PROJECT UPDATE: VISUALIZING CONVOLUTIONAL NEURAL NETWORKS

Amey Agrawal & Jaikumar Balani
{f2014148, f2014022}@pilani.bits-pilani.ac.in

1 INTRODUCTION

We are attempting to build a framework to visualise a CNN model, using multiple methods. Including deconvolution nets proposed by Zeiler & Fergus (2014) and activation based method proposed by Erhan et al. (2009).

2 PROGRESS SO FAR

The framework is built in python with Keras. Keras has been gaining popularity as deep learning framework due to its simplicity and flexibility. Keras can use both TensorFlow and Theano backends. We have developed an extensible driver module with command interactive line interface. It can take any keras model stored in hdf5 format as input. If the user does not provide an input model VGG16 checkpoint is used by default.

One of the two major methods of CNN visualisation has been implemented, where we maximise the average of norms of gradients in a given kernel using gradient ascent. A png of filters in the selected layer is generated as output.

3 WORK REMAINING

The deconvolution based algorithm will be implemented by 14th April. Deconvolution makes it possible to visualise particular neurones activated in given filter with the input. We want to develop a system where webcam feed can be used to see the output of deconvolution net in real time. The real-time nature of the input makes it challenging. t-SNE visualisations have been proven to be useful for visualizing high dimensional vector embeddings. For sake of completeness a t-SNE visualizer would be added to the toolbox to help visualize the final embeddings generated by the network.

REFERENCES

Erhan, Dumitru, Bengio, Yoshua, Courville, Aaron, and Vincent, Pascal. Visualizing higher-layer features of a deep network. *University of Montreal*, 1341:3, 2009.

Zeiler, Matthew D and Fergus, Rob. Visualizing and understanding convolutional networks. In *European conference on computer vision*, pp. 818–833. Springer, 2014.