先进设计与智能计算省部共建教育部重点实验室
Key Laboratory of Advanced Design and Intelligent Computing (Dalian University), Ministry of Education
（智能信息处理与网络技术重点实验室）

Second-order Attention Guided Convolutional Activations for Visual Recognition

Shannan Chen , Qian Wang, Qiule Sun,
Bin Liu, Jianxin Zhang and Qiang Zhang

Dalian University
Dalian Minzu University
Dalian University of Technology

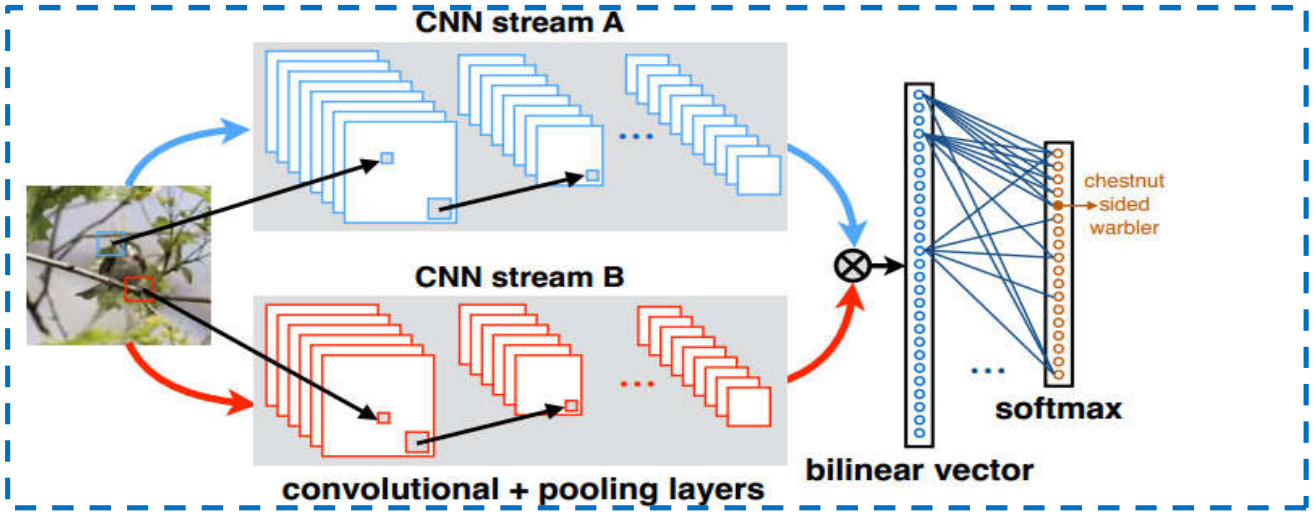
1th December, 2020

Contents

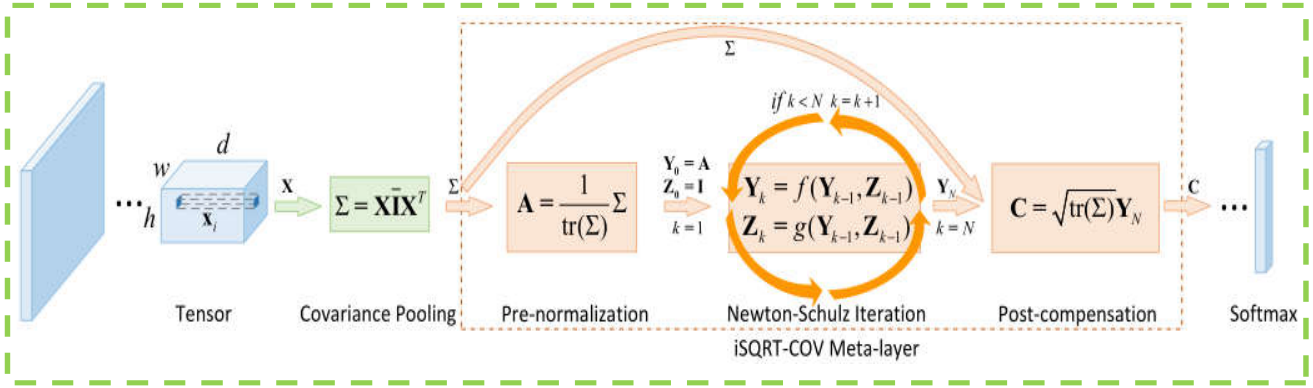
- Introduction
- Our methods
- Experiments
- Conclusions

• Introduction

B-CNN[1]



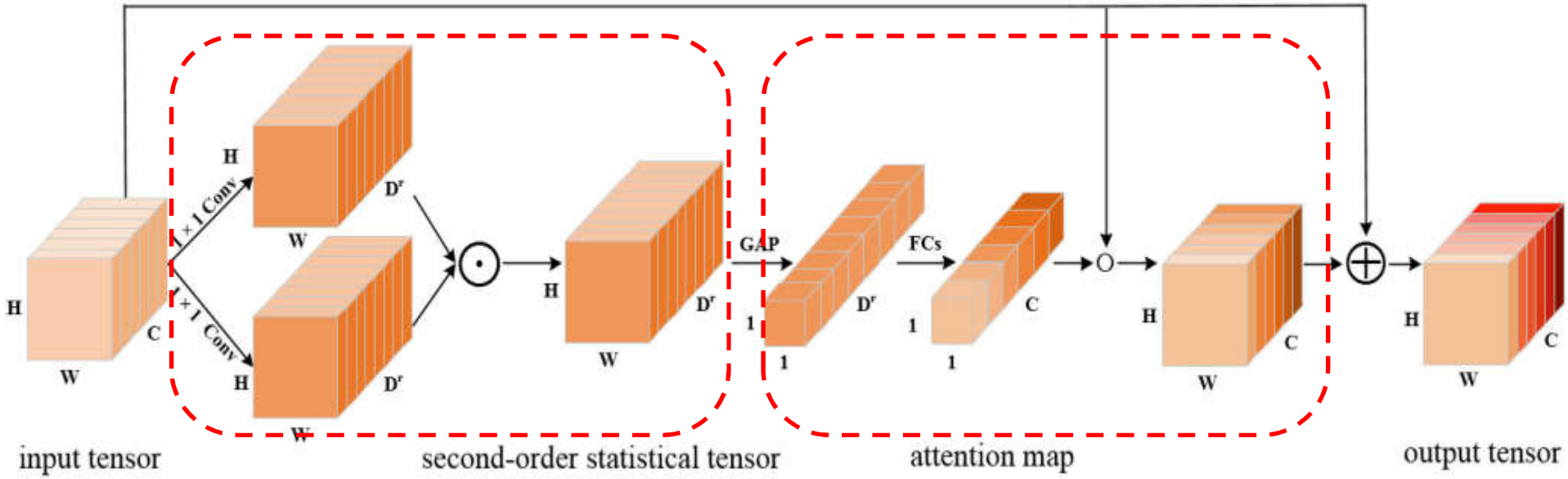
Fast-MPN[2]



---

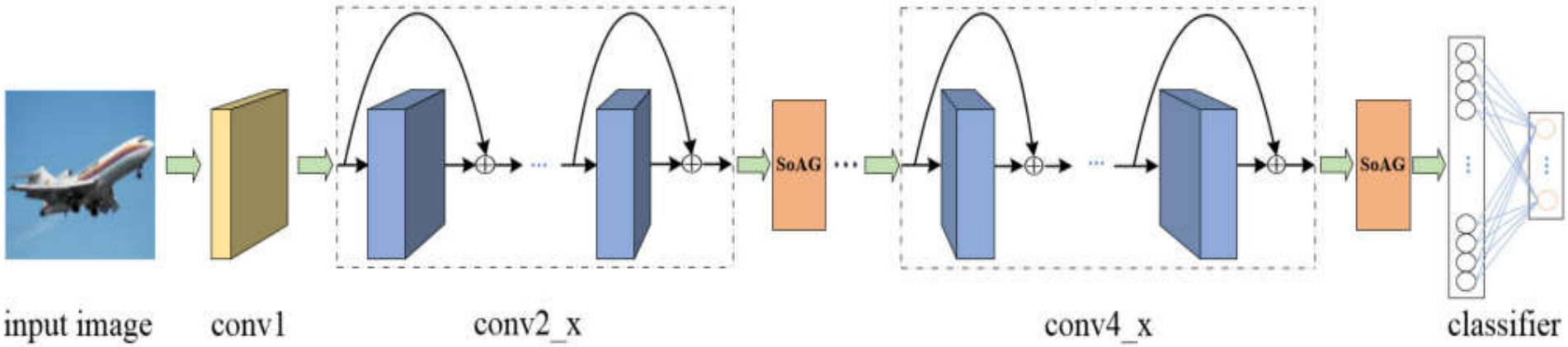
[1] Lin T Y, Roychowdhury A, Maji S. Bilinear CNN Models for Fine-Grained Visual Recognition. IEEE International Conference on Computer Vision. IEEE, 2016:1449-1457.
[2] Li, P., Xie, J., Wang, Q., Gao, Z., 2018. Towards faster training of global covariance pooling networks by iterative matrix square root normalization, in: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).

- Ours methods



- A novel Second-order Attention Guided Network (SoAG-Net) is proposed for visual recognition in an end-to-end.
- Ours model involves a group of SoAG modules seemingly inserted into intermediate layers of deep convolutional networks.
- Ours model captures second-order statistics of activations to predict attention map used for guiding the learning of networks

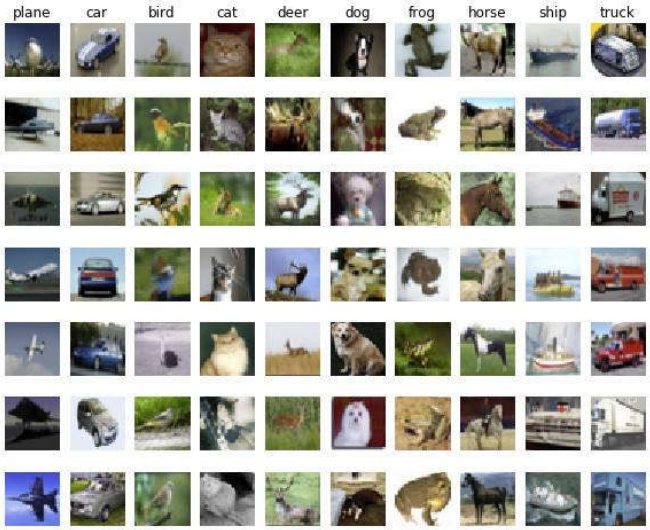
• Ours methods



- 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 ours module is used for generating image representations fed into a classifier.

• Experiments

1.Datasets



CIFAR-10/100

SVHN

- CIFAR-10/100: The CIFAR-10 and CIFAR-100 are single-label datasets containing 60,000 32×32 color images of 10 and 100 classes, respectively, both of which are split into 50,000 training images and 10,000 test images.
- SVHN: The Street View House Numbers dataset includes Street View images of 10 object categories with size of 32×32, containing 73,257 training images, 26,032 test images, and 531,131 extra training images.

• Experiments

2.Experimental results

Comparison of accuracy (%) with state-of-the-arts. C10 and C100 refer to CIFAR-10 and CIFAR-100 datasets respectively.

Method	Backbone	C10	C100	SVHN
ResNet [9]	ResNet-20	92.28	68.20	97.70
MS-SAR [7]		92.39	68.91	–
Online [3]		92.30	68.60	–
SW [25]		92.36	69.13	–
C3 [37]		–	69.34	–
SoRT [35]		92.65	68.35	97.74
SE [11]		92.63	69.07	–
RSoRT [39]		92.91	69.23	97.93
SoAG (Ours)		**93.53**	**72.72**	**98.07**
ResNet [9]	ResNet-32	93.17	69.72	97.46
MS-SAR [7]		93.32	70.39	–
SoRT [35]		93.67	70.39	97.78
SE [11]		93.33	71.55	–
RSoRT [39]		93.78	70.87	98.11
SoAG (Ours)		**94.07**	**72.87**	**98.29**

- The comparison results show that ours method exhibits competitive performance and showcases its effectiveness.

• Conclusion

- we presented a novel second-order attention guided network (SoAG-Net).
  - Contain conceptually simple yet effective SoAG modules conveniently plugged into earlier residual stages of ConvNet.
  - Ours module non-trivially guides the learning of convolutional activations by attention map computed from second-order statistics of activations themselves.

- In the future
  - Explore more discriminate high-order statistics of CNN features
  - Extend experiments on more datasets

Thank you !