Research on Small Sample Bird Recognition Based on B-CNN

Qiang Gao[a, *], Yutong Chen[b], Jiaxing Lei[c] and Xianyu Lu[d]

School of Airport Management, Guangzhou Civil Aviation College, Guangzhou, China

[a]281900404@qq.com, [b]yutungchan@qq.com, [c]104087066 @qq.com, [d]904824290@qq.com

*Corresponding author

Keywords: Bird Recognition, B-CNN, fine-grained image

Abstract: Bird recognition is a typical fine-grained image classification task. Bilinear-CNN (B-CNN) is a widely used fine-grained image recognition framework. Based on the Object Detection API in the TensorFlow framework, this paper combines the B-CNN algorithm, uses data set fusion to achieve data amplification, and then performs secondary training to solve the problem of small sample training difficulties. Finally, the high-risk airport is realized by B-CNN. Automatic recognition of bird images. The experimental results show that the image neural network detection model based on B-CNN can fully utilize the flexibility of TensorFlow and B-CNN, and effectively improve the stability, speed effectiveness and accuracy of image recognition. The average detection accuracy reaches 85.8. %.

1. Introduction

The problem of bird-attacking aircraft has always been a major problem that has plagued the world aviation community. It has not yet been fundamentally resolved. The International Air Transport Association has classified bird strike disasters as an "A"-level safety disaster in the aviation industry. The bird strike disaster has caused huge losses to the country and people's lives and property. In order to prevent bird strikes, various major methods of bird repellent are widely used in major airports, such as hunting, intimidation, sonic, optical, etc., but these methods do not distinguish birds, and there is instability and efficiency. Studies have shown that in order to achieve targeted bird repelling, different birds should use different methods of bird repelling, such as egrets, white-waisted Swifts and other high-altitude bird-repelling equipment such as titanium mines; night herons, grasshoppers, gray forest rafts, etc. Fireworks, tracer bullets, etc. are needed to stimulate their eyes [1]. Therefore, the realization of intelligent identification of birds and the corresponding driving method for intelligent selection of different birds can improve the efficiency and stability of bird repelling, and it is of great significance for reducing bird strike disasters and achieving safe operation of airports.

Convolutional Neural Networks (CNN) [2] is a special type of artificial neural network whose main features are Convolution Operators. CNN excels in many application fields, especially image-related tasks, such as image recognition [3], object detection [4], image semantic segmentation [5], image retrieval [6] and other computer vision problems. The so-called fine-grained image recognition is that the smart recognition objects often come from different sub-category categories of smaller granularity levels under the same category, and different types of images have certain similarities, such as different kinds of birds, dogs and cars. Therefore, bird recognition is a typical fine-grained image recognition problem.

Since the difference between fine-grained objects is only in the nuances, how to effectively identify the fine-grained-level objects in the foreground and discover important local area information becomes the key to solve the fine-grained image recognition algorithm. To this end, the researchers have proposed a number of novel solutions, such as PS-CNN [7], SWFV-CNN [8], BGL network [9], B-CNN [10], etc., among the above algorithms, B-CNN is one of the current highly accurate bird recognition algorithms and has been widely studied and applied.

2. B-CNN model

The B-CNN model is a bilinear convolutional network model proposed by Tsungyu Lin et al. in 2015. The model is shown in Figure 1.

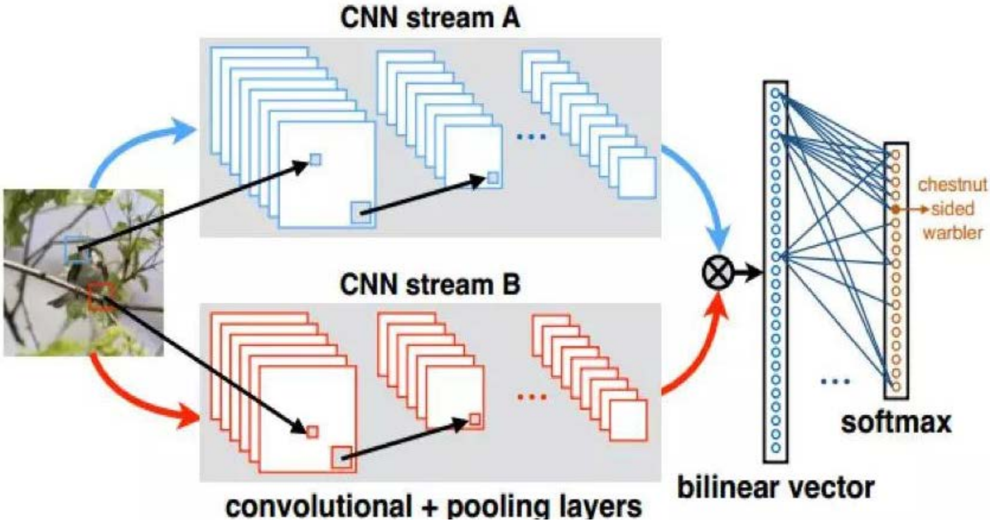

Figure 1. B-CNN model diagram

The intuitive interpretation of B-CNN is that network A is used to locate objects, and network B is used to extract features of objects detected by network A. The two networks coordinate with each other to accomplish two of the most important tasks in the fine-grained image classification process: area detection and feature extraction. The specific process is as follows:

(1) The input image is first processed to a size of 448 x 448, and then the features of the image are extracted using two neural networks A and B, respectively. At each position of the image, the two networks respectively generate features of 1 x 512 size, in each Position l performs an outer product operation on the features *A(l)* and *B(l)* extracted by the two networks. As shown in equation (1), the bilinear characteristic *X(l)* of this position is obtained, and the size is 512×512.

$$X(l) = A(l)^T B(l) \tag{1}$$

(2) Using the summation pooling method, the bilinear features obtained from all positions are summed as the characteristics of the image going to the image, and the size is also 512×512, as shown in equation (2).

$$X = \sum_l X(l) \tag{2}$$

(3) Next, the bilinear feature is subjected to square root and regularization operations. As shown in equations (3) and (4), the final feature Z of the image is obtained and used for classification by SVM or logistic regression.

$$\mathrm{Y} = \mathrm{sing}(X)\sqrt{|X|} \tag{3}$$

$$Z = \frac{Y}{\|Y\|_2} \tag{4}$$

The bilinear feature obtains better results in classification than the features extracted by a single convolutional network. The roles of the two convolutional neural networks are equivalent to region detection and feature extraction, respectively. Therefore, on the one hand, it avoids many manual marking operations in the conventional method, and at the same time, obtains a high accuracy.

Applying B-CNN to the actual scene is an end-to-end training process. The first half of the model is the basic convolutional neural network model, so only the gradient values of the second half are required to complete the training of the entire model. The known loss function gradient is dL/dX, and the gradient value of the loss function to the A and B network outputs can be obtained by the chain law, as shown in equation (5) and (6).

$$\frac{\mathrm{dL}}{\mathrm{d}A} = B\left(\frac{dL}{dX}\right)^{T} \tag{5}$$

$$\frac{\mathrm{dL}}{\mathrm{d}B} = A\left(\frac{dL}{dX}\right)^{T} \tag{6}$$

3. Experimental verification

3.1 Experimental platform and dataset

This experiment was performed on the Dell Precision Tower T7920 workstation using the python3.6+TensorFlow1.6 object API to implement the SSD algorithm. The workstation is configured as CPU: Intel Xeon Silver 4114, 10 cores and 20 threads, clocked at 2.2GHz; GPU: NVIDIA GeForce GTX 1080TI, memory 11G; memory: 64GB, operating system: Ubuntu 16.04 LTS (64-bit).

The data set used in this experiment is CUB-200-2011 and SOU-CN-BIRD. The CUB-200-2011 dataset has 200 types and 11788 bird pictures, of which the training set size is 5994 and the test set size is 5794. There are approximately 30 images in each class in the training set test set. The SOU-CN-BIRD data set is produced by the author. The bird images are from the high-risk birds commonly found in Guangzhou Baiyun Airport. They include 11 types of black-tailed tails, yellow tits, starlings, crows, white-waisted birds and long-tailed leafhoppers. 3254 images, the number of which is shown in Table 1. The data set is also divided into a training set and a test set. The training set size is 2277, and the test set size is 977. Some bird images in the data set are shown in Figure 2

Table.1. Name and numbers of birds in SOU-CN-BIRD dataset.

Birds’name	Numbers
Starling	369
White-waisted bird	402
Long-tailed leafhopper	311
Crow	374
siskin	298
Dicrurus macrocercus	388
Light-vented Bulbul	127
Masked Laughingthrush	232
Red-buttocked Quail	345
Tree lark	211
Turdus cardis	197

Figure 2. The birds’ image in SOU-CN-BIRD dataset

3.2 Experimental results

In order to solve the problem that the sample size is small and the model is over-fitting, the image in the SOU-CN-BIRD data set is first integrated into the CUB-200-2011 data set to realize the amplification of the data volume; The CUB-200-2011 dataset pre-trains the B-CNN, and finally used the SOU-CN-BIRD training set for secondary training and tests through the test set.

The training model used in this paper is based on the TensorFlow image recognition API published by Google, which includes the pre-trained B-CNN model. In order to solve the problem that the sample size is small and the model is over-fitting, the image in the SOU-CN-BIRD data set is first integrated into the CUB-200-2011 data set to realize the amplification of the data volume; The CUB-200-2011 data set retrains the B-CNN and uses the SOU-CN-BIRD training set for secondary training and testing through the test set. The entire operation flow is: loading of the library, setting of the environment, loading of B-CNN, preparation of the model, downloading the model, loading the model into memory, loading the data set, starting training, testing. The results of B-CNN identifying high-risk birds at airports are shown in Table 2.

Table.2. B-CNN identifies high-risk birds in the airport

Birds' name	Average Precision(AP)
Starling	90.1%
White-waisted bird	91.2%
Long-tailed leafhopper	89.7%
Crow	90.5%
siskin	85.9%
Dicrurus macrocercus	90.8%
Light-vented Bulbul	72.3%
Masked Laughingthrush	81.8%
Red-buttocked Quail	89.4%
Tree lark	83.4%
Turdus cardis	78.7%
Mean Average Precision(mAP)	85.8%

According to Table 2, it can be concluded that the recognition accuracy of B-CNN in the SOU-CN-BIRD data set reaches 85.8%, and the recognition accuracy is relatively high, but the recognition accuracy of birds with less data such as white-headed owl and black-ash worm is less. They are only 72.3% and 78.7% respectively, which are 13.5 and 7.1 percentage points lower than the average accuracy, which indicates that B-CNN is not robust to unbalanced samples.

4. Conclusion

This paper uses the TensorFlow API and B-CNN model to solve the problem of small sample data sets of high-risk birds in the airport through data set amplification and secondary training. Then, the bird identification of the high-risk bird data set in the airport is carried out. The B-CNN model can accurately achieve the bird identification, and the recognition accuracy reaches 85.8%. In addition, the robustness of B-CNN to unbalanced samples needs to be improved, which is also the direction of future research.

Acknowledgments

The authors would like to thank Science and Technology Program of Guangzhou, China (201904010226) and Young Innovative Talents Program in Universities and Colleges of Guangdong Province, China (2018GkQNCX080) and Special Funds for the Cultivation of Guangdong College Students' Scientific and Technological Innovation (“Climbing Program” Special Funds) (pdjh2019b0680).

References

[1] Liang Qiao, Zerong Shi. Investigation on the bird situation of Guangzhou Baiyun District Airport and the risk analysis of bird collision [J]. Science Technology and Engineering, 2014, 14 (03): 7-11.

[2] A. Krizhevsky, I. Sutskever, and G. E. Hinton. ImageNet classification with deep convolutional neural networks. In Advances in Neural Information Processing System,2012.

[3] S. Ren, K. He, R. Girshick, and J. Sun. Faster R-CNN: Towards real-time object detection with region proposal networks [J]. IEEE Transactions on Pattern Analysis and Machine Intelligence, 2017, 39 (6): 1137-1149.

[4] J. Long, E. Shelhamer, and T. Darrell. Fully Convolutional networks for semantic segmentation [C]. In proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp.3431-3440, 2015.

[5] A. Babenko and V.Lempitsky. Aggregating deep convolutional features for image retrieval [C]. In Proceedings of IEEE international Conference on Computer Vision, pp. 1269-1277, 2015.

[6] Nilsback M E, Zisserman A. Automated Flower Classification over many Classes [C] //In proceedings of IEEE Computer Society on Sixth Indian Conference on Computer Vision, Graphics & Image Processing, pp. 722-729, 2015 .

[7] Zhang X, Xiong H, Zhou W, et al. picking deep filter responses for fine-grained image recognition [C]. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1134-1142, 2016.

[8] Yili Zhao, Dan Xu. Fine-grained recognition of bird images based on joint semantic components [J]. Journal of Computer-Aided Design & Computer Graphics, 2018, 30 (08): 1522-1529.

[9] Xinye Li, Guangbiao Wang. Fine-grained bird recognition based on semantic detection of convolutional neural networks [J]. Science Technology and Engineering, 2018, 18 (10): 240-244.

[10] Lin T Y, Roy Chowdhury A, Maji S. Bilinear CNN models for fine-grained visual recognition [C] //Proceedings of the IEEE International Conference on Computer Vision. 2015: 1449-1457.