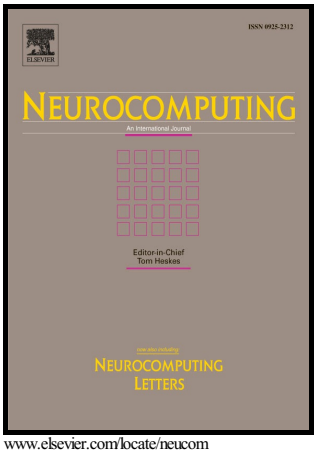

www.elsevier.com/locate/neucom

Combining Local and Global Hypotheses in Deep Neural Network for Multi-label Image Classification

Qinghua Yu, Jinjun Wang, Shizhou Zhang, Yihong Gong, Jizhong Zhao

PII: S0925-2312(16)31579-X
DOI: http://dx.doi.org/10.1016/j.neucom.2016.12.051
Reference: NEUCOM17882

To appear in: *Neurocomputing*

Received date: 7 August 2015
Revised date: 25 September 2016
Accepted date: 19 December 2016

Cite this article as: Qinghua Yu, Jinjun Wang, Shizhou Zhang, Yihong Gong and Jizhong Zhao, Combining Local and Global Hypotheses in Deep Neural Networ for Multi-label Image Classification, *Neurocomputing* http://dx.doi.org/10.1016/j.neucom.2016.12.051

Combining Local and Global Hypotheses in Deep Neural Network for Multi-label Image Classification

Qinghua Yu[a], Jinjun Wang[a,*], Shizhou Zhang[a], Yihong Gong[a], Jizhong Zhao[b]

[a]*Institute of Artificial Intelligence and Robotic, Xi'an Jiaotong University, Xian Ning West Road No.28 , Shaanxi, P.R. China, 710049*
[b]*School of Electronic & Information Engineering, Xi'an Jiaotong University*

Abstract

Multi-label image classification is a challenging problem in computer vision. Motivated by the recent development in image classification performance using Deep Neural Networks, in this work, we propose a flexible deep Convolutional Neural Network (CNN) framework, called Local-Global-CNN (LGC), to improve multi-label image classification performance. LGC consists of firstly a local level multi-label classifier which takes object segment hypotheses as inputs to a local CNN. The output results of these local hypotheses are aggregated together with max-pooling and then re-weighted to consider the label co-occurrence or interdependencies information by using a graphical model in the label space. LGC also utilizes a global CNN that is trained by multi-label images to directly predict the multiple labels from the input. The predictions of local and global level classifiers are finally fused together to obtain MAP estimation of the final multi-label prediction. The above LGC framework could benefit from a pre-train process with a large-scale single-label image dataset, e.g., ImageNet. Experimental results have shown that the proposed framework could achieves promising performance on Pascal VOC2007 and VOC2012 multi-label image dataset.

Keywords: Deep Learning, Convolutional Neural Network, Multi-label Classification

*Corresponding author