Bird Image Classification Competition

Cheng ZHANG
ENS Cachan
61 Avenue du Prsident Wilson 94235 Cachan CEDEX France
cheng.zhang@ens-paris-saclay.fr

Abstract

In this report, we present an approach using Convolutional Neural Network (CNN) for bird species classification problem. We compared the performance between training a self-defined simple four-layer CNN with fine-tuning pre-trained models on ImageNet [1]. We also applied data augmentation and ensemble learning methods to increase the accuracy. We tested the models on a subset of the Caltech-UCSD Birds-200-2011 bird dataset [8] and achieved an accuracy of 75.48% on the test set.

1. Introduction

Bird species identification is very important to maintain the biological diversity. As it is exhausting for experts to recognize bird species, this problem becomes a scientific and technical challenge for computer vision [5]. CNN has been proved very powerful for image recognition tasks [3] and in this report we focus on bird species classification by CNNs. The report is organized as follows: we present the whole approach in Section 2 and our experiments and results and some discussions in Section 3.

2. Approach Overview

First, we designed a simple convolutional neural network (SimpleNet) with the following network structure:

- Convolutional layer with 5 by 5 filters, 32 feature maps + 2 by 2 Max pooling + ReLU activation
- Convolutional layer with 5 by 5 filters, 64 feature maps + 2 by 2 Max pooling + ReLU activation
- Convolutional layer with 5 by 5 filters, 128 feature maps + 2 by 2 Max pooling + ReLU activation
- Convolutional layer with 3 by 3 filters, 256 feature maps + 2 by 2 Max pooling + ReLU activation

After each convolutional layer, we add Dropout and Batch Normalization layer to avoid overfitting and gradient saturation. In order to fit this network structure, the input images should be resized to 112×112. After convolutional layers, we add one fully connected layers with ReLU activation and the output layer.

As we do not have a large enough dataset, it is difficult to train a totally new deep neural network. So we also applied Transfer Learning [4] to train our model. We have tried ResNet model [2], VGGNet model [6], Inception model [7] and some other models. Among different variations of transfer learning, we also tried to fine-tune the whole pre-trained model, freeze several layers and freeze all feature layers. We also applied data augmentations (DA) to train on larger datasets. Finally, we applied an ensemble method to take the majority vote of different models' results as the final result, which slightly improved the accuracy.

3. Experiments and Results

Several models' results are listed in Table 1. We can see that all the pre-trained models achieve a greatly improved performance than SimpleNet. This is because of the simple structure and the lack of data. We can also see that data augmentation usually gives a better performance. ResNet model gives the best accuracy among all other models and ensembling all models slightly improved the performance.

	Accuracy (%)	
Method	without DA	with DA
SimpleNet	25.16	29.68
ResNet18	**72.26**	**74.19**
VGGNet	66.45	69.03
InceptionV3	67.74	70.32
Ensemble	75.48	

Table 1. Different models' results

We have also noticed some observations which are not listed in results, e.g. applying pre-trained models requires a small learning rate as we just need to slightly correct weights, varying learning rate is useful to make network continue converging, retraining the whole pre-trained model usually gives better performance than freezing layers.

References

[1] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei. ImageNet: A Large-Scale Hierarchical Image Database. In *CVPR09*, 2009.

[2] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[3] Y. Lecun, L. Bottou, Y. Bengio, and P. Haffner. Gradient-based learning applied to document recognition. *Proceedings of the IEEE*, 86(11):2278–2324, Nov 1998.

[4] S. J. Pan, Q. Yang, et al. A survey on transfer learning. *IEEE Transactions on knowledge and data engineering*, 22(10):1345–1359, 2010.

[5] C. Pang, H. Yao, and X. Sun. Discriminative features for bird species classification. In *Proceedings of International Conference on Internet Multimedia Computing and Service*, page 256. ACM, 2014.

[6] K. Simonyan and A. Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014.

[7] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich. Going deeper with convolutions. In *Computer Vision and Pattern Recognition (CVPR)*, 2015.

[8] C. Wah, S. Branson, P. Welinder, P. Perona, and S. Belongie. The Caltech-UCSD Birds-200-2011 Dataset. Technical Report CNS-TR-2011-001, California Institute of Technology, 2011.