Second-order Attention Guided Convolutional Activations for Visual Recognition

Shannan Chen[1], Qian Wang[1], Qiule Sun[2,*], Bin Liu[3], Jianxin Zhang[1,4,*], Qiang Zhang[1,5]

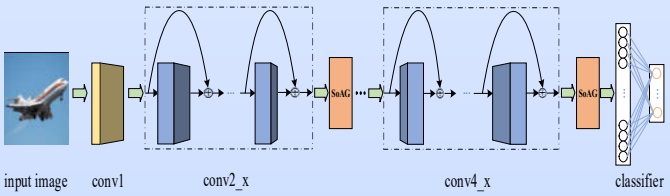
[1]Key Lab of Advanced Design and Intelligent Computing (Ministry of Education), Dalian University, Dalian, China
[2]School of Information and Communication Engineering, Dalian University of Technology, Dalian, China
[3]International School of Information Science and Engineering (DUT-RUISE), Dalian University of Technology, Dalian, 116622, China
[4]School of Computer Science and Engineering, Dalian Minzu University, Dalian, 116600, China
[5]School of Computer Science and Technology, Dalian University of Technology, Dalian, 116024, China
*Corresponding author: jxzhang0411@163.com, qiulesun@163.com

INTRODUCTION

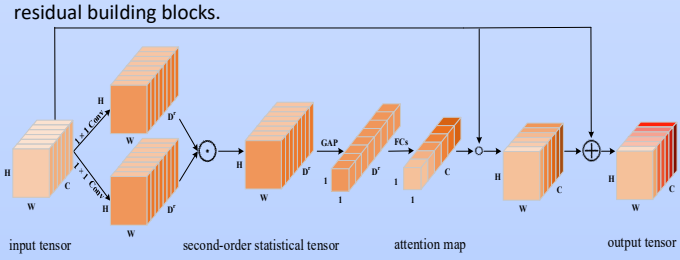
Recently, modeling deep convolutional activations by global second-order pooling has shown great advance on visual recognition tasks. However, most of the existing deep second-order statistical models mainly capture second-order statistics of activations extracted from the last convolutional layer as image representations connected with classifiers. They seldom introduce second-order statistics into earlier layers to better adapt to network topology, thus limiting the representational ability of deep convolutional Networks (ConvNets). To address this problem, this work makes an attempt to exploit the deeper potential of second-order statistics of activations to guide the learning of ConvNets and proposes a novel Second-order Attention Guided Network (SoAG-Net) for visual recognition.

METHOD

Ours paper proposes a second-order attention guided network (SoAG-Net) to make better use of second-order statistics and attention mechanism for visual recognition. Intuitively, thesecond-order statistics are collected across from lower to higher layers, and then used for predicting attention map, guidingthe learning of activations.



(a) Overview of SoAG-Net. The proposed second-order attention guidance (SoAG) module can be seemingly inserted into intermediate layers of ConvNet, forming our SoAG-Net. The global average pooling (GAP) after the last SoAG module is used for generating image representations fed into a classifier. Blue cuboids denote multiple consecutive convolutional layers. One blue cuboid and an identity shortcut connection form a residual building block. Each residual stage (e.g., conv2_x) has same number of residual building blocks.



(b) SoAG module. Given an input tensor, SoAG module captures the second-order statistics by element-wise product operator that receives two activation tensors outputted from two 1 × 1 convolutional (conv) layers performed on input tensor. The second-order statistical tensor is then passed through GAP and two fully connected layers (FCs) to predict attention map scaling input tensor along channel dimension. An identity shortcut connection is further employed to add input tensor to scaled one to boost information propagation.

Figure 1: The proposed second-order attention guided network (SoAG-Net). The overall of SoAG-Net is illustrated in Fig. 1(a) and its core SoAG module is given in Fig. 1(b). We take a step to exploit the potentials of second-order statistics of activations from earlier layers to guide the learning of deep ConvNets in a channel attention fashion. (Best viewed in color)

EXPERIMENTS

1. Effect of number of 1×1 convolutional filters:

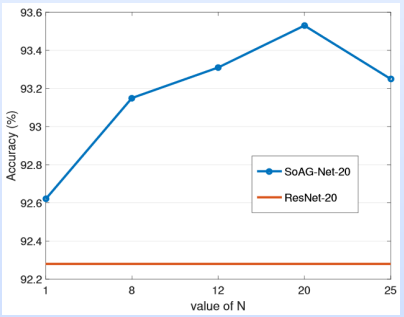


Figure 2: Accuracy rate with varied values of N. As a reference, the performanceof vanilla ResNet-20 is reported.

2. Comparison with state-of-the-art methods:

Method	Backbone	C10	C100	SVHN
ResNet [12]	ResNet-20	92.28	68.20	97.70
MS-SAR [23]		92.39	68.91	–
Online [25]		92.30	68.60	–
SW [24]		92.36	69.13	–
C3 [34]		–	69.34	–
SoRT [19]		92.65	68.35	97.74
SE [2]		92.63	69.07	–
RSoRT [26]		92.91	69.23	97.93
SoAG (Ours)		**93.53**	**72.72**	**98.07**
ResNet [12]	ResNet-32	93.17	69.72	97.46
MS-SAR [23]		93.32	70.39	–
SoRT [19]		93.67	70.39	97.78
SE [2]		93.33	71.55	–
RSoRT [26]		93.78	70.87	98.11
SoAG (Ours)		**94.07**	**72.87**	**98.29**
ResNet-56 [12]		93.70	71.75	97.51
SENet-56 [2]		94.52	72.91	98.15
ResNet-110-SW [24]		94.31	73.52	–
ResNet-110-C3 [34]		–	73.36	–
DenseNet-250-BC [29]		94.81	80.36	98.26
DenseNet-250-MCA [30]		94.62	76.22	98.34

Table 1: Comparison of accuracy(%) with state-of-the-art methods under both ResNet-20 and ResNet-32 back-bones. C10 and C100 refer to CIFAR-10 and CIFAR-100 dataset respectively. The comparison results show that SoAG-Net exhibits competitive performance and showcases its effectiveness.

CONCLUSION

We presented a novel second-order attention guided network(SoAG-Net), which contains conceptually simple yet effective SoAG modules conveniently plugged intoearlier residual stages of ConvNet. The SoAG module non-trivially guides the learning of convolutional activations byattention map computed from second-order statistics of acti-vations themselves. With such module throughout a network,SoAG-Net simultaneously gets stronger representation powerand provides more nonlinearity. The experimental results onvarious datasets manifest that SoAG-Net achieves good per-formance across different network depths.