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

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

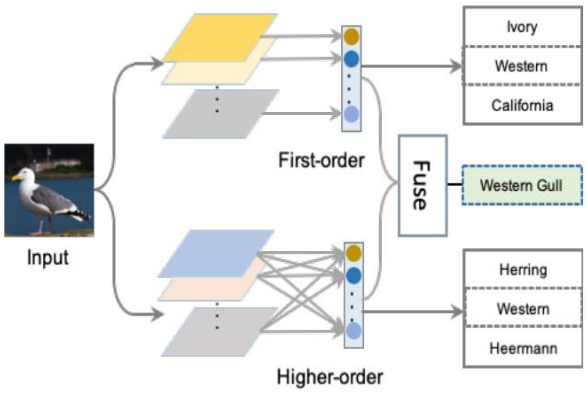
School of Computer Science and Technology, Hangzhou Dianzi University
School of Computer Science, FEIT University of Sydney
School of Computer Science, FEIT University of Technology Sydney

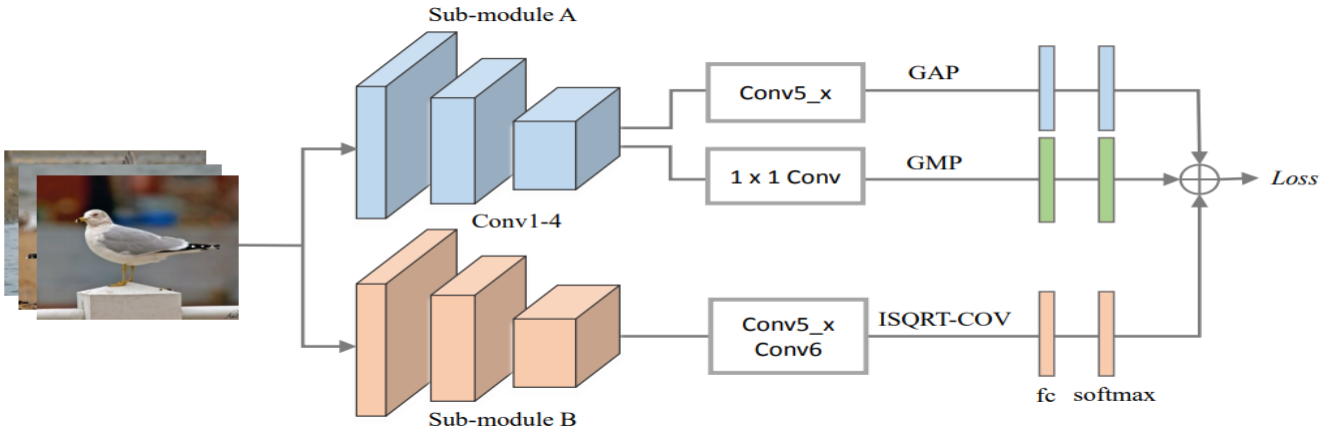
## BACKGROUND

Existing methods produce high-level features by performing first-order or second-order pooling to tackle the fine-grained categorization problem.

## CONTRIBUTION

we propose a multi-order feature statistical method (MOFS), which learns fine-grained features characterizing multiple orders.




| 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** |

## METHOD

The MOFS consists of two sub-modules: (i) a firstorder module modeling both mid-level and high-level features. (ii) a covariance feature statistical module capturing high-order features.

### Results

We evaluate the proposed method on CUB-200-2011, Stanford Cars, and FGVC-Aircraft. Compared with state-of-the-art methods, experiment results exhibit superior performance in recognizing fine-grained objects.