Proceedings of the 5th International Conference on Computing and Data Science
DOI: 10.54254/2755-2721/15/20230831

Image classification based on CNN with three different networks

Wenyan Sun

Shanghai Lixin University of Accounting and Finance, Shanghai 201209, P.R.China

201330122@stu.lixin.edu.cn

Abstract. Image classification refers to classifying images based on the different features reflected by the information in each image. Image classification is a fundamental issue in computer vision and has important significance. It has a wide range of applications, such as autonomous driving, face recognition, image retrieval, and other fields. This article first gives a bird's-eye view of the development of image classification and briefly introduces the factors that affect the accuracy of convolutional neural networks. Experiments and comparative analysis are conducted on the effects of convolutional layer numbers and optimizers on the accuracy of convolutional neural networks. Three convolutional neural networks with different convolutional layer numbers are built, their structural design is introduced in detail, and their structural diagrams were given. Different models are used to classify images from the CIFAR 10 dataset. The experimental results show that with the continuous increase of the convolutional layer, the accuracy rate is continually improving, but the program running time is continually increasing. In addition, using different optimizers can lead to changes in accuracy.

Keywords: image classification, deep learning, convolutional neural network.

1. Introduction

Image classification refers to distinguishing images based on their different characteristics reflected in the image information. It is a fundamental issue in computer vision. In addition, image recognition has important significance in daily life, and is applied in many fields, such as automatic driving, face recognition, automatic album classification, image retrieval, and so on.

Before proposing deep learning algorithms, traditional algorithms such as SVM and random forest classifiers are often used to solve image classification problems. Using traditional image classification algorithms usually includes several stages, such as feature extraction, feature coding, classifier design, and image classification. In 1998, Lecun Y et al. proposed using the LeNet5 algorithm to classify handwritten data sets (MNIST) and found that convolutional neural networks are superior to other techniques in processing two-dimensional shapes [1]. Since then, the application of convolutional neural networks in image recognition has entered people's vision, and related research has also begun to expand, such as image recognition, character recognition, gesture recognition, and so on. Hinton G E and Salakhutdinov R R (2006) first proposed the concept of deep learning and found that deep networks can better achieve dimensionality reduction of data than principal component analysis [2]. In 2006, deep learning algorithms received widespread attention again. Alex Krizhevsky et al. (2017) constructed and trained a large-scale deep convolutional neural network, and added a "discard" regularization method.

DOI: 10.54254/2755-2721/15/20230831

In the ILSVRC-2012 competition, they obtained a high accuracy rate and won first place, with a difference of nearly 10% from the second place [3]. The significant improvement in the accuracy of convolutional neural network algorithms and the development of discarding layers have caused more and more people to pay attention to convolutional neural networks. With this disruptive initiative, convolutional neural networks have begun to be widely used in various fields.

Convolutional neural networks (CNN) are mainly used in computer vision and natural language processing. Convolutional neural networks have long been one of the core algorithms in the field of image recognition. Chaganti S Y et al. (2020) used support vector machines (SVM) and CNN to classify images, and compared the accuracy of classification, proving that CNN has better results than traditional machine learning algorithms in image classification under large data sets for large-scale image classification problems [4]. In addition, in 2022, Tripathi, S, and Singh, R used CNN to classify cat and dog images, and found that CNN can automatically extract features without requiring feature engineering, and CNN has better accuracy in image classification than other algorithms [5].

Actually, the performance of convolutional neural networks (CNN) is affected by many factors. In 2015, Nielsen MA mentioned in Neural Networks and Deep Learning that increasing the number of convolutional layers used in convolutional neural networks (CNN) may improve the accuracy of image recognition. In addition, increasing the number of full connections, adding regularization, waiver, pooling, and other methods may improve the accuracy of the model [6]. In 2014, Kingma D and Ba J first proposed the Adam algorithm, and compared with other random optimization methods, Adam performed better in experiments. In addition, they also proposed a variant algorithm of Adam, Adamax [7]. Ruder S (2016) studied different variants of gradient descent, summarized the difficulties encountered, and introduced common optimization algorithms [8].

According to the existing literature, it can be found that CNN has good effects in image classification, image recognition, and other fields. However, the performance of CNN is affected by many factors, such as the number of convolution layers, the number of fully connected layers, pooling layers, and descending algorithms. However, there are few comparative experiments and studies on different convolution levels and different descent algorithms in existing papers. Based on this, this article will discuss and compare the impact of different convolution levels and descent algorithms on the accuracy rate when using the same dataset.

2. Method

2.1. Brief introduction of the CIFAR-10 dataset

The CIFAR-10 dataset consists of 10 categories of 60000 color images, including cats, dogs, airplanes, and so on, and these categories are completely mutually exclusive. Each class has 6000 images, with an image size of 32 * 32. The dataset is divided into five training batches and one test batch. Each training batch consists of a total of 10000 randomly selected images from each category, and randomly selected images are not repeatedly selected [9].

2.2. Data pre-processing

This time, 50000 training set data are used. For this 50000 data, first, convert all images into specific numbers, and then divide the data into training data and verification sets. The training set is used to train the model, and the verification set is used to verify the quality of the model and the calculation accuracy. There are 40000 training data and 10000 verification data. The dataset is classified using random grouping and the random_split() function. Then, use the DataLoader() function to group and bundle the data in the two datasets, and process the data in small batches each time. All the batch sizes in this model are 128.

DOI: 10.54254/2755-2721/15/20230831

2.3. Model structure adopted

2.3.1. Model construction and training

The model uses a convolutional neural network. First, the class ImageClassifier() is defined, which classifies images. For the Cifar10 dataset, images are divided into 10 categories. Establish a convolutional neural network, and establish a model based on the initial number of input channels of 3 and the number of convolutional layers to be established. The activation function used in this convolutional neural network is the ReLU function. Then, using the BCELoss loss function, select an appropriate descent algorithm, test the loss value of the training dataset, and continuously train the model through cycles to continuously reduce the loss value, achieving the goal of model optimization.

2.3.2. Model testing and calculation method of accuracy

According to the above training model, the previously separated test set is substituted into the trained model for testing. Through variable cycling, each image is predicted and then compared with a given type to determine whether the prediction is correct, thereby calculating the accuracy of the model prediction.

2.3.3. Experimental process of different models

The general structure of the model is described above. The main structures of the following different models are basically the same. In Model 1 and Model 2, only the number of layers of the convolution layer will be changed, keeping the optimization function, activation function, and so on unchanged, using the Adam optimization algorithm, In model 3, different optimization functions will be used for comparison, keeping the convolution layer number, activation function, etc. unchanged.

2.3.4. Model 1 – three convolution layers

The convolutional neural network in Model 1 is a 3-layer convolutional layer, which maintains the previously described data processing process, only changing the number of convolutional layers. Set the kernel size of each layer of the convolutional layer to 3 * 3, and the padding to 1. Change the number of input channels, output channels, and stride. The convolution layer of the first layer has an input value of 3, an output value of 96, and a default value of 1, The input channels of the second layer of the convolutional layer are the same as the out channels of the first layer of convolutional layer, which is 96. The out channels of the second layer are defined as 384, with a side of 2, Similarly, the in channels of the third layer of the convolution layer is 384, the out channels are 256, and the side is 2, Finally, add a pooling layer and select the MaxPool2d function where kernel size is 3.The structural diagram of model 1 is shown in Figure 1:

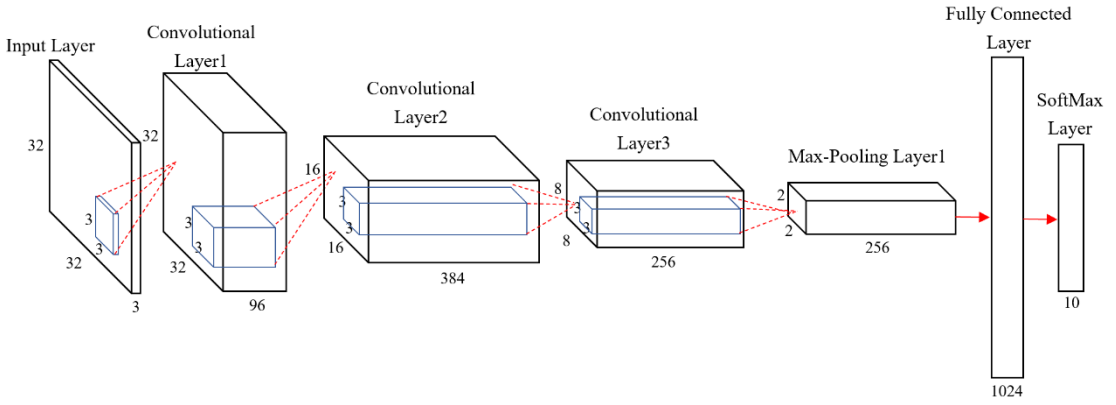

Figure 1. CNN Model 1 Structure Diagram

DOI: 10.54254/2755-2721/15/20230831

2.3.5. Model 2 – four convolution layers

The convolutional neural network in model 2 is a 4-layer convolutional layer, maintaining the same basic assumptions as model 1. The convolution layer of the first layer is the same as model 1, The input channels of the second layer of convolution layer 6 are the same as the output channels of the first layer of the convolution layer, and the output channels of the second layer are defined as 256, Add a pooling layer and select the MaxPool2d function, where the kernel_ Size is 2, Similarly, the input channels of the third layer of convolution layer are 256, and the output channels are 384, The fourth layer of convolutional layer has 384 input channels and 256 output channels, Finally, add a pooling layer and select the MaxPool2d function where kernel size is 2.The structural diagram of model 2 is shown in Figure 2:

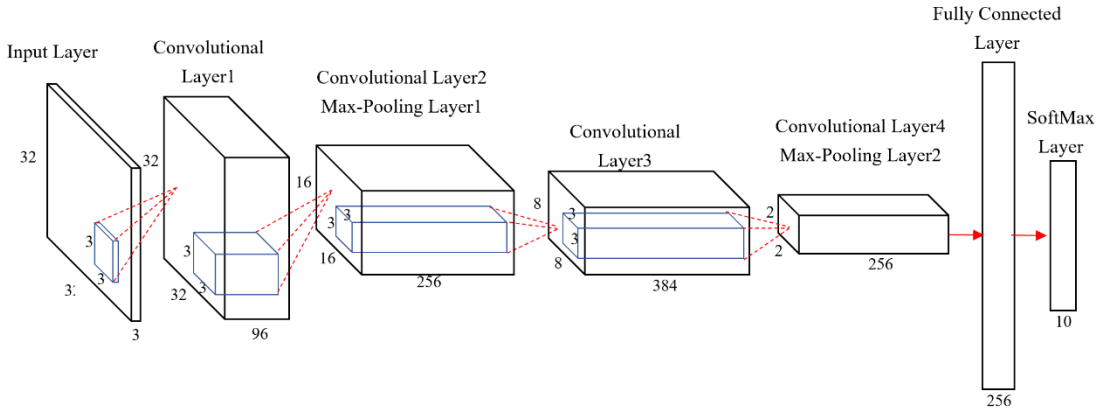

Figure 2. CNN Model 2 Structure Diagram

2.3.6. Model 3 – five convolution layers

The convolutional neural network in model 3 is a 5-layer convolutional layer, maintaining the same basic assumptions as model 1. The first and second convolution layers are the same as model two, Add a pooling layer and select the MaxPool2d function, where kernel size is 3, Similarly, the Input channels of the third layer of the convolution layer are 256, and the Output channels are 384, Add a pooling layer and select the MaxPool2d function, where kernel size is 3, The Input channels and Output channels of the fourth volume layer are 384 and 384 respectively, The Input channels of the last convolutional layer are 384, and the Output channels are also 256, Finally, add a pooling layer and select the MaxPool2d function, where kernel size is 3.The structural diagram of model 3 is shown in Figure 3:

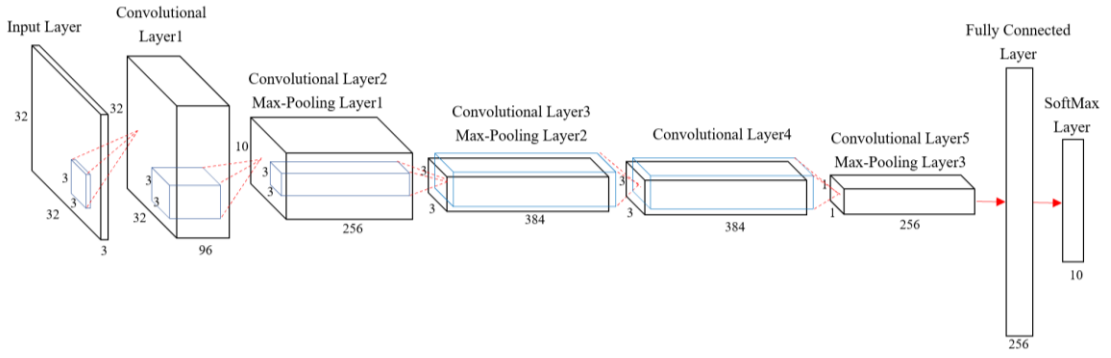

Figure 3. CNN Model 3 Structure Diagram

2.4. Model Results and Comparison

According to the above experimental results of different models that change the convolution layer (see Table 1), it can be seen that as the convolution layer continues to increase, the final calculated loss

DOI: 10.54254/2755-2721/15/20230831

continues to decline, and the accuracy rate continues to improve. However, the running time of the program is also increasing, and GPU is also used in this experiment. Without the use of GPU, the program ran for more than 2 hours, which took too long to use. GPU was used in this experiment [10].

Therefore, convolutional neural networks should specifically select several layers of neural networks, not only considering the accuracy but also considering the time consumed. Improving accuracy while reducing time is still a worthwhile research direction.

Table 1. Prediction results of different convolution layer models

Model	Accuracy	Times	Loss
CNN_3	68.54%	11minutes	0.071
CNN_4	76.68%	15minutes	0.016
CNN_5	81.37%	23minutes	0.004

Changing the descent algorithm based on Model 3 can obtain experimental results (see Table 2 and Figure 4). Here, five different optimizers are selected [11], and it can be found that changing the optimization algorithm will change the accuracy and loss of the model. In addition, the accuracy rate of SGD is the lowest, while the accuracy rate of other optimizers is around 81%, with Adamax having the highest accuracy rate, up to 82.75%. By observing the Loss column, it can be seen that for most models, the lower the Loss value, the higher the accuracy rate. However, by observing Adadelta and SGD, it is found that the Loss values of both models are similar, with around 0.32, but the accuracy rate is significantly different. As shown in the figure, it can be clearly seen that different loss values continue to decline with the increase in training times. The declining trend of Adadelta and SGD loss values is basically the same, but when the epoch reaches about 10 times, it basically remains at 0.32 and no longer decreases. The remaining three optimizers maintain a steady downward trend.

Table 2. Prediction Results for Different Optimizer Models

Optimizer	Accuracy	Loss
Adam	81.37%	0.005
SGD	13.44%	0.328
Adamx	82.75%	0.001
Adadelta	82.01%	0.327
AdamW	81.68%	0.004

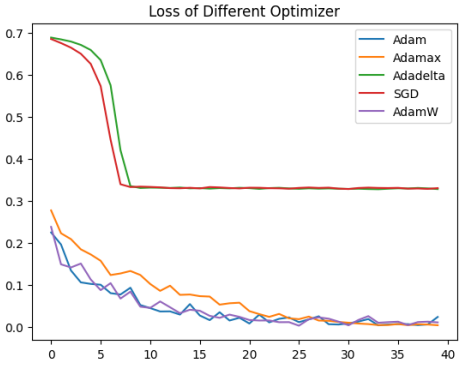

Figure 4. Loss of Different Optimizer (CNN_5)

3. Conclusion

This article first gives a bird's-eye view of the development of image classification and briefly introduces the factors that affect the accuracy of CNN. Next, this article conducted experiments and comparative

DOI: 10.54254/2755-2721/15/20230831

analysis on the impact of convolution layers and optimizers on CNN accuracy. Three convolutional neural networks with different convolution layers were built, their structural design was introduced in detail, and their structural diagrams were given. Different models were used to classify images from the CIFAR 10 dataset. Finally, the accuracy results under different convolution levels and optimizers are given and compared. Research has found that as the volume layer continues to increase, the final calculated loss continues to decline, and the accuracy rate continues to improve. However, the running time of the program is also increasing, so there is still much room for improvement. In addition, it was also found that using a five-layer convolutional neural network for the CIFAR 10 dataset and using different optimizers can lead to changes in the accuracy rate, with Adam, AdamW, and Adamax having the highest accuracy rate and the best prediction effect.

In fact, the model still has many shortcomings. Firstly, with the increase of the convolutional layer, the running time of the program increases, which is an inevitable problem for many neural network models. Current solutions mostly use local or cloud GPUs. Another problem that has always been faced is the selection of optimizers, which is often compared through some experience and continuous parameter adjustment. Currently, there is no good method to select the appropriate optimizer and parameters for a specific dataset or model. These are all possible directions for future development, which are worth our in-depth consideration and exploration.

References

[1] LeCun Y, Bottou L, Bengio Y and Haffner P 1998 Gradient-based learning applied to document recognition. *Proceedings of the IEEE.* 86(11):2278-2323.

[2] Hinton GE and Salakhutdinov RR 2006 Jul Reducing the dimensionality of data with neural networks. *Science (New York, N.Y.)*.313(5786):504-507.

[3] Krizhevsky A, Sutskever I and Hinton G E 2012 ImageNet classification with deep convolutional neural networks. *Communications of the ACM, 60*, 84 - 90.

[4] Tripathy S and Singh R 2022 Convolutional Neural Network: An Overview and Application in Image Classification. *Advances in Intelligent Systems and Computing*.

[5] Chaganti S Y, Nanda I, Pandi K R, Prudhvith T G N R S N and Kumar N 2020 Image Classification using SVM and CNN, *2020 International Conference on Computer Science, Engineering and Applications (ICCSEA)*, Gunupur, India, pp. 1-5.

[6] Nielsen M A 2015 Neural networks and deep learning[M]. San Francisco, CA, USA: Determination press.

[7] Kingma D and Ba J 2014 Adam: A Method for Stochastic Optimization[J]. *Computer Science.*

[8] Ruder S 2016 An overview of gradient descent optimization algorithms[J].

[9] Krizhevsky A 2009 Learning Multiple Layers of Features from Tiny Images.

[10] László E, Szolgay P and Nagy Z 2012 Analysis of a GPU based CNN implementation, 2012 13th International Workshop on Cellular Nanoscale Networks and their Applications, Turin, Italy, pp. 1-5.

[11] Ruder S 2016 An overview of gradient descent optimization algorithms.