

Building High-level Features Using Large Scale Unsupervised Learning

Anuar Maratkhan

I. SUMMARY

The focus of the research is to develop a feature learning network that will classify high-level unlabeled data using only unlabeled data. In particular, the paper investigates the possibility of building a face detector by training the model to detect such class-specific features on images without labels. Generally, the large datasets of images with labels like contains face or not are used for training face detectors. However, the work was inspired by neuroscience hypothesis that there exist so called "grandmother neuron" in the temporal cortex of human brain which is selective for faces, hands, and sometimes specific people. In other words, a baby, the neuron, learns to group faces to one cluster because it has seen many faces but not because of the reward or guidance. To build such class-specific detector a set of approaches such as restricted Boltzman machines (RBM), autoencoders, sparse coding, and K-means was used. The problem of recent studies is that these feature learning algorithms learn low-level features only (e.g. edges, blobs). The purpose of this study is to learn more complex features. The model training part includes 10 million 200x200 images generated from YouTube videos on a 9-layered network that has a billion of connections. The model parallelism and asynchronous stochastic gradient descent (SGD) was used for training on 1,000 machines with 16,000 cores during three days. As a result, the study revealed that it is possible to train high-level class-specific detector using unlabeled data. The accuracy of the model shows relatively high improvement in comparison to state-of-the-art baselines, K-means and deep autoencoders on a test set containing 37,000 images, and on ImageNet (16M images, 20K categories). The last one showed a leap of 70% improvement over the state-of-the-art.