Thesis Title: Investigation of Key Issues in Deep Learning Models for Visual Pattern Classification and Retrieval

Name: Sadaqat ur Rehman
Public e-mail: engr.sidkhan@gmail.com
Home institute address: Sarhad University
Host institute address: Tsinghua University

Abstract: This thesis is an investigation of unsupervised pre-train filter learning, training optimization and embedding of semantic information in deep learning models for visual pattern classification and retrieval. To this end, an unsupervised pre-training filter learning algorithm for CNN is proposed, called Convolution Sparse Filter Learning (CSFL) to obtain rich and discriminating features of an image. The main idea of the proposed CSFL method is that in this work, an effective and efficient algorithm is proposed, which only need the number of learning features. Different from the traditional methods, the proposed CSFL model tries to optimized the cost function –*l2* normalized features, despite of model construction for data distribution. Moreover, CSFL is able to leverage high dimensional inputs, and also has the potential to extract key features in the subsequent layers. The features extracted by CSFL algorithm are used to initialize the first CNN layer, and then these features are further used in feed forward manner by the CNN to learn high level features for classification. The linear regression classifier (SoftMax classifier) is used to serve as the output layer of CNN for providing the probability of an image class.

Leading to an optimized CNN, a modified resilient backpropagation (MRPROP) algorithm to improve the convergence and efficiency of CNN training is also designed. Particularly, a tolerant band is introduced to avoid network over training, and incorporated with the global best concept for weight updating criteria to allow training algorithm of CNN to optimize its weights more swiftly and precisely.

In addition, this thesis also contributes a new large-scale dataset for image and text retrieval, namely FB5k. The proposed dataset has 5130 image-feeling pairs randomly crawled from Facebook. To our knowledge, this is the first effort to collect a dataset of high-level concepts with small semantic gaps based on users' semantic descriptions i.e. image-feeling relationships. Furthermore, a novel multimodal approach is introduced by using Optical Character Recognition (OCR), explicit incorporation of high-level semantic information, and a new similarity measurement in the embedded space, which significantly overcomes the conventional Euclidean distance and improve retrieval performance.