Universe Scientific Publishing

Multi-Attention Mechanism Fusion for Fine-Grained Image Classification

Rong Du, Dongmei Ma
Electronic and Information Engineering, Northwest Normal University, Lanzhou 730000, China.

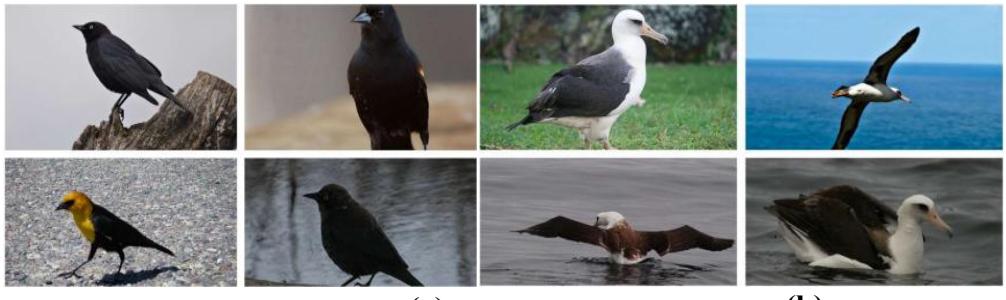
Abstract: In recent years, image classification has developed to the fine-grained level, which has become a new research hotspot. Compared with the traditional image classification task, the fine-grained image classification task has small difficulties due to the influence of image shooting scenes. So focus on mechanism has been widely used in fine-grained image classification problems, but the traditional attention focus on mechanism has the characteristics of first positioning and after processing, the model needs to run step by step and the attention focus on method is single. To further improve the performance of deep convolutional neural networks on the fine-grained image classification task, this paper studies the end-to-end weakly supervised fine-grained image classification model with multiple attention mechanism fusion. In this paper, a fine-grained image deep convolutional network model embedded with four attention focusing mechanisms, they are including: class activation mapping CAM attention focusing method, channel attention CA focusing method, spatial attention SA focusing method and channel spatial confusion attention and CSCA focusing method. On the fine-grained image classification dataset CUB-200-2011, Stanford-dogs, Stanford-cars, the results show that the four attention focusing methods can focus on local features and improve the performance of convolutional network classification performance, among which the channel spatial confusion attention focus method is the most significant improvement in the model classification performance.

Keywords: Fine-Grained Image Classification; Convolutional Neural Network; Attention Focus; Multi-Scale Learning

1. Introduction

The FGIC task of image classification mainly has the following problems: (1) the shooting angle, location, background, light, weather and other factors;(2)the small class and class differentiation caused by the identification of different subclasses in the same category;(3)some models need to be run step-by-step by positioning and after processing; (4) some models rely heavily on manual pre-annotation information. In order to build a model with high accuracy, end-to-end learning, and no manual annotation information, this paper needs to sort out the common problems of the above fine-grained image data .

This paper studies the content four attention embedding Multiple attention suggestions include four attention suggestion methods, Various attention mechanisms are CAM (Class Activation Mapping) local attention attention mechanism, CA (Channel Attention) attention mechanism, SA (Spatial Attention) attention mechanism, and CSCA (Channel-Spatial Confusion Attention) attention mechanism.



(a) **(b)**

Figure 1 (a) Small differences between classes of fine-grained images
(b) Huge differences between Within-class fine-grained images

2. Related Works

2.1 Experimental design

The ResNet-50 was used as the base network for the experiment. The proposed FCN and CAM attention attention modules, channel attention CA, spatial attention CA, and channel space confusion attention CSCA are embedded in the residual structure unit of ResNet-50.Respectively, we should build a deep FCN residual attention network (FCN-ResNet50), Deep channel attention residual attention network (CA-ResNet50), Deep space attention network (SA-ResNet50), deep channel space confusion attention residual attention network (CSCA-ResNet50). Finally, using the above four deep residual attention networks embedded with PyTorch through different attention attention mechanisms, the officially implemented original ResNet50 was tested in the CUB-200-200-2011 bird dataset, Stanford-dogs dog dataset, and Stanford-cars automotive dataset. The purpose is to verify whether the above four improvements are effective, focusing on improving the above classification performance of the baseline CNN network in response to the FGIC problem of the fine-grained image classification task.

2.2 Training tricks for ImageNet

Pretraining weight with Image enhancement.The present experiment uses a pretrained model from ImageNet as an initialization method for ResNet50, Where directly loaded are the pretrained weight parameters given from the PyTorch official ResNet50 implementation. Due to the use of the pretrained weight parameters, It can be said that the experiment of the ResNet50 benchmark network contains the idea of partial transfer learning , For the fine-grained image classification task, the FGIC commonly used fine-grained image classification dataset, such as the Stanford-dogs dog dataset, is itself a subdataset of the ImageNet. This allows the pretrained data from the ImageNet to contain much information that can be used for fine-grained feature identification. After moving the above weight parameters, It helps the network to extract the features of fine-grained images, And it can speed up the training speed, Make the network complete the convergence in a shorter time. In this experiment, the transforms tool under torchvision was used for random enhancement and weakening of brightness, contrast, saturation and tone using ColorJitter, and for random horizontal flipping of images using Random Horizontal Flip and random angle rotation of images using Random Rotation. Its purpose is to complete the image amplification, to increase the number of the original image through the above means of image preprocessing, and to play to the purpose of expanding the data set, so that the network requiring training can obtain more learnable information.On the other hand, in making the above random image amplification means learning the characteristics of the model will have stronger generalization ability, such as one, the original image in dark weather, after the brightness contrast saturation and tone adjustment, it may learn the performance of the same image in different weather scenarios, greatly enhance the robustness of the model and make the model after the change of application scenarios or data set.Therefore they still have good classification performance. Finally, the present model replaces the partially fully connected FC layer by introducing the FCN fully convolutional neural network, and the benefit is that the number of parameters can be quickly reduced or increased after using the FCN.Meanwhile the FCN can perform a pixel-level response to the image. The FCN with 1×1 convolution kernels can be considered as a pixel-level FC layer.

3. Results

Table1 Comparison of the four attention networks with the baseline networks

DataSet Accuracy(%)	ResNet50	FCN-ResNet50	CA-ResNet50	SA-ResNet50	CSCA-ResNet50
CUB-200	76.4	76.5	79.1	78.4	84.7
Stanford-dogs	73.9	73.8	77.3	77.1	83.5
Stanford-cars	87.6	87.1	89.7	90.2	92

4. Conclusion

The CSCA-ResNet50 network presented in this paper achieves good classification performance, with improvements of 8.3%, 9.6%, and 4.4% on the CUB-200-2011 dataset, Stanford-dogs dataset, and Stanford-cars dataset respectively.

References

[1] Berg T, Liu J, Lee S W, et al. Birdsnap: Large-Scale Fine-Grained Visual Categorization of Birds[A]. Computer Vision and Pattern Recognition[C]. Piscataway, NJ : IEEE, 2014: 2019-2026.

[2] Akata Z, Reed S, Walter D J, et al. Evaluation of output embeddings for fine-grained image classification[A]. Computer Vision and Pattern Recognition[C]. Piscataway, NJ : IEEE, 2015: 2927-2936.
[3] Wah C, Branson S, Welinder P, Perona P, Belongie S. The Caltech-UCSD Birds-200-2011 Dataset. [DB/OL]. (2011). Available from: http://www.vision.caltech. edu/visipedia/CUB-200-2011.html.
[4] Lin Z, Mu S, Huang F, et al. A Unified Matrix-Based Convolutional Neural Network for Fine Grained Image Classification of Wheat Leaf Diseases[J]. IEEE Access, 2019: 11570-11590.
[5] Sun Z, Yao Y, Wei X S, et al. Webly Supervised Fine-Grained Recognition: Benchmark Datasets and An Approach[C]//Proceedings of the IEEE/CVF International Conference on Computer Vision. 2021: 10602-10611.
[6] Zhang L, Huang S, Liu W. Intra-class Part Swapping for Fine-Grained Image Classification[C]//Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2021: 3209-3218.
[7] Wang Y, Morariu VI, Davis LS. Learning a Discriminative Filter Bank Within a CNN for Fine-Grained Recognition[C]// 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, 2018.