Regional Gating Neural Networks for Multi-label Image Classification

Rui-Wei Zhao[1]
rwzhao14@fudan.edu.cn
Jianguo Li[2]
jianguo.li@intel.com
Yurong Chen[2]
yurong.chen@intel.com
Jia-Ming Liu[3]
james.liu.n1@gmail.com
Yu-Gang Jiang[1]
ygj@fudan.edu.cn
Xiangyang Xue[1]
xyxue@fudan.edu.cn

[1] Shanghai Key Lab of Intelligent Information Processing,
School of Computer Science,
Fudan University,
Shanghai, China

[2] Intel Labs China
Beijing, China

[3] Department of Control Science and Engineering
Tongji University
Shanghai, China

In this paper we propose a novel deep learning framework named as regional gating neural networks (RGNN) for multi-label image classification. It mainly focuses on integrated contextual object region selection. The motivation arises from the fact that successful global CNN features ignore the underlying context information among different image objects. However, when people attempt to use information from objectness regions, current objectness region proposal algorithms usually produce too many irrelevant or even noisy regions as well. Thus it is meaningful to study how to effectively select useful contextual regions for image classification in the deep architecture.

The proposed RGNN is an end-to-end deep learning framework that can automatically select contextual region features with specially designed gate units, which are then fused for better classification. The feed-forward path of RGNN consists of 5 steps: (1) For each image, object proposals are used to generate multiple candidate regions. (2) Shared Conv + ROI pool + FC layers are then applied to obtain feature representations of regions. (3) Region/feature level gate units are imposed on each regional representation to control whether to be turned on/off so as to select useful contextual region features. (4) Multi-scale cross region pooling are further applied to get contextual image level feature representation. (5) Fused contextual representation are fed into FC layers to predict image labels.

The whole network is optimized with multi-label loss. When object level bounding box annotations are available, we further define a localization loss to aid effective region selection and optimize the network with multi-task learning. Because the gate units and the classifier are integrated in the same deep neural network pipeline, we can learn parameters of the network simultaneously.

Method	VOC'07	VOC'12
VGG-16+19 [1]	89.7	89.3
HCP-VGG [2]	90.9	90.5
HCP++ [2]	-	93.2
RGNN-RL	93.7	93.4
RGNN-FL	93.7	93.3

Table 1: Classification results (AP in %) comparison on VOC'07 and VOC'12 benchmarks.

We evaluate on PASCAL VOC 2007/2012 and MS-COCO benchmarks, and results show that RGNN is superior to existing state-of-the-art methods. Partial comparison results with state of the arts are displayed in Table 1. We can see from these results that our proposed networks with region level gate (RGNN-RL) and feature level gate (RGNN-FL) outperform the global VGG-16+19 networks. Compared to existing algorithms based on objects information like HCP-VGG and HCP++ (with multiple models fusion), RGNNs also work better thanks to the effective integrated contextual region selection in the deep networks. We also find that the introduced localization loss can effectively improve RGNN performances from our ablation studies on VOC 2007 data set.

[1] Simonyan et al. Very Deep Convolutional Networks for Large-Scale Image Recognition. *arXiv.org*, September 2014.

[2] Wei et al. HCP: A Flexible CNN Framework for Multi-label Image Classification. *IEEE TPAMI*, pages 1–8, 2015.