2020 13th International Symposium on Computational Intelligence and Design (ISCID) | 978-1-7281-8446-3/20/$31.00 ©2020 IEEE | DOI: 10.1109/ISCID51228.2020.00053

Dog Breed Classification Based on Deep Learning

Changqing Wang, Jiaxiang Wang, Quancheng Du, Xiangyu Yang
School of Electronic and Electrial Engineering
Henan Normal University
Xinxiang, China
wjx_0119@163.com

Abstract—**Deep learning is part of the field of artificial intelligence. It has powerful feature extraction and learning capabilities. Because of its various advantages, it has been applied in many fields. Object detection is an important technology in deep learning, and object detection based on deep learning has also been studied by many people. With the gradual improvement of people's living standards, pets have gradually received people's love, among which pet dogs occupy the majority. Different types of pet dogs will bring different problems. For example, large pets may be more aggressive and cause problems for city management. If dangerous pets can be distinguished in time, it can bring more security to people and avoid some people being bitten by pet dogs. In the deep learning algorithm, YOLOv3 has better object detection performance, but it only targets different species and objects, and the classification of different categories of specific species is not good enough. In daily life, the body of the pet dog is sometimes hidden by the complicated background, which makes it difficult to extract the overall characteristics of the pet dog. At this time, the facial features of the pet dog can be fully utilized to distinguish the pet dog. In order to solve this problem, this paper proposes an improved yolov3 model for face detection and categorization of pet dogs. This paper establishes a data set of 8 different kinds of pet dogs. The data set is divided into training set and a test set, and the training set is sent to the established model for training. Finally, we use the test set to verify the effect of the model. This paper establishes a data set of 8 different kinds of pet dogs. Pet dog types include Akita, Golden Retriever, Poodle, Pomeranian, Samoyed, Corgi, Pug, and Husky. Experiments show that this paper can realize the detection and classification of pet dogs with high detection speed and accuracy, and mAP(mean Average Precision) can reach 94.91%.**

Keywords-deep learning; yolov3; object detection; dog breed categorization

I. INTRODUCTION

Object detection is an important issue in the field of image processing and computer vision, which has attracted widespread attention and has become a research hotspot in recent years. Object detection has a wide range of applications, such as image classification[1], visual tracking[2], semantic segmentation[3] and so on. Object detection algorithms can be roughly divided into two types, one is traditional object detection, and the other is object detection algorithm based on deep learning. Traditional object detection algorithms require manual feature extraction, which is time-consuming. Using sliding windows to select candidate frames makes redundant windows too many, the accuracy is relatively low, and the real-time performance is not good enough. With people's continuous research, deep learning has been developed rapidly. The object detection algorithm based on deep learning has many advantages over traditional methods, and its detection speed and accuracy have been improved. It has slowly become the future development trend of target detection algorithms.

Convolutional Neural Network (CNN) is one of the representative algorithms of deep learning. Most of the current target detection is based on CNN, which is divided into one-stage and two-stage methods. Two-stage is divided into two steps. First, a sliding window is used to obtain the candidate area on the picture, and then a classifier is used to identify the target, such as R-CNN[4], Fast R-CNN[5], Faster R-CNN[6], etc. One-stage is based on classification and regression methods, such as SSD[7], YOLO[8], etc. YOLOv3[9] has the characteristics of fast speed and high precision, so it is widely used in various fields. After comparing these different detection algorithms, I decided to use the yolov3 model as the basis and combine it with the dog data set I made for training. After training, the position and type of the pet dog's face can be detected in the image.

II. YOLOv3 OBJECT DETECTION FRAMEWORK

The Yolov3 backbone network uses Darknet53. The main feature of this network is the use of residual network[10] and feature pyramid network[11]. The residual convolution is to perform 3x3 convolution with a step size of 2 and save it, and then perform a 1x1 convolution and one time 3x3 convolution, add these two results as the final output result. The residual network uses jump connections, which effectively solves the problem of gradient explosion and gradient disappearance during training. The residual network structure is shown in Figure 1.

YOLOv3 uses a feature pyramid network for prediction. The feature pyramid is used to extract the three feature layers of the backbone network. The three feature layers are located in the middle, lower and bottom layers. Some of these feature layers are used to output prediction results, and some are up-sampled and fused with other feature layers. The detection result will have three outputs of different scales to predict the position and category respectively, and obtain the preselected boxes of different scales. Finally, the non-maximum suppression algorithm[12] is used to screen the prediction boxes to obtain the final detection result. The feature pyramid network is shown in Figure 2.

2473-3547/20/$31.00 ©2020 IEEE
DOI 10.1109/ISCID51228.2020.00053

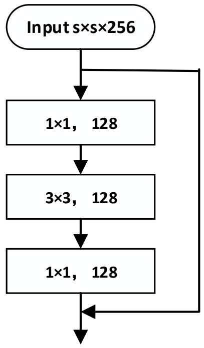


Figure 1. residual network structure

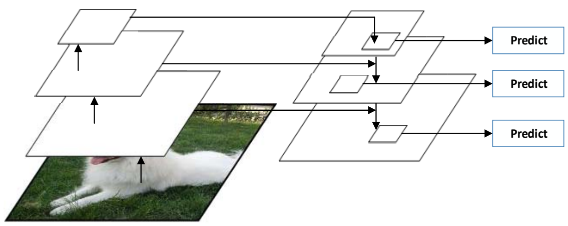


Figure 2. feature pyramid network

III. EXPEREMENTAL PROCESS AND RESULTS

The computer hardware configuration in this experiment is as follows: CPU is Intel(R) i5-7500 @ 3.40GHz and memory is 16GB. The GPU is NVIDIA RTX 2080 8GB. The operating system is windows 7, which depends on Python3.6.12, tensorflow-gpu1.13.2, keras2.1.5, numpy1.17.4, opencv-python4.4.0.42, etc. The neural network parameter epochs is set to 78, and the mAP calculation uses the voc2007 method.

A. Preparing Data Set

Part of the data set produced in this article comes from the Stanford Dogs Dataset[13-14], and the other part comes from pictures of some public websites. Stanford Dogs Dataset includes images of 120 dog breeds in different parts of the world, most of which are not in our country. Therefore, in order to produce a better data set suitable for our country's pet dogs, I have selected several Stanford Dogs Dataset as references. There are generally 8 kinds of pet dogs in our country, including Husky, Golden Retriever, Samoyed, Akita, Corgi, Pomeranian, Pug, and Poodle. I only selected these dog pictures from the Stanford dog data set and public websites as the data set, and then labeled these pictures with labelImg. 1280 images in the data set are used for training and 143 images are used for testing. Different dog training sets and test sets are shown in Table Ⅰ.

TABLE I. DATA SET DISTRIBUTION

Breeds of Dogs	Training Sets	Test Sets
Husky	172	15
Golden Retriever	134	16
Samoyed	194	24
Akita	105	12
Corgi	166	15
Pomeranian	195	24
Pug	179	21
Poodle	135	16
Total(Class 8)	**1280**	**143**

B. Experimental results

In the evaluation of the object detection model, the precision rate P and the recall rate R usually measure the important indicators of the model. The calculation formula is as follows:

$$P = \frac{TP}{TP + FP} \tag{1}$$

$$R = \frac{TP}{TP + FN} \tag{2}$$

TP means that a positive sample is tested and the test is correct. FP means that a positive sample was detected but the detection was wrong. FN indicates that it was originally a positive sample, but it was wrongly classified as a negative sample. P and R cannot be used to measure the quality of a model alone. This paper uses the area enclosed by the PR function and the coordinate axis to measure the performance of the trained model, and this area is mAP. Mean average precision represents the average accuracy of all detected targets. The mAP of a single target is AP. The calculation formula of AP and mAP is as follows:

$$AP = \frac{\sum \Pr ecision}{Num\left(Total \ \ \mathrm{Im}ages\right)} \tag{3}$$

$$mAP = \frac{\sum AP}{Num\left(Classes\right)} \tag{4}$$

The specific content of this paper is as follows: First, create a picture library of pet dogs, and select the faces of pet dogs of various categories through the labeling tool labelImg to generate your own data set. Then divide the data set into training set and test set. We input the images to be trained into the improved yolov3 network model for training. Finally we get our own model. This model can detect 8 types of pet dogs and frame the dog's face. The final detection accuracy of the model in this paper is shown in Figure 3. The results show that the accuracy of Akita and Corgi in our test set is 100%. The accuracy of Golden Retriever is 98.55%, and the accuracy of Husky is 94.12%. The detection results show that the method can achieve the face detection of pet dogs and the accuracy of Akita and Corgi have higher accuracy.

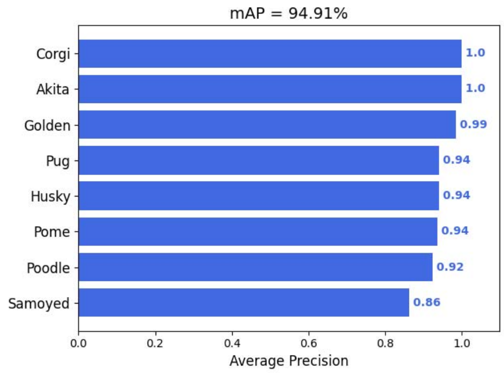


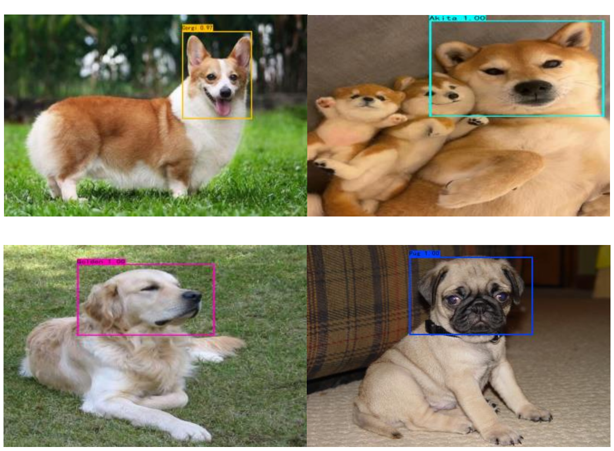
Figure 3. Detection accuracy of dog breeds

Some studies combine the appearance of dogs with geometric information to predict the type of dog[15-16]. The accuracy of our method is compared with others as shown in Table Ⅱ.

TABLE II. EXPERIMENTAL RESULTS COMPARISION

Method	Accuracy
Liu et al.*[15]	67.0%
Liu et al. [15]	77.2%
Parkhi et al. [16]	75.1%
Our method	94.91%

The detection results of some images are shown in the Figure 4. The experiment shows that the model established in this paper has a better effect of selecting the face of a pet dog, with fewer missed and false detections. For example, in the Akita dog in Figure 4, only the pet dog was detected, but the toy dog was not detected. The positioning and detection accuracy of the pet dog's face is very high.



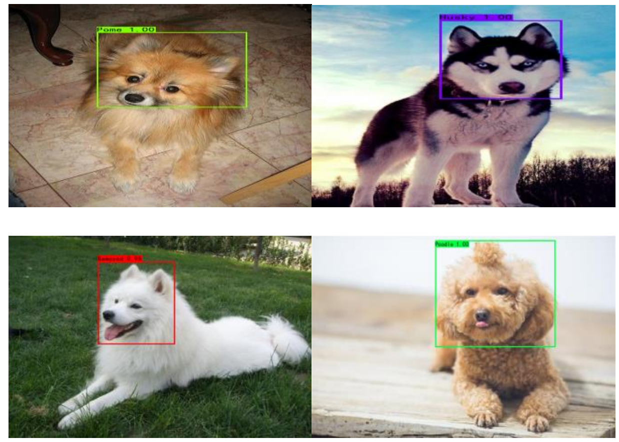


Figure 4. Detection results of dog breeds(Corgi, Akita, Husky, Golden Retriever, Pug, Pomeranian, Samoyed, Poodle)

IV. CONCLUSION

Since the current object detection generally detects different species, the research on object detection is rarely conducted on the subordinate classification of the same type of species. Aiming at this problem, we propose a pet dog classification model based on yolov3 using more common pet dogs as samples. The model does not use the overall characteristics of the pet dog, but by detecting facial features, just like face detection, select the face of the pet dog and mark the dog category. Experimental results show that this method can quickly and accurately locate the face position of a pet dog, and it solves the problem of pet dog detection and classification. By detecting the category of the pet dog, the pet dog can be better managed, which provides a good guiding significance for the management of the pet dog in today's society. In the follow-up work, the model proposed in this paper is used to track the dog in the video, so as to track the position of the dog in real time.

ACKNOWLEDGMENT

This work is supported by Key Laboratory Optoelectronic Sensing Integrated Application of Henan Province.

REFERENCES

[1] Zhang K, Guo Y, Wang X, et al. Channel-wise and feature-points reweights densenet for image classification[C]//2019 IEEE International Conference on Image Processing (ICIP). IEEE, 2019: 410-414.

[2] Hong S, You T, Kwak S, et al. Online tracking by learning discriminative saliency map with convolutional neural network[C]//International conference on machine learning. 2015: 597-606.

[3] Zhang Z, Cui Z, Xu C, et al. Pattern-affinitive propagation across depth, surface normal and semantic segmentation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019: 4106-4115.

[4] Girshick R, Donahue J, Darrell T, et al. Region-based convolutional networks for accurate object detection and segmentation[J]. IEEE transactions on pattern analysis and machine intelligence, 2015, 38(1): 142-158.

[5] Girshick R. Fast r-cnn[C]//Proceedings of the IEEE international conference on computer vision. 2015: 1440-1448.

[6] Ren S, He K, Girshick R, et al. Faster r-cnn: Towards real-time object detection with region proposal networks[C]//Advances in neural information processing systems. 2015: 91-99.

[7] Liu W, Anguelov D, Erhan D, et al. Ssd: Single shot multibox detector[C]//European conference on computer vision. Springer, Cham, 2016: 21-37.

[8] Redmon J, Divvala S, Girshick R, et al. You only look once: Unified, real-time object detection[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 779-788.

[9] Redmon J, Farhadi A. Yolov3: An incremental improvement[J]. arXiv preprint arXiv:1804.02767, 2018.

[10] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]//Proceedings of the IEEE conference on computer vision and pattern recognition. 2016: 770-778.

[11] Krizhevsky A, Sutskever I, Hinton G E. Imagenet classification with deep convolutional neural networks[C]//Advances in neural information processing systems. 2012: 1097-1105.

[12] Neubeck A, Van Gool L. Efficient non-maximum suppression[C]//18th International Conference on Pattern Recognition (ICPR'06). IEEE, 2006, 3: 850-855.

[13] Khosla A, Jayadevaprakash N, Yao B, et al. Novel dataset for fine-grained image categorization: Stanford dogs[C]//Proc. CVPR Workshop on Fine-Grained Visual Categorization (FGVC). 2011, 2(1).

[14] Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database[C]//2009 IEEE conference on computer vision and pattern recognition. Ieee, 2009: 248-255.

[15] Liu J, Kanazawa A, Jacobs D, et al. Dog breed classification using part localization[C]//European conference on computer vision. Springer, Berlin, Heidelberg, 2012: 172-185.

[16] Parkhi O M, Vedaldi A, Zisserman A, et al. Cats and dogs[C]//2012 IEEE conference on computer vision and pattern recognition. IEEE, 2012: 3498-3505.