probabilistic and deep learning methods. I am currently writing about most notable models of Deep Learning that are possibly applicable for our problem and might in future be embedded into it by preserving recurrence. This includes convolutional neural network (CNN) as a feature descriptor layer. Furthermore, benchmarking parameters have been precisely defined for comparison and evaluation with state-of-the-art models.

3. Future

3.1 Model & Implementation

I am planning to finish the implementation model and testing. This includes finalizing implementation of training algorithm on multi-dimension and training on widely used benchmarking dataset MNIST by early February. Further on, construct final model, which either will use CNNs or superpixel segmentation as preprocessing layer and train on available datasets such as Make3D, NYU Depth v2 or KITTY and benchmark results. Upon successful completion of this stage we will expand the model to support spatiotemporal input including heavy optimizations that is estimated that will make the code 100 times faster and create synthesized dataset for training on image sequences, as no appropriate dataset is available.

3.2 Literature Review & Final report

I plan to finish literature review by the end of this week and start documentation of current model and implementation including benchmarking techniques that will support objective evaluation of the model to aim to have the final report by the end of March.

[1] https://github.com/davidbuniat/DepthRNN