CIFAR-10 Image Classification Based on Convolutional Neural Network

Xiyun Lv

College of Electrical & Information Engineering, Southwest Minzu University, Chengdu 610041, China
Email: 1045106108@qq.com

Abstract. Aiming at the problems of gradient diffusion and network redundancy caused by the degradation of network performance during the training of most convolutional neural image classification models, the ResNet neural network model was improved, and the data expansion technology was used to expand the data and use SGD to fine-tune the depth Convolutional network to avoid gradient dissipation. In the CIFAR-10 test set, experiments show that the test accuracy of CIFAR-10 in the model data set reaches 90.85%. Compared with other models, it improves the accuracy of image classification. And by manually observing and comparing the classification effects of the 10 types of objects of the model, the model can distinguish each object more accurately.

Keywords: image classification, ResNet, data augmentation, CIFAR-10

1   Introduction

Image classification is one of the basic research topics in the field of computer vision recognition. Its research goal is to predict the category label of the input image for a given image and a set of classification labels. In the entire classification process, feature extraction and selection are collectively referred to as feature expression, and good feature expression plays a vital role in improving the accuracy of image classification. Therefore, for image classification tasks, deep learning has gained widespread attention and applications with its powerful feature extraction capabilities. Convolutional neural network is one of the important models of deep learning. Its weight sharing network structure greatly reduces the complexity of the network model and reduces the number of parameters, thus avoiding the complex feature extraction and data reconstruction in traditional recognition algorithms process.

As a very practical and effective machine learning algorithm, deep learning has made breakthrough progress in image classification tasks. However, with the continuous increase of the number of layers of deep convolutional networks, the network training process is often accompanied by gradient diffusion, which will make the model training speed very slow, or even fail to converge, and affect training efficiency. It will cause network redundancy problems. An ultra-deep network can be constructed by simply stacking more network layers. However, as the scale of the structure increases, a large number of parameters are generated in the network, which increases the redundancy of the network, which in turn leads to the degradation of network performance.

In response to the above problems, this paper improves the ResNet neural network model on the existing basis, uses data augmentation technology to expand the CIFAR-10 data set with a small data scale, and uses small batch SGD with additional momentum coefficients Fine-tune the deep convolutional network to avoid gradient dispersion. Then, the data contains a large number of features, noise, and the CIFAR-10 data set with different proportions of objects to be identified is carried out for verification experiments.

2   Related Works

Bengio et al. [1] replaced the restricted Boltzmann machine in each layer of the network with an autoencoder in 2007, and proposed a deep network structure of stacked autoencoding, and obtained it on the MNIST data set. 98.6% of image classification accuracy. Alex Krizhevsky et al. [2] proposed the deep convolutional network model Alex Net in 2012 and won the championship with a significant advantage in the Image Net competition, and reduced the top-5 classification error rate on the Image Net dataset to 16.4%, Karen et al. [3] from the University of Oxford proposed the VGG model, which first fixed other

parameters in the structure, and imitated a larger receptive field effect by using multiple 3×3 convolution kernels. Szegedy et al. [4] proposed a new network model, Goog Le Net, to solve the problem of over-fitting caused by excessive parameters and gradient dispersion caused by network deepening. The Highway network proposed by Srivastava et al. [5] successfully trained deep convolutional networks by using shortcut connections. Inspired by the Highway network, He et al. [6] proposed a deep residual network model (Deep Residual Network, Res Net), which uses the concept of shortcut connections in the residual unit to perform residual learning on the deep convolutional network, which can effectively use the depth It also improves the generalization ability and further accelerates the training speed of the network. The model won the championship with a top-5 classification error rate of 3.57% on the Image Net dataset in the ILSVRC-2015 competition.

3   Dataset

The CIFAR-10 data set [7] is provided by Krizhevsky et al. of the Hinton team. Among them, 50,000 images are used for training and 10,000 images are used for testing. There are a total of 60,000 RGB three-channel images with a size of 32×32. In addition, the data set consists of 10 classes: plane, car, bird, cat, deer, dog, frog, horse, ship, truck. Each category has 5000 training images and 1000 test images. Among them, the test batch contains 1000 randomly selected images from each category, and the training batch contains the remaining images in random order. However, some training batches may contain more images from one category, which makes the number of various types of images included in each batch vary. Therefore, in order to ensure the balance of sample probabilities, all training data contains 5000 images per class. The following partial images are classes in the data set, and each class contains 10 random images, as shown in Figure 1.

Data sets are highly relevant and suitable for our goals because they are all real images, not just digital or categorical inputs. Data preprocessing is done by using the transform function through a python script. Shuffle the data to ensure that these classes are not clustered together, and normalize them to reduce data redundancy and improve data integrity by 0.5 and 0.5.
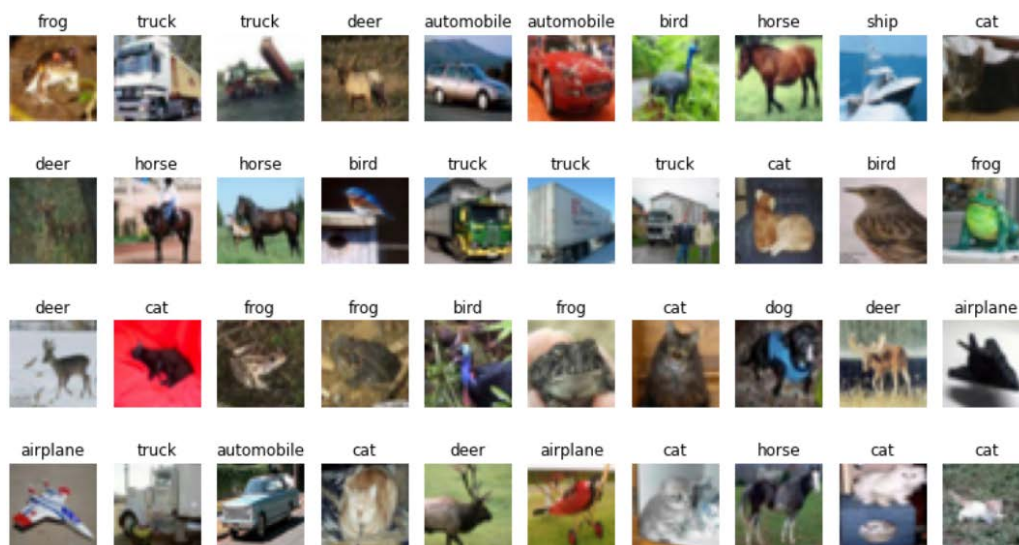

Figure 1. Sample image of CIFAR-10 data set

4   Experiment and Result Analysis

4.1   Data Augmentation

In actual deep learning projects, the demand for data sets is very large. A large amount of data training makes the generalization ability of the model stronger, and to a certain extent restrains the appearance of

overfitting. Data augmentation is an effective means to multiply data sets, and it was successfully applied to AlexNet for the first time to achieve great results. According to the data enhancement methods proposed in related papers, the original data obtained is usually flipped horizontally, flipped up and down, whitened, changed contrast, and randomly cropped to increase the data sample size of this article. The picture after data augmentation is similar to the original picture distribution and is a brand new picture to the model, so as to achieve the purpose of obtaining more training data and labeling.

As a mature and convenient deep learning framework, Keras provides a very efficient API for image data augmentation. This model follows the standard data extension strategy used for the CIFAR dataset [8, 9, 10, 11, 12], using Keras ImageDataGenerator interface for data augmentation. Image flipping is divided into horizontal flipping and upside down flipping. The reason why the image is flipped is because in many image recognition problems, the flipping of the image should not affect the recognition result. Therefore, when training the neural network model for image recognition, the training image can be randomly flipped, so that the trained model can recognize entities from different angles. We first standardize the data. Fill 4 pixels on each side of the image, then randomly select a 32×32 crop from the image or its horizontal flip, and use the data equalization strategy described in [13] to perform small batch sampling to compensate for the category Uneven distribution, and then continue to use SGD training to remove data augmentation, and then train until convergence. Figure 2 shows the image after data augmentation.
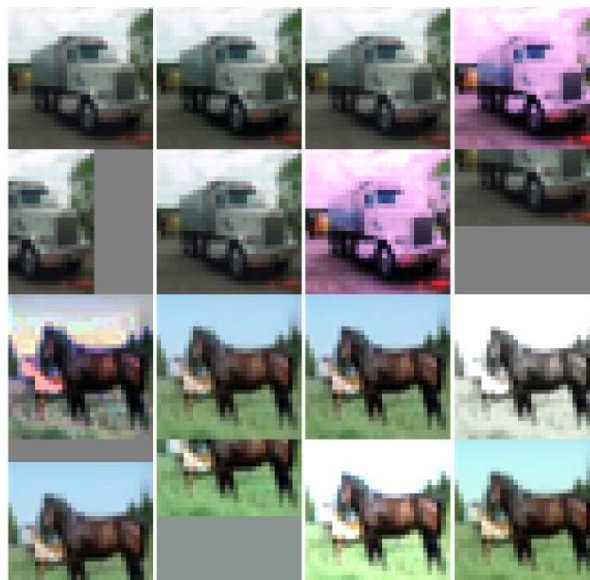

Figure 2. The image after data augmentation

4.2 ResNet Convolutional Network Model

The classic model in image classification is the convolutional neural network CNN, but as the number of layers increases, CNN shows a degradation problem, that is, the performance of the deeper network is not as good as the shallower network; this is not caused by overfitting, because the degradation gap is shown on the training set. ResNet can better solve this problem.

Residual neural network ResNet (Residual Network), won the 2015 ImageNet championship. The main idea of ResNet is the residual block. Kaiming He et al. designed a skip connection (or shortcut connections) structure to make the network have stronger identity mapping (identity mapping) capabilities, thereby expanding the depth of the network, and also Improved network performance. The structure of the residual block is shown in Figure 3.
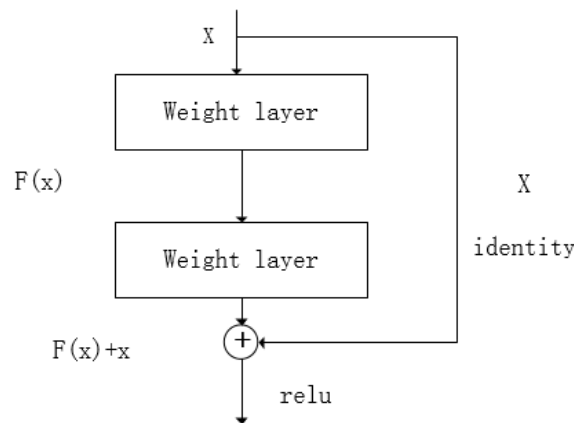
Figure 3. Residual block structure

$$F(x) = H(x) - x \hspace{3cm} (1)$$

Formula (1) is a deep residual of ImageNet, where x is the output of the shallow layer, H(x) is the output of the deep layer, and F(x) is the transformation represented by the two layers sandwiched between the two. Assuming that the accuracy of the shallow network in the network has reached saturation, any changes to the feature x will increase the loss, so through the "shortcut connections" method, the input x is directly passed to the output as the initial result, That is, x will continue to be passed from the identity mapping path, at this time F(x) will automatically tend to learn to become 0. In this way, the initial purpose is achieved without increasing the computational cost: in the forward process, when the output of the shallow layer is sufficiently mature (optimal), the layers behind the deep network can realize the role of identity mapping. The example residual block is shown in Figure 4.
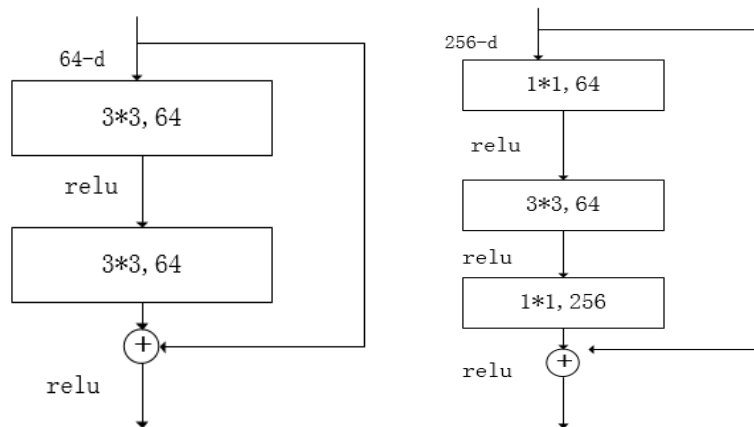

Figure 4. Deep residual block structure

ResNet can solve the degradation problem (degradation problem) caused by too many layers in the deep network. The model used in this article is an improvement based on the ResNet model. When training the CIFAR10 data set, it is trained without parameters shortcut connection. The input of the network is a 32×32 picture, and the average value is subtracted from all points after preprocessing. The first layer is a 3×3 convolutional layer. After that, there are a total of 6n 3×3 convolutional layers, of which there are 2n convolutional layers for the features maps {32, 16, 8}. The residual network is finally followed by a global average pooling and a full connection of 10 categories.

4.3 Experimental Setup and Result Analysis

The model is trained using stochastic gradient descent, and the initial learning rate is 0.1. The model was trained for 200 epochs and the batch size was 32. When 10 epochs (patience = 10) verify that the accuracy does not increase by at least 0.01% (min delta = 0.0001), the learning rate is divided by 0.1 (factor = 0.1). Use ModelCheckpoint to save the model with the highest verification accuracy. When the verification accuracy does not increase by at least 0.01% (min delta = 0.0001) in 25 epochs (patience = 25), the training will stop. It can be seen from Table 4-1 that after 200 epochs training, the test accuracy of our model reaches 90.85%. Compared with other models, this model has the best effect and improved performance. Figure 5 below shows the model's training and testing accuracy and loss rate.
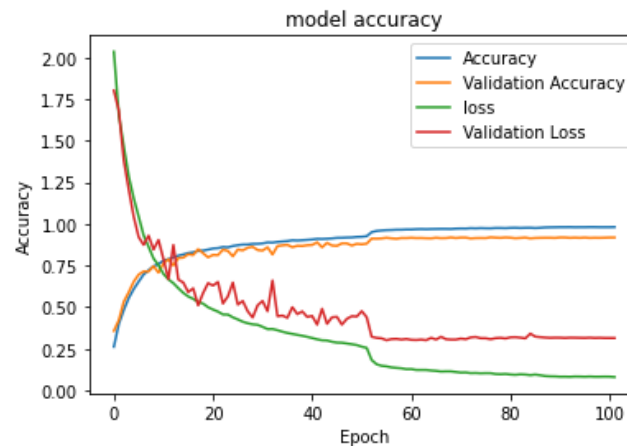

Figure 5. Model accuracy

Table 4-1. Compare recognition accuracy on CIFAR-10 dataset

Algorithm	Test accuracy
Yang Mengzhuo et al [14]	89%
Zhou Yuepeng et al [15]	86.74%
Pang Sisi et al [16]	86%
Ours	90.85%

Table 4-2. Classification accuracy of each object

object	precision	recall	F1-score	support
b'airplane'	0.91	0.93	0.92	1000
b'automobile'	0.95	0.96	0.96	1000
b'bird'	0.88	0.88	0.88	1000
b'cat'	0.80	0.81	0.80	1000
b'deer'	0.89	0.93	0.91	1000
b'dog'	0.87	0.83	0.85	1000
b'frog'	0.94	0.94	0.94	1000
b'horse'	0.96	0.92	0.94	1000
b'ship'	0.95	0.95	0.95	1000
b'truck'	0.95	0.95	0.95	1000

Table 4-2 shows the classification scores of the 10 types of objects in the CIFAR-10 data set by the improved model. It can be seen that the F1 score of the car is the highest, 0.96, which means that the model has the highest accuracy in its classification. The F1 score of the cat is the lowest, only 0.80, and the classification accuracy is not high.

In order to better observe the effect of the model on image classification, we display the visual test set of any 10 prediction results. As shown in Figure 6, it can be seen that the prediction of the 9th picture is wrong, and the other predictions are correct.
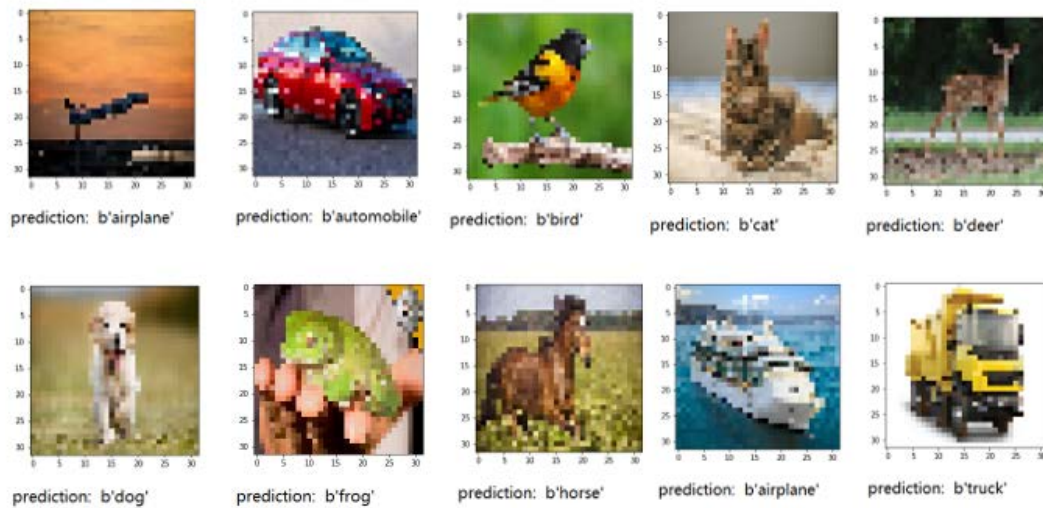

Figure 6. The predicted picture

5   Conclusion

This paper improves on the basis of the ResNet neural network model, uses data augmentation technology to expand CIFAR-10, trains and tests on the CIFAR-10 data set, uses TensorBoard to visualize the improved model on the classification of objects in the data set, and the experiment is The training accuracy of CIFAR-10 in the model data set reached 98.79%, and the test accuracy reached 90.85%, which improved the accuracy of image classification. Although we have made some progress, there are still some problems. Our model cannot accurately identify relatively similar pictures such as airplanes and ships. In the future, we will consider adding an attention mechanism to optimize the model to obtain better performance to meet actual application requirements.

References

1. Y. Bengio, P. Lamblin, D. Popovici, et al. Greedy layer-wise training of deep networks[C]. In Proceedings of the 20th Annual Conference on Neural Information Processing Systems, 2006: 153-160.
2. A. Krizhevsky, I. Sutskever, E. G. Hinton. Image Net classification with deep convolutional neural networks[J]. Communications of the ACM, 2017, 60(6): 84-90.
3. K. Simonyan, A. Zisserman. Very deep convolutional networks for large-scale image recognition[J]. ar Xiv preprint ar Xiv: 1409.1556, 2014.
4. S. Christian, W. Liu, Y. Q. Jia, et al. Going deeper with convolutions[C]. In Proceedings of the 28th IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2015: 1-9.
5. R. K. Srivastava, K. Greff, J. Schmidhuber. Training very deep networks[C]. In Proceedings of the 29th Annual Conference on Neural Information Processing Systems, 2015: 2377-2385.
6. K. M. He, X. Y. Zhang, S. Q. Ren, et al. Deep residual learning for image recognition[C]. In Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.
7. Y. Yamada, M. Wamura, K. Kise. Deep pyramidal residual networks with separated stochastic depth[J]. ar Xiv preprint ar Xiv: 1612.01230, 2016.

8. K. M. He, X. Y. Zhang, S. Q. Ren, et al. Deep residual learning for image recognition[C]. In Proceedings of the 29th IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770-778.

9. S. Zagoruyko, N. Komodakis. Wide residual networks[C]. In Proceedings of the 27th British Machine Vision Conference, 2016: 1-12.

10. T. Zhang, G. J. Qi, B. Xiao, et al. Interleaved group convolutions for deep neural networks[J]. ar Xiv preprint ar Xiv: 1707.02725, 2017.

11. G. Huang, Z. Liu, L. Van Der Maaten, et al. Densely connected convolutional networks[C]. In Proceedings of the 30th IEEE Conference on Computer Vision and Pattern Recognition, 2017: 2261-2269.

12. L. M. Zhao, J. D. Wang, X. Li, et al. On the connection of deep fusion to ensembling[J]. ar Xiv preprint ar Xiv: 1611.07718, 2016.

13. L. Shen, Z. C. Lin, Q. Huang. Relay backpropagation for effective learning of deep convolutional neural networks[C]. In Proceedings of the 14th Computer Vision European Conference, 2016: 467-482.

14. Yang Mengzhuo, Guo Mengjie, Fang Liang. Research on Image Classification Algorithm Based on Keras Convolutional Neural Network[J]. Science and Technology Wind, 2019, 000(023):117-118.

15. Zhou Yuepeng, Lu Xili. CNN Application Research Based on Combined Activation Function[J]. Journal of Shaoguan University, 2019(12):24-30.

16. Pang Sisi, Huang Chengcheng. Research on Image Classification Based on Convolutional Neural Network[J]. Modern Computer, 2019(23).