Multi-Order Feature Statistical Model for Fine-Grained Visual Categorization

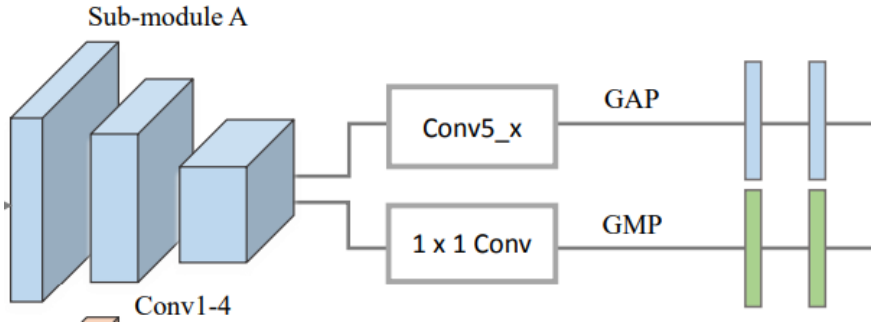
Qingtao Wang, Ke Zhang, Shaoli Huang, Lianbo Zhang, Jin Fan

School of Computer Science and Technology, Hangzhou Dianzi University
School of Computer Science, FEIT University of Sydney
School of Computer Science, FEIT University of Technology Sydney
{qingtao.wang, ke.zhang, fanjin}@hdu.edu.cn

Background

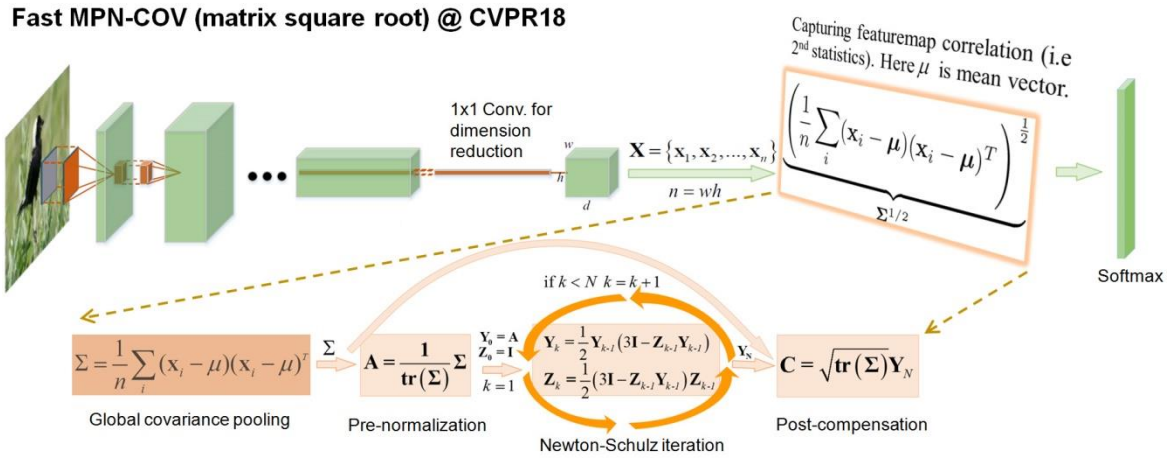
The key to address the problem of fine-grained categorization is to learn a discriminative representation that captures the subtle difference of similar class.

Existing methods generate high-level feature mostly by performing global **first-order** pooling, such as global average pooling (GAP), global max pooling (GMP).



The first-order module in proposed MOFS, using GMP and GAP.

Higher-order method

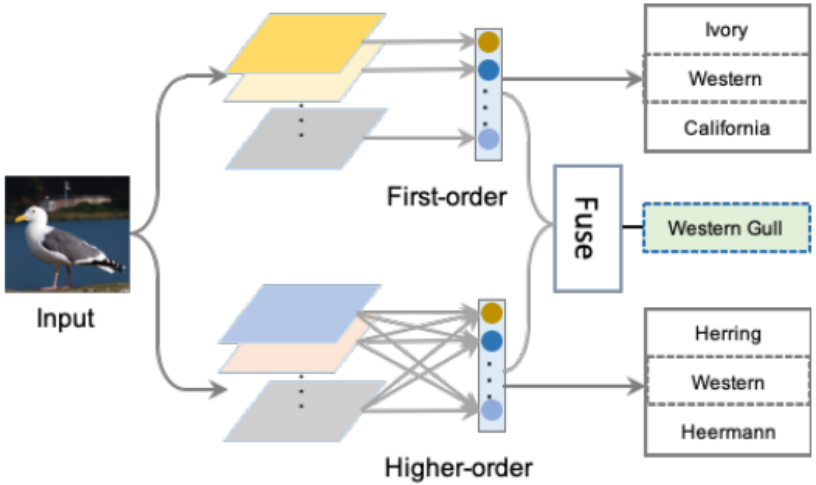


Li's method(iSQRT) obtains an impressive result on fine-grained datasets.
Our MOFS model adopts this method as the second-order extractor.

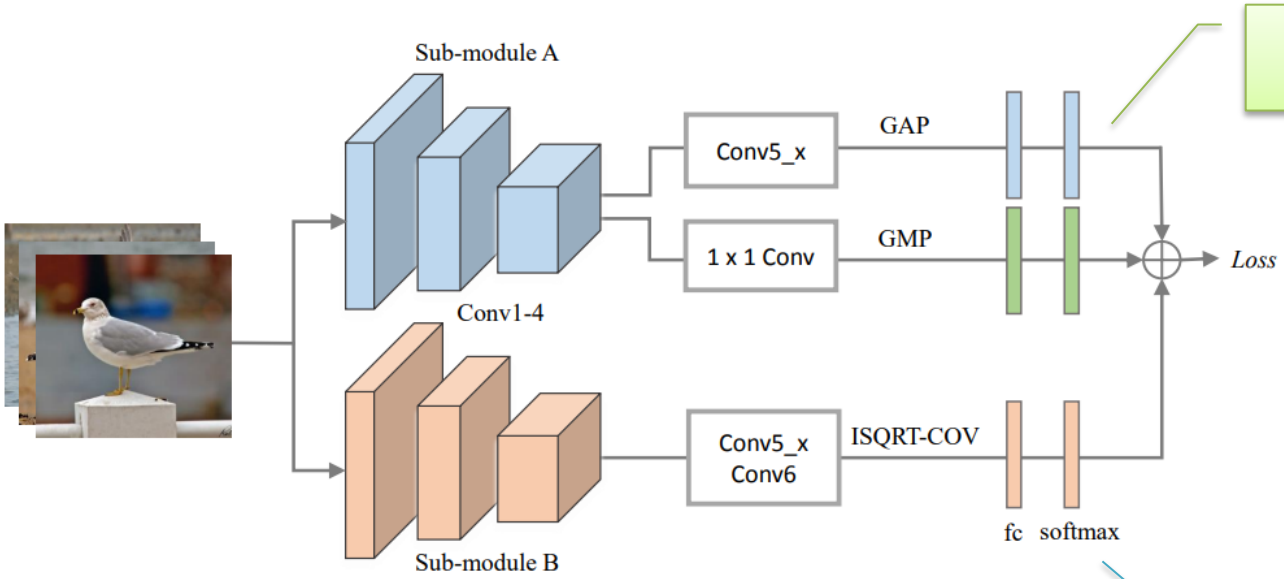
[1]Towards Faster Training of Global Covariance Pooling Networks by Iterative Matrix Square Root Normalization
Li, Peihua and Xie, Jiangtao and Wang, Qilong and Gao, Zilin.

Contribution

- We proposed a multi-order feature statistical method that integrates both first-order and covariance statistics to build strong representation.
- Our method consistently outperforms the state-of-the-art fine-grained method.



The Multi-Order Feature Statistical model



Sub-module A extracts high-level and mid-level first-order feature statistics by tow branches with GAP and GMP layers respectively. Sub-module B obtain second-order statistics following iSQRT-COV method

Results

- No part or bounding box annotations are used during training and testing.
- Compare to existing methods that only extract first-order or higher-order feature statistics, our approach with different order pooling layers obtain a higher accuracy on three datasets for fine-grained.

Method	Backbone	Accuracy(%)		
		CUB-200-211	FGVC-Aircraft	Stanford-Cars
VGG-19	VGG-19	77.8	-	84.9
ResNet-50	ResNet-50	85.4	90.3	91.7
ResNet-101	ResNet-101	86.8	-	91.9
RA-CNN[28]	VGG-19	85.3	88.2	92.5
MA-CNN[6]	VGG-19	86.5	89.9	91.5
B-CNN[16]	VGG-16	84.1	84.1	91.3
Compact B-CNN[17]	VGG-16	84.0	-	-
Low-ran B-CNN[18]	VGG-16	84.2	87.3	90.9
Kernel-Activation[19]	VGG-16	85.3	88.3	91.7
Kernel-Pooling[24]	VGG-16	86.2	86.9	92.4
MG-CNN[6]	VGG-19	82.6	86.6	-
RAM[29]	ResNet-50	86.0	-	-
MAMC[30]	ResNet-101	86.5	-	93.0
DFL-CNN[21]	ResNet-50	87.4	91.7	93.1
DFL-CNN[21]	VGG-16	87.4	92.0	93.8
NTS-Net[31]	ResNet-50	87.5	91.4	93.9
iSQRT-COV[22]	ResNet-50	88.1	90.0	92.8
iSQRT-COV[22]	ResNet-101	88.7	91.4	93.3
MOFS(ours)	ResNet-50	88.8	91.7	94.7
MOFS(ours)	ResNet-101	**89.2**	**93.0**	**94.9**

Ablation Study

For understanding the importance of each branch and model on the decision of interest, we draw the attention map by Grad-CAM method.

Method	Accuracy(%)		
	shared	separate	MOFS
GAP	80.7	85.1	85.7
GMP	76.8	78.7	81.1
GAP+GMP	83.2	84.3	86.2
ISQRT	87.8	88.0	88.0
GAP+GMP+ISQRT	87.7	87.9	**88.8**