DOI: 10.54254/2755-2721/5/20230594

Recognition of feline images based on deep learning

Zhuchang Gu

Department of computer science and technology, Wenzhou-Kean University, Wenzhou, Zhejiang, 325000, China

guzhu@kean.edu

Abstract. As the study of felines continues, image recognition technology is helping researchers identify species more and more. However, the research on cat image recognition based on deep learning is not deep enough in the past, which leads to the low efficiency of researchers and the inability to achieve the purpose of accurate recognition. According to researchers, there are more than 500 million pet cats in the world, of about 40 breeds. Predictably, the popularity of cats has led to an increasing amount of research on cats. On the basis of the existing background, this paper uses the deep learning network based on the PyTorch framework to study this topic and conducts in-depth exploration which finally has high efficiency and accuracy. It is hoped that this study can improve the research efficiency of related researchers.

Keywords: Images recognition, Deep learning, Feline.

1. Introduction

In this 21st, it is the time of information era, individuals and organizations use computer to work and they are common to use the emails to chat with each other's and share knowledge or their wish or discuss about business [1]. However, Spam is harassing our lives with this situation. Spam is a kind of email which be delivered to a large number of recipients without ask the wish of receiver [2]. Spam is waste our time to read useless information and occupation our email address and network flow. The problem of spam is become more and more important which has been increasing over years. In recent research, 40% of emails are spam.15.4 million spam are sent in every day which cost user about 355 million dollar per year [3]. In order to get rid of spam, using machine learning's method to use some models so computer can distinguish spam automatic to prevent spam into email. Machine learning is a branch of computational algorithms which is aimed to designed to simulate human's mind by learning from the nearby environment and predict the results [4]. Machine learning is also the core of Artificial Intelligence (AI). Nowadays, lots of people are searching on this subject and successful design some effective model. However, Chinese and English have lots of different in orthography, syntax, semantics, and phonetics are different, so the method of English language spam is hardly achieved in Chinese language [5]. So, it is important to design a model to identify Chinese language model. There are many methods have been proposed to handle spam by using machine learning. The methods of identify spam can be distinguished into two groups. One is called static methods and the other is called as dynamic methods. Static methods record the email address to identify spam and dynamic methods are consider about the content of email to identity normal email or spam [1]. In this essay, dynamic

DOI: 10.54254/2755-2721/5/20230594

methods are used to identify spam. After studying the relevant background and consulting materials, the author finds that the traditional classification of felines is identified, classified, and counted by professionals according to their shape, size, color, and other characteristics. The heavy workload of manual identification is easy to make the operator tired and affects the accuracy of identification and statistics, which seriously affects the efficiency and reliability of the investigation of felines. At present, most of the feline biological image processing software used by relevant researchers at home and abroad can only perform simple geometric measurements and counting of plankton images, and the specific species identification and classification still need manual intervention, which brings trouble to researchers without sufficient knowledge reserve [3]. In recent years, digital image processing technology is applied to implement the cat as a biological image automatic recognition and classification has aroused people's concern, but no scholar working in this field was carried out, so this area is worth us to study based on the depth and in learning, thus, to make up the gap.

The topic of this chapter is to analyze and recognize felis based on deep learning, and the method uses the PyTorch framework, ResNet network, and classification algorithm for the construction and training of neural networks [4].

Based on deep learning, to carry out image recognition of feline animals. Firstly, the related image recognition algorithms, models, databases, and other necessary data before the experiment are theoretically described and discussed. Secondly, based on the deep learning framework, the relevant training of feline cat images was carried out, and then the experiment was completed.

This paper uses Python software to carry out the image recognition of felines, mainly using the model analysis method

First need to make the model of learning needs of cat class image data collection, after the cats have cat-like animals which are data set, the need to establish a relevant learning model, then is to classify the relevant data sets, the division of data set follows the general principle, namely the ratio of eight binary, then the author needs to import data set, After the data set is imported, there will be the corresponding prediction results, and then the prediction results will be compared by the loss function to update the parameters, so as to make them fit the model. After fitting the model, the established model can be used for image recognition.

2. Algorithm Overview

2.1. Artificial Neural Network (ANN)

Since the emergence of artificial neural networks, more and more researchers have begun to pay attention to this field, and it is very hot in the research and application of artificial intelligence. Its concept is to simulate the processing of received information outside the human brain and then construct the network composed of nerves and neurons in the brain. and the relevant model is established. In academic circles, researchers refer to them as neural networks. A neural network is a model, and just as the human brain is made up of nerves, the model is made up of closely interconnected neurons. Each neuron can also be called a node, and a nerve is composed of connections between nodes, because the signal weights in the connections are equivalent to the memory of Ann. The output of the network changes due to many factors, such as changes in connectivity and activation functions.

In previous studies, the research of neural networks has been improved and made great progress. In recognition of images, robotics, economic forecasting, medicine, economy, and other aspects, especially the deep learning application of neural networks have become the focus of Internet data and artificial intelligence.

2.2. Convolutional neural networks (CNN)

There is a class of computer machines, although they do not have very advanced and developed attributes, that can use this attribute to build more advanced attributes, realize the characteristics of different learning levels, and then complete the automation to learn the construction of their own

DOI: 10.54254/2755-2721/5/20230594

characteristics, which is one of the most famous models, the convolutional neural network [5,6]. CNN is a feed-forward neural network, which consists of a multilayer network structure and can perform convolutional computation in training. It is a representative deep learning algorithm [7]. It can automatically classify the input message with the corresponding structure of the classification, which shows that it has a strong learning ability [8-10]. Convolutional neural network is a deep learning technique based on machine learning, and each layer of it uses a different function to transform one activation value into another. In general, a convolutional neural network has a convolutional layer, a pooling layer, and a fully connected layer (FC layer), which is usually preceded by an activation function that stacks all the layers so that a complete framework can be built.

Table 1. Comparison of relevant methods.

Models	Characteristics	advantages	drawbacks
AlexNet	AlexNet selects the nonlinear non-saturated RELU function as an activation function and adopts local response normalization (LRN).	In the training stage, the gradient attenuation speed is much faster than the traditional model	Only a part of the features are obtained by convolution kernel, and the generalization ability of the model decreases
ResNet-50	The residuals between layers jump together to introduce forward information and reduce gradient disappearance.	Avoid the problem of gradient dissipation and degradation	Most network layers prevent model degradation, and the error is relatively large
InceptionNet	Use different sizes of convolution kernels, perception of ascension	The multi-size convolution is used to broaden the network structure, and the convolution operation is used to reduce the number of parameters	Training may become less efficient in later stages, especially when the number of network layers becomes deeper.

The convolutional layer is mainly used to extract local features in the image, and when the number of convolutional layers deepens, the extracted information becomes more complex; the pooling layer is mainly used to downscale the parameters in the convexity; and the fully connected layer is used to combine the extracted image features and deliver the desired result. In traditional deep learning, as the number of network layers increases, the loss rate of the training set decreases until it becomes saturated. When the network degenerates, the shallow network is better trained than the deep network, but this does not mean that overfitting has occurred, and the deeper layers of the network do not make the loss function higher. In the convolutional neural network model, the training accuracy is often improved by controlling the perceptual field and setting appropriate initialization parameters during the convolution process. However, as the convolutional layers are deepened, However, when researchers train the model with backpropagation, some problems may occur, such as gradient disappearance, gradient explosion, etc. After finding relevant problems, in order to avoid these problems, the researchers developed the Residual Network (ResNet) after research and finally developed it.

The main feature of the ResNet is that it can avoid gradient disappearance, and gradient explosion problems when deepening the network to improve the model performance. The main principle of the residual network is the internal use of jump connections, i.e., the fusion of the input and output features of the current residual block, to maximize the preservation of the target information in the feature map. Because of its advantages in building deep models, residual networks are now used in network models for target detection, image recognition, and target segmentation. ResNet50 is one of the classical network structures, and the following figure shows the ResNet grid structure.

In traditional neural networks, the initialization of parameters and batch normalization is used to keep the feature information unchanged without losing model accuracy. Residual network compared to the previous model because a lot of features are more powerful than before, such as the classification performance, the corresponding detection ability improved, and its structure has changed, such as increasing the number of more network layers, solving the problem such as gradient problem puzzled researchers, and more widely applied to the back of the study. of 1×1 convolution in the residual can reduce the number of parameters, which can also reduce the amount of computation to a certain extent. In addition to the initial convolutional pooling layer and the final pooled fully connected layer, the ResNet50 network uses jump connections and constant mapping to reduce the impact of the gradient problem on model training.

This residual function has two main advantages, the first is that it can make the network more layers and make it deeper due to the different structure, and the second advantage is that such a residual function can achieve more easily optimized functions. The reason is that at each layer, it takes the corresponding reference (X) at the input and learns to form the residual function. In the structure used in this paper, namely ResNet, a total of two types are applied, namely residual module and convolutional network. The residual module is composed of 3*3 convolutional networks, and the structure of the convolutional network is a total of three, 1*1, 3*3 for cooperation, and 1*1 concatenated together as a residual module. In this paper, the author uses a 50-layer (Resnet). The first table below shows the characteristics, advantages, and disadvantages of ResNet and two other models. [7]. The following figure 1 is the Residual structure.

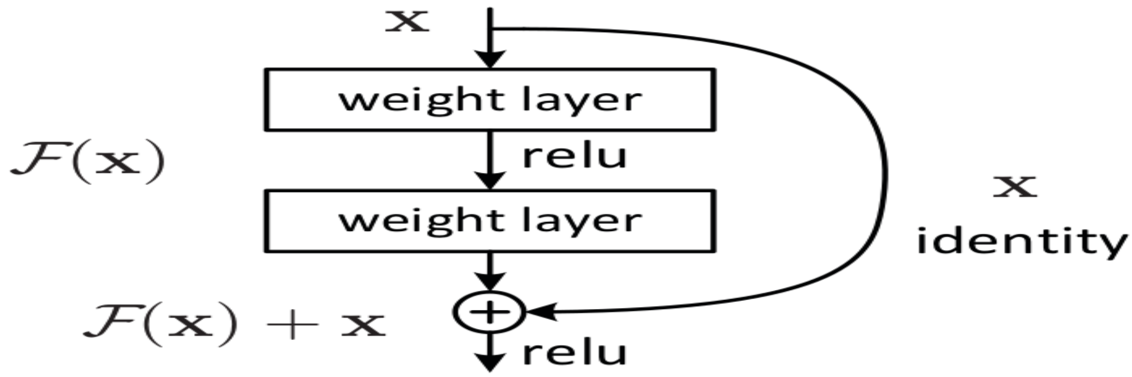

Figure 1. Figure of the Residual structure.
(Reference: https://doi.org/10.48550/arXiv.1512.03385)

2.3. Cross-entropy loss function

Nn.CrossEntropyLoss(), when training for classification, can be very efficient in some cases. During training, for each class weights are assigned, and the optional parameter weights should be a 1D tensor. This is very useful when you have an unbalanced training set. It is very useful when doing training for classification (specific classes) If researcher want to get some key data, such as very important, whether the expected output is very similar to the actual output. If we have k class samples during training, and the output value is also 0, we can infer that this is in line with the expected, in line with the desired output. If their proximity is high or extremely high, the labels can be inferred, as well as the differences in the data samples used by the researchers, and finally the network parameters can be updated through the backpropagation mentioned earlier

This function mainly illustrates the relationship between the two mentioned above, that is, the gap between the desired output value and the actual output value (both are probabilities), and the specific performance is that the higher the value is, the greater the probability gap is. The following icon gives the definition of this formula.

2.4. Data set and corresponding preparation

In image processing, the quality of the dataset has an important impact on the experiment. A better quality not only reduces the difficulty and time required for image pre-processing but also has a great impact on the loss rate and correctness of the model training. Therefore, the cat dataset used in this experiment was selected from the publicly available Baidu and Google image libraries. After considering the time and hardware and software environment, 20 feline species were selected to create the dataset. In the cat image classification, due to various factors, the image features extracted from different cats may be very similar, while the features extracted from the same cat species may be very different. Due to the small number of images in the dataset and the depth of the network used, the collected cat images were pre-processed in the experiment. First, the dataset was divided into two parts in the ratio of 8:2: the training set and the validation set. Second, images were recognized mainly by their shape and texture, and in this paper, the length and width of the images were randomly cropped to reduce the interference of the background (Figure 2).

Figure 2. Cropped image schematic.
(Photo Credit: Original)

The main advantage of PyTorch is that it supports GPU-accelerated model training and also provides the ability to dynamically change the network model, i.e., to add or remove relevant layers to the network model at will. PyTorch can be used to quickly build the desired neural network model. The hardware and software environments for the experiments are shown in Table 1.

The experimental data are first preprocessed by color transformation, angle adjustment, and mirroring. Then the data set is divided into training and validation sets according to 8:2. Then a convolutional neural network model is constructed by PyTorch to extract information from the targets in the images. In this experiment, only 20 feline species are classified and identified, so the model is more streamlined. The final fully connected layer and the average pooling layer were removed from

DOI: 10.54254/2755-2721/5/20230594

the network and replaced by a fully connected layer with 1024 nodes and 5 nodes to classify the targets more quickly. In the neural network algorithm, one of the factors affecting the training performance and recognition results is the hyperparameters. Therefore, in this experiment, the learning rate is set to 0.000 1, and the training results are stable when the number of training samples is 16, the number of images in the training set is 33, the number of images in the validation set, is 364, and the training epoch is set to 50. The training results and the training model are saved. Finally, the results are used to test the validation set, and the accuracy of the validation set, is obtained.

3. Result

Validation test and application test: To verify the validity and stability of the algorithm, the dataset is identified. The model training and testing in this paper are completed based on the PyTorch framework.

The results show that the system of examination of 20 kinds of cat's average recognition rate reached 92.1%, and in the picture below, namely angoraCat, bobtail, blackCat, blueCat, meseCat five cat species recognition rate reached 94.21%, the overall accuracy is higher. The recall rate reached 92.3%. FIG.3 shows that the recall rate reached 94.0% in the detection of five sampled cat pictures. In addition, the recognition rates of the test samples of each feline species also reached more than 86%, with little difference in recognition rates. There was no high recognition rate of one type of test species and low recognition rate of other species, which indicated that the classification features selected by the system were stable among various kinds of differences. Therefore, the system can make the classifier more optimized by increasing the number of training images, so as to improve the discrimination rate of the system. However, the number of training images of each species in the training set should be kept as small as possible to avoid misjudgment. Based on the domestic and foreign research on the images of real flies, the system test results of a large number of cat's image show that the recognition accuracy of the system is high, with an average recognition rate of 92.1%., the system can recognize more feline species by adding species to the training set. The increase in training sample size and keeping the balance between the number of training samples of each species are conducive to further improving the recognition rate of the system (Figure3-4).

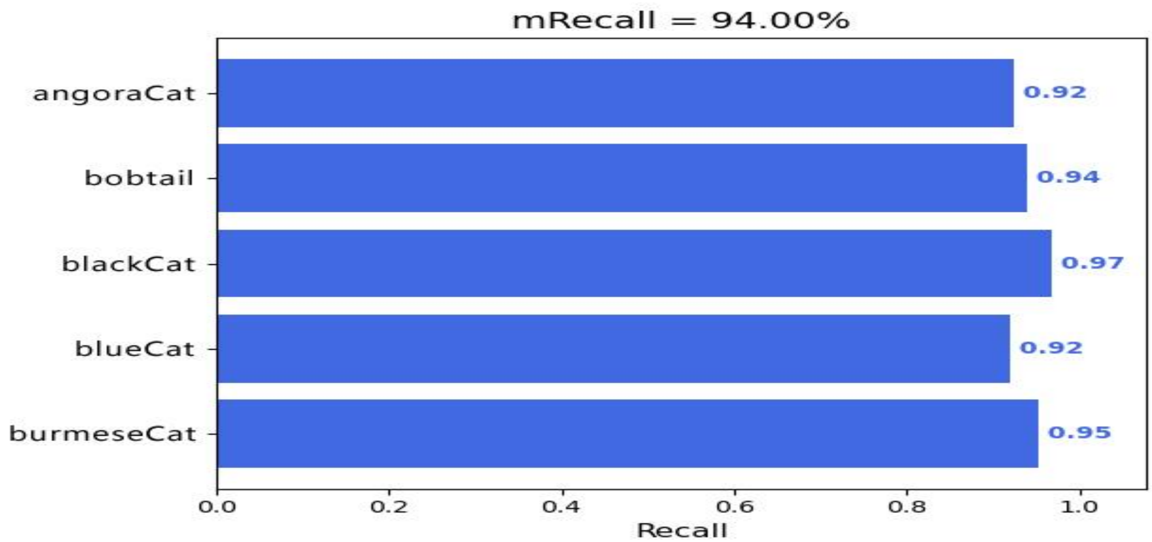

Figure 3. The recall of the model.
(Photo Credit: Original)

DOI: 10.54254/2755-2721/5/20230594

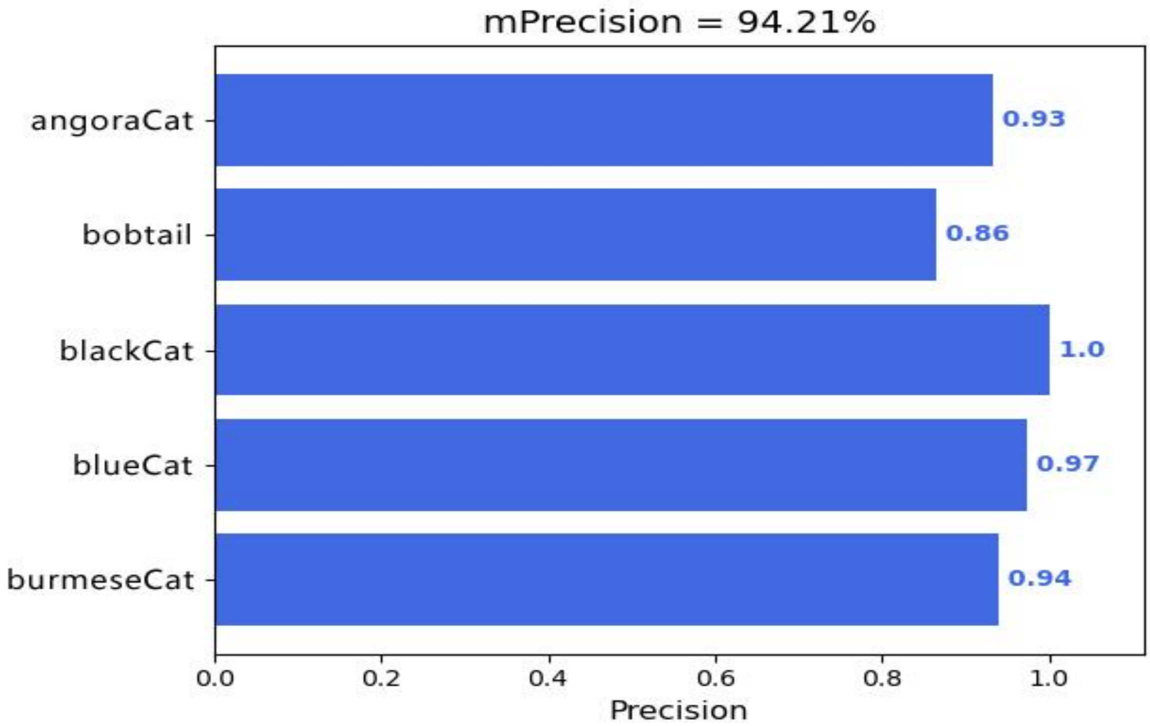

Figure 4. The precision of the model.
(Photo Credit: Original)

4. Conclusion

As research on felines continues to grow, image recognition techniques are increasingly helping researchers to identify species. However, past research on deep learning-based image recognition of cats has not gone far enough, leading to inefficiencies for researchers to achieve accurate identification. According to the researchers, there are currently over 500 million pet cats in the world, with approximately 40 breeds. Predictably, the popularity of cats has led to an increasing number of studies on cats. Based on the existing background, this paper uses a deep learning network based on the PyTorch framework to study this topic and explores it in depth, ultimately with a high degree of efficiency and accuracy. It is hoped that this study will improve the efficiency of related researchers' research. Through the recognition of cat images based on PyTorch architecture, high accuracy and recall rate are achieved, and the efficiency of classification is improved, but there are still some parts that need to be improved. For example, classification features can be added, because, with the increase of recognition types, it may be necessary to introduce other types of features, such as body shape, skin, etc., to achieve the purpose.

References

[1] Rochlitz, I 2005 The welfare of cats Vol. 3 Dordrecht: Springer.
[2] Frith-Macdonald, C 2008 Encyclopedia of cats. New York.: Parragon Books Ltd.
[3] Wu, M., and Chen, L. Image recognition based on deep learning. *In 2015 Chinese Automation Congress (CAC)*, 542-546.
[4] He, K., Zhang, X., Ren, S., and Sun, J 2022 Deep Residual Learning for Image Recognition. Retrieved 23 September 2022, from https://doi.org/10.48550/arXiv.1512.03385.
[5] Wang, Y., Ke, S., Wang, S., and Zheng, Z 2022 A Grapevine Virus Disease Detection Method Based on Convolution Neural Network. *In 2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA)*, 36-40.

DOI: 10.54254/2755-2721/5/20230594

[6] Fukushima, K., and Miyake, S 1982 Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. *In Competition and cooperation in neural net*s 267-285.
[7] De Boer, P. T., Kroese, D. P., Mannor, S., and Rubinstein, R. Y 2005 A tutorial on the cross-entropy method. *Annals of operations research*, 134, 19-67.
[8] LeCun, Y. 1998 The MNIST database of handwritten digits. http://yann. lecun. com/exdb/mnist/.
[9] Canziani, A., Paszke, A., and Culurciello, E 2016 An analysis of deep neural network models for practical applications. arXiv preprint arXiv:1605.07678.
[10] Ho, Y., and Wookey, S 2019 The real-world-weight cross-entropy loss function: Modeling the costs of mislabeling. *IEEE Access*, 8, 4806-4813.