ABSTRACT

Convolutional Neural Networks have shown excellent results on many visual classification tasks. However, in our usual classification problems, with the exception of ImageNet, the datasets are carefully crafted such that objects are well-aligned, located in the center of the whole image and at similar scales. Intuitively, classification gets more challenging as the amount of variation in the data increases, as the models have to learn to be invariant to certain changes in appearance. Sharing identical weights among adjacent units, in a certain level, allows CNN to be invariant to the transformations in the data. In this thesis, we investigate the change in model performance of CNN when it faces variation in the data that are unseen in training, such as rotating, scaling, shifting and adding random noises. We compare the difference between first convolutional layer’s filters from a model seen certain variation before and from a model does not, show how they evolved during training and compare the corresponding feature maps created. Via experimental approach, we explore the two possible ways of increasing model complexity to enhance the model performance—adding filters and adding layers. We make comparison of the efficiency of these two methods and validate the conclusion by another problem setting—image with noise.