Improved Image Classification Algorithm Based on Convolutional Neural Network

Xin Li[1,2], Luyu Dong[1,2], Mengting Li[1,2]

[1]Department of Information Science and Engineering, Guilin University of Technology, Guilin, China
[2]Guangxi Key Laboratory of Embedded Technology and Intelligent System, Department of Information Science and Engineering, Guilin University of Technology, Guilin, China
Email: 1996019@glut.edu.cn

Abstract

This article mainly introduces the image classification algorithm research based on the improved convolution neural network and some improvement ideas for the research of the classification based on the convolution neural network. The VGG16 model was improved in this article. By adding a dropout layer and a feature extraction layer, and performing L2 regularization on the loss function at the end, it deepens the model depth and improves the fit of the entire model. The experimental results show that the improved model can greatly improve the detection accuracy rate 98.7% than the traditional CNN algorithm 96.97% on the cat and dog data set.

Subject Areas

Computing Methodologies, Machine Learning, Machine Learning Approaches, Neural Networks

Keywords

Deep Learning, Image Classification, Convolutional Neural Network, VGG16 Model

1. Introduction

In recent years, with the rapid development of new information technology related to artificial intelligence, computer vision, as the intelligent field of artificial intelligence development, has also become more rapid with the improvement of computer performance [1]. Computer vision itself is through some photography, image information obtained from the camera equipment, and then through some software on the computer to convert it into digital signals for further

processing, making the image more intuitive and clearer, so as to realize the automated analysis of the target. Image classification technology is currently one of the most important applications of computer vision, widely used in people's lives and academic research [2]. As a classic image classification problem, two classification is an important research field. With the rapid development of deep convolutional neural networks, the accuracy of image classification and detection speed have been greatly improved. However, due to the complexity of the picture, the feature extraction of different pictures is very problematic in actual problems, and the problem of image distortion may occur, which makes it difficult to accurately identify and detect image classification under different models.

In the 21st century, most image classifications are still using traditional methods such as Decision Tree and Support Vector Machine [3]. The advantages of decision tree algorithms are that they are simple, easy to understand, and easily achieved on two classification problem, and the general-purpose type is better. But because of a relatively simple problem, when there are too many data in the data set, the problem of overfitting is easy to occur. At the same time, the effect is not satisfactory in the case of multi-classification. As a more scientific machine learning method, the SVM algorithm is a linear classifier with the largest interval in the feature space. It uses high-latitude mapping to solve some linear inseparable problems. It is also used in text, sound, image and other classification problems. Very good results have been achieved in processing. But like the decision tree algorithm, when the data is too large, the amount of calculation will increase exponentially, and it will eventually lead to overfitting, and eventually lead to the inefficiency of the model. Moreover, these two algorithms are based on manual feature extraction. Manual feature extraction depends on the person who extracts the feature, their familiarity with the task, and professionalism. Different people perform feature extraction, which often affects the subsequent model fitting degree. The level. In recent years, the successful application of deep learning algorithms represented by convolutional neural networks in the computer field has laid the foundation for CNN in image classification. Through the experiments of countless researchers, they have found that the accuracy of using convolutional neural networks to achieve various image classification tasks is higher than that of using SVM and decision tree algorithms. Experiments by many researchers have shown that the extraction of pneumonia image features by convolutional neural networks is more complete than manually extracted features, more fitting models, and more expressive. However, in the early stage, the structure of the convolutional neural network is relatively simple, and the ability to learn features for image classification tasks is also limited, and the accuracy of image classification needs to be improved.

Although the current transfer learning method has achieved better results in the task of image classification. However, migration learning cannot fit all image classification tasks. For example, the model parameters that are fitted on the pneumonia image are not very effective in the classification of flowers and plants. Because they did not improve the transfer learning model accordingly.

Therefore, this article proposes a deep neural network model based on the results of the VGG16 network to improve, and conduct experiments on the cat and dog data set.

This article studies the classification on the cat and dog data set. The main work is as follows: 1) Improve the VGG16 network, add the Dropout layer and add the Batch Normalization layer. 2) Construct a feature fusion layer and combine it with the improved VGG16 network structure to obtain more diverse nonlinear features.

2. Image Classification and Recognition Architecture

### 2.1. Convolutional Neural Network Model

When traditional image classification algorithms classify images, they often need to manually participate in feature processing, so the features obtained can only be used in this task. Moreover, the model trained in this way often has poor performance when fitting other tasks, does not have a general purpose, and there is still a lot of room for performance improvement.

Compared with traditional simple learning, deep learning is different in that deep learning can use a multi-layer non-linear structure to "autonomous" learning to characterize the characteristics of data. Most of the latter need to manually extract feature information, and because the manually extracted features often cannot well represent the essence of things, it is difficult to improve the learning effect. Compared with simple learning methods such as decision trees and SVM, deep learning learns the characteristics of data hierarchically through a multi-layer network structure, and then fine-tunes the classification information of each layer separately, which can realize the completion of complex samples with a simple structure. description. Convolutional neural network is a successful application of deep learning in the field of image processing. As a feedforward neural network with deep structure, it is generally composed of convolutional layer, pooling layer, fully connected layer and output layer. Compared with the traditional artificial neural network, the main difference between the two is that the former greatly reduces the number of parameters and the difficulty of training through convolution kernel and weight sharing. Combined with the training of a large amount of data, the convolutional neural network breaks through the limitations of traditional networks and has made great progress in the field of image processing [3].

2.2. Network Structure Design

In this paper, a feature fusion layer is designed, which can perform non-linear combination of the feature information learned by the VGG16 network, so as to obtain more non-linear combination results.

Figure 1 is a traditional model diagram of VGG16, and Figure 2 is an improved model diagram of the VGG16 network designed in this paper. Compared with the traditional VGG16 network model, it adds a Dropout layer and a Batch
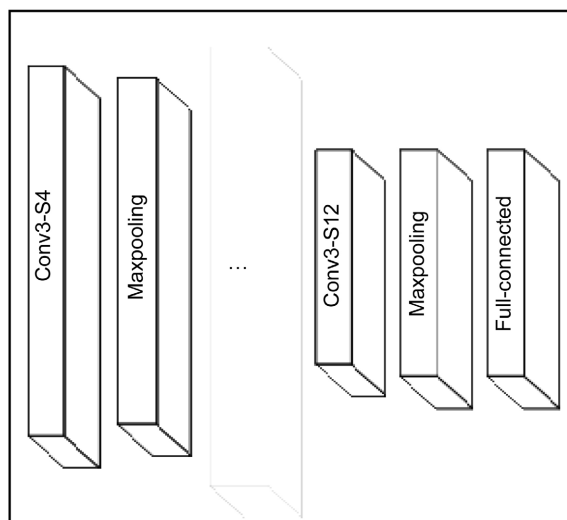
Figure 1. VGG16 network model diagram.
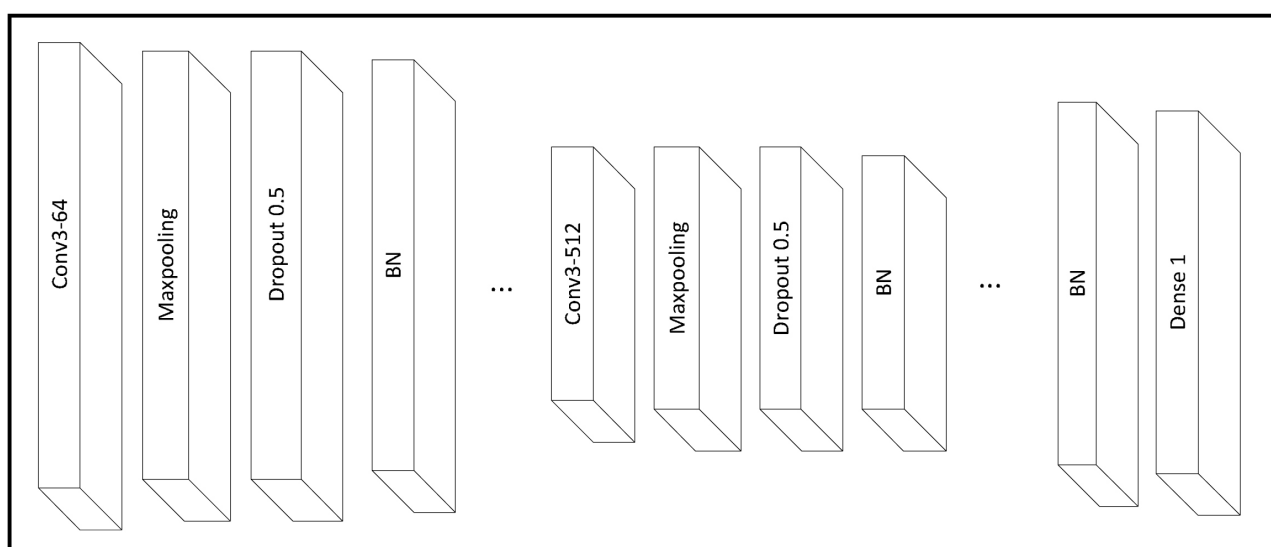
Figure 2. Improved model of VGG16.

Normalization layer. The feature fusion layer includes 3 fully connected layers, 2 batches of standardized layers, and 1 global pooling layer [4].

In the construction of the network model, the convolution kernels of the convolution layer follow the regulations of VGG16, which are 64, 128, 256, and 512, respectively, and the convolution kernels in the feature fusion layer are 1024, 512, and 1 in turn [5]. The node hiding rate parameter of the Dropout layer is generally selected between 0.3 - 0.7. This paper conducts many experiments between the best hiding rate, and finally sets it as 0.5 [6].

In the model training process, at the end of each iteration, the backpropagation algorithm (BP) is used to adjust the relevant parameters of the feature fusion layer. The calculation formula of the fully connected layer is:

$$Y_{out} = f\left(WX_{in} + b\right) \tag{1}$$

where Y is the output of the fully connected layer, and X is The input of the fully connected layer, W is the weight of training, $f(\cdot)$ is the activation function ReLU [7].

$$f(x) = \begin{cases} 0, x \le 0 \\ x, x > 0 \end{cases} \tag{2}$$

ReLU function has unilateral inhibition, which meets the requirements of the loss function in this experiment.

In the process of neural network training, the problem of gradient explosion or disappearance often occurs, which makes network training difficult. And as the network deepens, training will become more and more difficult, and convergence will become slower and slower. Therefore, the function of the BN layer is to keep the input of each layer of neural network in the same distribution in the process of deep learning network [8].

The Dropout layer is to suppress the problem of overfitting, so that after some layers of feature extraction in the deep learning network, it makes him randomly discard a part of the neurons, thereby suppressing the problem of overfitting [9].

L2 regularization is a commonly used technology in machine learning, and its main purpose is to control the complexity of the model and reduce overfitting. The basic regularization method is to add a penalty term to the original objective function to punish the model with high complexity [10].

3. Experiments

The experiment was conducted under Windows 10 system, the CPU was AMD2600X, and the graphics card was NVDIA 2060. The experimental environment is based on the Windows-based TensorFlow2 framework, and the experiment runs under the GPU environment.
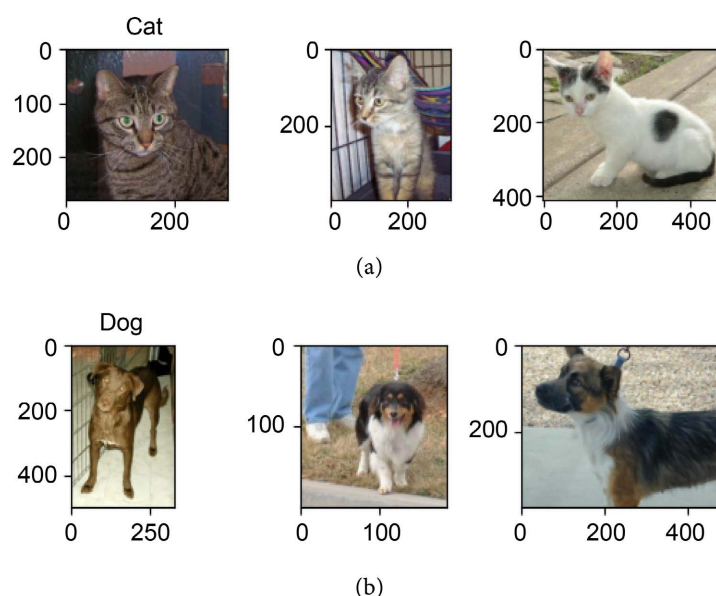
3.1. Data Set

This article uses a question from the Kaggle competition: the cat and dog data set [11]. The data set includes two independent cat and dog image data sets, 25,000 in the training set and 1000 in the test set. Each data set includes two categories of images. Cat pictures, Dog pictures. Figure 3 shows the realization of data related to cats and dogs.

The data distribution of cat and dog pictures is shown in Table 1.

3.2. Image Preprocessing

Image preprocessing is the process of reducing unnecessary image information and restoring or marking effective information. Commonly used methods include normalization and spatial transformation. The input image is generally smoothed in the scale space through a Gaussian blur kernel. After that, one or more features of the image are calculated by local derivative operation. Otherwise, in order to ensure the training effect, this paper randomly cut and reverse the data before passing it into the model.

(a)

(b)

Figure 3. Cat and dog data set display. (a) Cat data set; (b) Dog data set.

Table 1. Distribution of cat and dog data sets.

Data Set	Cat	Dog	Total
Training Set	13,446	11,554	25,000
Test Set	500	500	1000

3.3. Experimental Process

Step 1: Build the VGG16 model (Figure 1).

The network structure includes 13 convolution layers, 3 fully connected Layers, and 5 pool layers. The 13 convolution layers are a medium-sized 3 × 3 matrix with a moving step of 1. The number of convolution kernels gradually increased from 64 in the first layer to 128 to 256 and then to the last 512. The size of the convolution kernel of the pooling layer is 2 × 2, and the step size is 2.

Step 2: After building VGG16 model, this article adds a Dropout layer and Batch Normalization, and sets the Dropout layer parameters to 0.5 to randomly discard 50% of the neurons.

Step 3: The number of model iterations is set to 40, the activation function of the convolutional layer is set to ReLU, the activation function of the last layer of fully connected layer is set to sigmoid, and because it is a binary classification problem, the final loss function is set to binary cross-entropy.
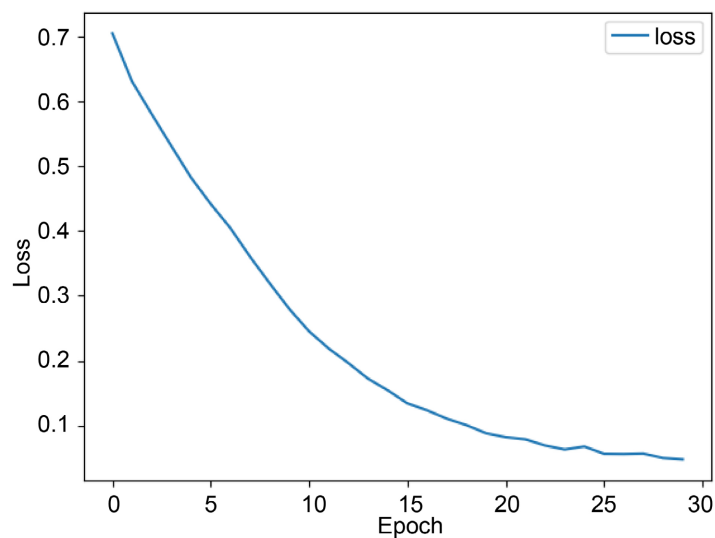
4. Experimental Result and Analysis

This paper uses the accuracy rate and loss function to evaluate the performance of the model. Figure 4 is the accuracy rate of the training set and test set of the model, and Figure 5 is the decrease rate of the loss function.

The loss function used in this article is:
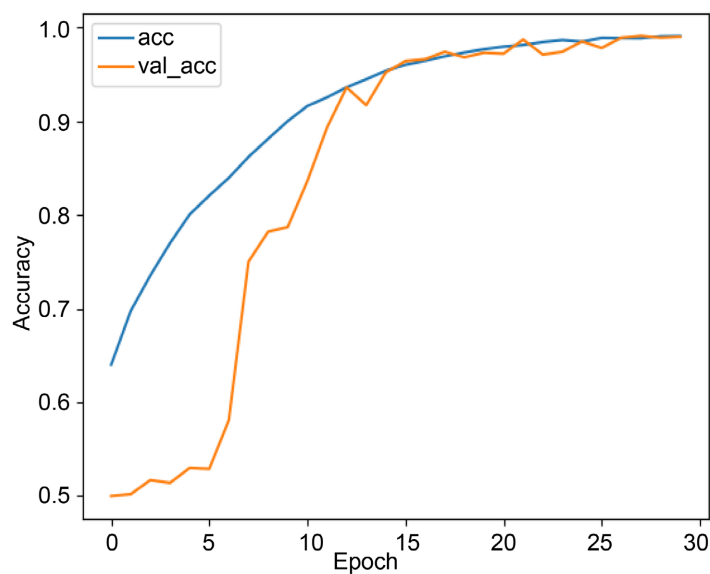
Figure 4. Correct rate of training set and correct rate of test set.

Figure 5. Loss function descending gradient.

$$\text{MSE} = \frac{\sum_{i=1}^{n}\left(y_i - y_i^p\right)^2}{n} \tag{3}$$

where n is the number of samples, and y_i is the real data.

The accuracy calculation formula is:

$$\text{ACC} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{4}$$

where TP is the predicted value of the sample matches the true value and both are positive, and FP is the sample predicted value is positive and the true value is negative, FN is the predicted value of the sample is negative and the true value is positive, TN is the predicted value of the sample matches the true value and both

Table 2. Comparison of accuracy rates of VGG16 and VGG16 improved models.

Experiment number	Recognition methods	Accuracy
1	SVM	86.90%
2	VGG16	96.97%
3	VGG16 Improve the model	98.7%

are negative.

This paper uses the VGG16 model for comparative experiments. The experimental results are shown in Table 2. As can be seen from Table 2, in comparison with its own VGG16, the accuracy rate has increased by 2 percentage points, while the loss function shown in Figure 4 has decreased the gradient also meets the experimental requirements. The experimental results show that the improved VGG16 model in this paper has a good detection effect.

5. Conclusion

This paper proposes an improved VGG16 convolutional neural network model to realize the classification of cat and dog images. Based on the VGG16 model, this paper adds a dropout layer, batches standardization layer, and constructs a feature fusion layer to make the model obtain more diversified non-linear feature representations, improve accuracy, and reduce the ability of the model to overfit. In comparative experiments, the training set and test set of the model in this paper have achieved better recognition results and generalization capabilities [12]. In the follow-up work, the model will be considered, the number of layers of the model will be improved, and the capsule network will be added as a feature fusion layer. Carry out three classifications of animals, and use semantic segmentation to segment and locate the image.

Acknowledgements

Conflicts of Interest

The authors declare no conflicts of interest.

References

[1] Lu, D. and Weng, Q. (2007) A Survey of Image Classification Methods and Techniques for Improving Classification Performance. *International Journal of Remote Sensing*, 28, 823-870. https://doi.org/10.1080/01431160600746456

[2] Lecun, Y. and Bottou, L. (1998) Gradient-Based Learning Applied to Document Recognition. *Proceedings of the IEEE*, **86**, 2278-2324. https://doi.org/10.1109/5.726791

[3] Abiyev, R.H. and Ma'aitah, M.K.S. (2018) Deep Convolutional Neural Networks for

Chest Diseases Detection. *Journal of Healthcare Engineering*, **2018**, Article ID: 4168538. https://doi.org/10.1155/2018/4168538

[4] Cicero, M., Bilbily, A., Colak, E., *et al.* (2017) Training and Validating a Deep Convolutional Neural Network for Computer-Aided Detection and Classification of Abnormalities on Frontal Chest Radiographs. *Investigative Radiology*, **52**, 281. https://doi.org/10.1097/RLI.0000000000000341

[5] Jun, S.H., Park, B.H., Seo, J.B., et al. (2018) Development of a Computer-Aided Differential Diagnosis System to Distinguish Between Usual Interstitial Pneumonia and Non-specific Interstitial Pneumonia Using Texture- and Shape-Based Hierarchical Classifiers on HRCT Images. *Journal of Digital Imaging*, **31**, 235-244. https://doi.org/10.1007/s10278-017-0018-y

[6] Da Silva, F.L. and Costa, A.H.R. (2019) A Survey on Transfer Learning for Multiagent Reinforcement Learning Systems. *Journal of Artificial Intelligence Research*, **64**, 645-703. https://doi.org/10.1613/jair.1.11396

[7] Szegedy, C., *et al.* (2015) Going Deeper with Convolutions. 2015 *IEEE Conference on Computer Vision and Pattern Recognition* (*CVPR*), Boston, MA, 7-12 June 2015, 1-9. https://doi.org/10.1109/CVPR.2015.7298594

[8] Haykin, S. and Kosko, B. (2009) Gradient Based Learning Applied to Document Recognition. Wiley-IEEE Press, Piscataway, 78-83.

[9] Wang, H. (2020) Garbage Recognition and Classification System Based on Convolutional Neural Network VGG16. *Proceedings of the* 2020 3*rd International Conference on Advanced Electronic Materials*, *Computers and Software Engineering* (*AEMCSE*), Shenzhen, 24-26 April 2020, 252-255. https://doi.org/10.1109/AEMCSE50948.2020.00061

[10] Swasono, D.I., Tjandrasa, H. and Fathicah, C. (2019) Classification of Tobacco Leaf Pests Using VGG16 Transfer Learning. *Proceedings of the* 2019 12*th International Conference on Information & Communication Technology and System* (*ICTS*), Surabaya, 18 July 2019, 176-181. https://doi.org/10.1109/ICTS.2019.8850946

[11] https://www.kaggle.com/chetankv/dogs-cats-images

[12] Xu, B., He, X. and Qu, Z. (2021) Asian Food Image Classification Based on Deep Learning. *Journal of Computer and Communications*, **9**, 10-28. https://doi.org/10.4236/jcc.2021.93002