An Application of Machine Learning Algorithms on the Finger Image Prediction

Yifei Liu[*]

Big Data Finance Experimental class of Chumin College, Shanxi University, Taiyuan, Shanxi Province, China

ABSTRACT: In recent years, using machine learning (ML) algorithm to analyze a picture, obtain its features, and finally identify what the picture is about is getting more and more important. This is because with the popularity of the electronic equipment and high-performance computing equipment, people began to pursue a more convenient and automatic life. The science of the image recognition frees people's hands to a certain extent through the training of algorithms, thus making people's life more convenient. This paper presents a comparison of two ML algorithms: Multi-layer Perceptron (MLP), and Convolutional Neural Network (CNN) with three different optimization methods on the data-set by measuring their test accuracy and their running time. The said data-set consists of a training-set of 1080 pictures (64 by 64 pixels) of signs representing numbers from 0 to 5 (180 pictures per number) and a test set of 120 pictures (64 by 64 pixels) of signs representing numbers from 0 to 5 (20 pictures per number). For the implementation of the ML algorithms, the data-set was partitioned in the following fashion: 90% for training phase, and 10% for testing phase. The hyper-parameters used for all the classifiers were manually assigned. Results show that most of the presented ML algorithms performed not bad with a test accuracy over 80%, and the CNN algorithm performed best among all the implemented algorithms with a test accuracy about 91.04%.

1. INTRODUCTION

Convolutional neural network (CNN) is a feedforward neural network. Its artificial neuron can respond to some surrounding elements in the coverage area, and has an excellent performance in the large-scale image processing. CNN, including the convolution layer and pool layer, is an efficient recognition method developed in recent years. In the 1960s, Hubel and Wiesel found that the unique network structure of neurons in the cat skin layer for local sensitivity and direction selection can effectively reduce the complexity of the feedback neural network [1]. Then CNN with convolution calculation and a deep structure was proposed. Generally speaking, CNN is composed of the input layer, convolution layer, pooling layer, full connection layer, and output layer. The input layer can process multi-dimensional data and use gradient descent algorithm for learning. The convolution layer can extract the features of the input data. After the feature extraction, the output feature map will be transferred to the pooling layer for feature selection and information filtering. The upstream of the output layer in the convolutional neural network is usually the full connection layer, so its structure and working principle are the same as that in traditional feedforward neural network. However, with the increase of CNN depth, the parameters and features learned by CNN are more clear and complete, so as to achieve a higher accuracy than traditional methods.

So the image recognition technology can develop fast and well, and obtain a huge success. In this research, the author uses different kinds of ML algorithms such as CNN and MLP, trying to get a better effect and a higher accuracy rate for the image identification. This paper applys the image recognition algorithms and different optimization algorithms to the same image database, intending to find the best algorithm by comparing their accuracy. Since using self-driving technology is becoming more and more common with the development of modern society, this research can help increase the accuracy rate of the picture identification and provide some help on the development of the picture identification technology.

2. METHOD

2.1. Machine Intelligence Library

Python's TensorFlow library and python's H5PY, Numpy, Math, and Matplotlib libraries are used to implement this algorithm.

2.2. The Dataset

The machine learning algorithms were trained to decide what kind of number it signs in the image using the finger image data-set. This dataset contains 1200 64 by

---

*Corresponding author. Email: Aptx486982512022@163.com

64 color gesture images that have been digitized. The gallery below contains six gesture types in this gesture dataset. There are 1200 finger images in the dataset.

Accordingly, the dataset features are as follows: (1) zero, (2) One, (3) two, (4) three, (5) four, (6) five.
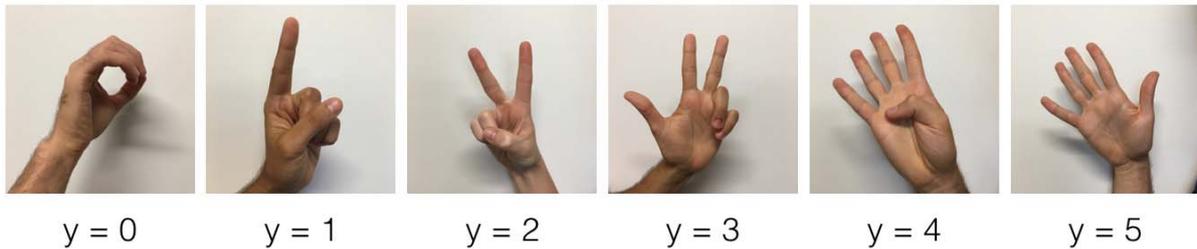

$y = 0$ $y = 1$ $y = 2$ $y = 3$ $y = 4$ $y = 5$

Figure 1 Six gesture pictures representing different numbers [12].

2.3. Dataset Preprocessing

To reduce the computation and improve the accuracy, in addition to flattening the dataset and dividing by 255 to normalize the data, there is also a need to convert each label into a unique heat vector. The formula is shown below.

X_train = X_train_flatten / 255

X is a three-channel gray matrix, and 255 is the Maximum gray level of the matrix. The one-hot encoding matrix is implemented using convert_to_one_hot of tensorflow. In addition, the whole matrix X is made by the formula below.

X_train_flatten=X_train_orig.reshape(X_train_orig.shape[0],-1).T

(Each column of the matrix is the gray data of a sample)

2.4. Machine Learning (ML) Algorithms

This section introduces the machine learning algorithms used in this article.

2.4.1. Convolutional neural network (CNN)

Convolutional neural network (CNN) is one of the most representative deep learning algorithms. Another name for CNN is "shift invariant artificial neural network (SIANN)", because CNN can represent, learn, and translate the invariant to classify the input information based on its hierarchical structure This structure obtains the relationship characteristics among the data in the range of "receptive field" through convolution kernel. In a picture, the adjacent pixels obviously have a stronger correlation. Compared with the fully connected neural network, CNN highlights the characteristics of this adjacent relationship, so it can obtain the useful information in the picture more accurately [2].

$$\text{n}_H = (n_{H_{prev}} - f + 2 \times pad) \div stride + 1 \tag{1}$$

Where f is the number of filters, and pad is the number of filters to fill in the matrix. Using the cost given by the cross entropy, the learning parameter W of CNN is propagated through the back propagation. Then the computational cost is minimized by Adam optimization. The optimization algorithm is also used for

MLP (section 2.4.2). When implementing the algorithm, the following steps and parameters are needed to be used:

Conv2d: Step: 1, filling mode: "SAME"

Activation: ReLU

Max Pool: Filter size: 8x8, pace: 8x8, padding method: "SAME"

Conv2d: Step: 1, filling mode: "SAME"

Activation: ReLU

Max Pool: Filter size: 4x4, pace: 4x4, filling method: "SAME"

Then One-dimensional the output of the previous layer

Full connection layer (FC): Use full connection layer without nonlinear activation function. Do not call SoftMax here, this will result in 6 neurons in the output layer, which will then be passed to SoftMax [3].

In TensorFlow, the Softmax and cost functions are grouped into one function, and different functions are called when calculating the cost.

2.4.2. Multilayer Perceptron (MLP)

Rosenblatt (1958) came up with the perceptron model on the basis of the MC CULLOCH & Pitts (1943) neuron model. It can be composed of multiple layers of perceptrons (such as MLP). MLP can have various hidden layersis apart from the input and output layers. It is also called the artificial neural network (ANN). The simplest MLP, which is referred to as the neural network later in this paper, contains the input layer, hidden layer, and output layer [4]. Generally speaking, the neural network is a technology based on the bionic neural network. Through the connection of various eigenvalues and a conbination of the linear and nonlinear, a goal can be achieved and then used to identify if the picture is a cat, a dog or which distribution it belongs to.

The activation function ReLU is used in this paper for MLP There are three hidden layers, and each layer contains different number of nodes in the forward propagation layer. Additionally, the cross entropy function is used to measure the cost.

2.5. Optimization algorithm

2.5.1. AdaGrad Optimization algorithm

The cumulative sum of squares of each parameter gradient in each iteration is calculated, and the cumulative value is divided by the basic learning rate to take the root, so as to realize the dynamic update of the learning rate. Therefore, not only the change of the gradient with time (the number of iterations), but also the change of the parameter state (the difference between the current gradient and the previous gradient) are taken into account by the dynamic update of the learning rate. At this time, the batch gradient update size is proportional to the primary gradient and inversely proportional to the sum of the secondary gradient, that is, the gradient update is based on the quadratic approximation formula [5].

Advantages:

In the early stage of training, the accumulated value is small and the gradient updating amplitude is large. In the late training period, the cumulative value increases and the gradient updating amplitude is small, which can realize the constraint on the updating amplitude of the learning rate. This stage is suitable for processing the sparse gradient (when the neuron enters the negative saturation region of the ReLU activation function, there is no update of the sparse gradient).

Disadvantages:

At the later stage of training, the cumulative value is extremely large, resulting in a rather small learning rate, which makes the model have basically no gradient update. There is still a need to customize the global learning rate.

2.5.2. RMS Optimization algorithm

Geoffrey E. Hinton came up with an optimization algorithm called Root Mean Square Prop (RMSProp), which solved the problem of a large swing amplitude preliminarily. In order to further optimize the overswing problem by updating the loss function, and to further accelerate the convergence function, the RMSProp algorithm uses the differential square weighted average weight W of the gradient and bias B, assuming t round is in the first iteration. The formulas are shown below [6].

$$s_{dw} = \beta s_{dw} + (1-\beta)dW^2 \tag{2}$$

$$s_{db} = \beta s_{db} + (1-\beta)db^2 \tag{3}$$

$$W = W - \alpha \frac{dW}{\sqrt{s_{dw}} + \varepsilon} \tag{4}$$

$$b = b - \alpha \frac{db}{\sqrt{s_{db}} + \varepsilon} \tag{5}$$

RMSProp adds an attenuation coefficient to control how much historical information to obtain, namely to set the global vector and make it divided by the global vector gradient parameters every time of the calculation. Through controlling the square root of the sum of squares by history of the attenuation coefficient, the influence of each parameter vector on the direction of different game parameter space is more moderate, thus improving the training speed.

2.5.3. Adam Optimization algorithm

Through the combination with the RMSProp and momentum algorithm, the first and the second moment estimation of gradient is used to adjust the learning rate of each parameter. After bias correction, the learning rate of each iteration has a certain range (dynamic constraint is carried out), making the parameters more stable [7].

Advantages:

AdaGrad performs well in handling the sparse gradients, while RMSProp in non-stationary targets. Moreover, AdaGrad algorithm has high computational efficiency and small memory requirement. It can calculate the adaptive learning rate for various parameters, making it quite fit for large datasets and high-dimensional space.

2.6. Data Analysis

Two phases of experiment, namely the training phase and the testing phase, are contained in this paper. The dataset is partitioned by 90% of training phase and 10% of testing phase. The parameters considered in the experiments are: Test Accuracy, Epochs, Learning rate, Tme which they take, Batch size, and Number of hidden layers.

3. RESULTS AND ANALYSIS

All experiments in this study were conducted on a laptop computer with Intel (R) Core (TM) i7-8550U CPU @ 1.80 GHz 1.99 GHz. Table 1 and 3 show the manually-assigned hyper-parameters used for the ML algorithms. Table 2 and 4 summarize the experiment results. Finally, this paper compares the test accuracy and consumption time of each combination to get a better one.

It can be clearly seen from Table 1 and Table 2 that if the Adam algorithm is used for optimization in convolutional neural network, it can reach the accuracy close to other algorithms in a short time. However, although the RMS algorithm takes longer time than the Adam algorithm, it can obtain higher accuracy than the Adam algorithm under the same hyper-parameter condition.

Compared with CNN algorithm, MLP algorithm has more training epochs, more time and a larger learning rate. On the contrary, the accuracy of the test set is 10 percentage points lower than the average accuracy of CNN. It can be seen from Table 3 and Table 4 that under the same super parameter conditions, the AdaGrad algorithm has higher accuracy and a better effect.

Table 1 Hyper-parameters used for the CNN algorithms.

Hyper-parameters	CNN (Adam optimizer)	CNN (Adagrad optimizer)	CNN (RMS optimizer)
Minibatch size	64	64	64
Learning rate	0.005	0.005	0.005
Epoch	60	60	60
Number of hidden layers	3	3	3

Table 2 Summary of experiment results on the CNN algorithms.

Parameter	CNN (Adagrad)	CNN (Adam)	CNN (RMS)	CNN (Adadelta)
Train accuracy	0.62685186	0.997	0.9982	0.15
Test accuracy	0.59166664	0.8916	0.9416	0.1916
The time it takes to complete a program	94.33509564399719 s	76.665 s	163.34	195.78
Loss of algorithm	1.5155144035816193	0.00911993844783865	0.004855376289924607	1.9598064944148064

Table 3 Hyper-parameters used for the MLP algorithms.

Hyper-parameters	MLP (Adam optimizer)	MLP (Adagrad optimizer)	MLP (RMS optimizer)
Minibatch size	64	64	64
Learning rate	0.009	0.009	0.009
Epoch	1300	1300	1300
Number of hidden layers	3	3	3

Table 4 Summary of experiment results on the MLP algorithms.

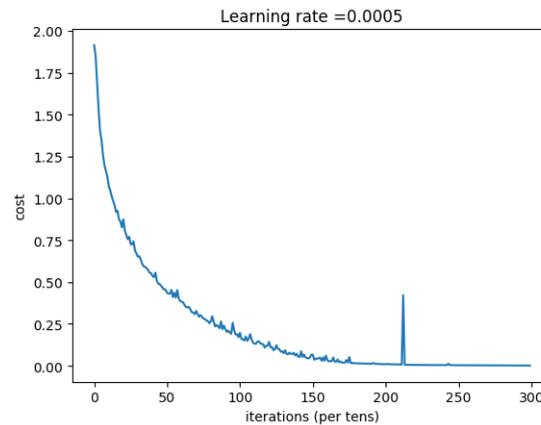| Parameter	MLP (Adagrad)	MLP (Adam)	MLP (RMS)	MLP (Adadelta)
Train accuracy	0.98796296	0.86	0.876	0.3361111
Test accuracy	0.84166664	0.737	0.752	0.775
The time it takes to complete a program	251.72092550000002 s	260.05442400000004 s	275.0544244 s	243.44839770000002

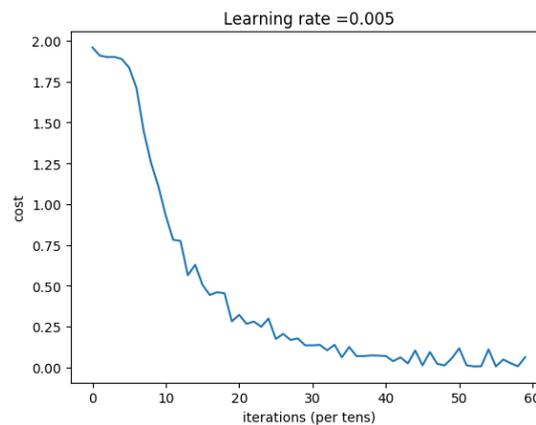**Figure 2** Cost of the MLP algorithms on finger image (Plotted using matplotlib).

Figure 3 The best effect of the loss function of the algorithm's cost and the test accuracy on finger image (Plotted using matplotlib).

After averaging the accuracy of test sets of different algorithms in Table 1, the average accuracy of CNN algorithm is 87.42%, while that of MLP algorithm d in Table 3 is 72.4%. At the same time, it can be seen from Figure 4 that the result obtained by the optimal combination of CNN algorithm is that the accuracy of the highest test set in this experiment is 94.166%, and the minimum time is 123.911s.

```
train accuracy: 1.0
test accuracy: 0.94166666
WARNING:tensorflow:From C:/Users/asus/PycharmProjects/pythonProjec
 .Saver instead.

Time taken by the CPU 123.91195797920227 s
```

Figure 4 The time and the test accuracy of the algorithm that works best.

4. DISCUSSION

The convolutional neural network uses original images as input. It has a high efficiency in learning the corresponding features from abundant samples, and can avoid the complexity in the process of feature extraction [8]. CNN has been widely applied in processing images due to its function of directly processing two-dimensional images, and many research achievements have been achieved. Through a simple nonlinear model, more abstract features are extracted from the original images, and only a small number of people are required to participate in the whole process [9].

CNN has a high efficiency in complexity reduction of the feedback neural network (traditional neural network). Common CNN structures include Lenet-5, AlexNet, ZFNet, VGGNet, GoogleNet, ResNet and so on. Among them, the network level of LVSVRC2015 champion ResNet is 20 times more than that of AlexNet, and 8 times more than that of VGGNet. As for these structures, the increase of layers can be regarded as one direction for CNN development. In this way, the approximate structure of the objective function can be obtained by using the increased nonlinearity, and a better feature expression can be obtained at the same time. However, this way leads to the increase of the overall complexity of the network, making the network more difficult to optimize and easy to over-fit [10].

The structure of CNN and traditional neural network is similar and different in various ways.

First, CNN includes the data input Layer, convolutional Layer, ReLU excitation Layer, pooling Layer, full connection Layer, and Batch Normalization Layer (which does not necessarily exist), while the traditional neural network mainly contains the data input layer, data output layer, and one or more hidden layers. Therefore, it can be seen that the hierarchical structure of traditional neural network is still applied by CNN. Besides, for CNN, different layers have different functions, whereas for traditional neural network, the linear regression is performed by each layer on the features of the previous layer, and then there is a nonlinear transformation. Moreover, CNN applies ReLU as the activation function (excitation function), whereas traditional neural networks applies Sigmoid. What is more, the pooling layer of CNN can reduce the dimension of data and extract the high frequency information of data, which is unachievable for traditional neural networks.

In addition, it is clear from Figure 2 and 3 that the time for CNN algorithm to reach the lowest cost is generally less than that of MLP, avoiding the disadvantage of MLP being easy to overfit [11]. At the same time, it is obvious from this experiment that the general image recognition accuracy obtained by CNN is a lot higher than the limit that multilayer neural network can achieve. Therefore, CNN algorithm is more suitable for gesture recognition in image recognition than MLP.

5. CONCLUSION

From the various optimization and prediction algorithms used in this experiment, it can be seen that CNN algorithm combined with Adam optimization algorithm can obtain better accuracy in a short time, but CNN algorithm combined with RMS optimization algorithm can obtain higher accuracy in a longer time. At the same time, it can be seen from the experimental data that the average accuracy of the test set that MLP algorithm can achieve under the condition of all possible optimization algorithms is much lower than that of CNN algorithm, so it can be known that the effect of CNN algorithm is generally superior to MLP algorithm in image recognition.

However, the types and quantities of the image data used in this experiment were not enough, which may lead to the universality of experimental results. At the same time, the running neural network does not have too many layers, which may also affect the experimental conclusion. In the later research, a deeper neural network can be used to recognize the image, and different optimization algorithms can be compared to select the best combination, so as to improve the accuracy and efficiency of the image recognition. Moreover, under the above conditions, more abundant image data can be used for prediction, so as to make the results more universal. At the same time, images with richer background can be used for recognition, so as to strengthen the anti-interference of the neural network prediction.

REFERENCES

1. J.M. Feng, F.A. Zhao, M. Ye, W.C. Sun, "The Auxiliary System of Cleaning Vehicle Based on Road Recognition Technology", SAE International, 2021.

2. J. Cai, J.Y. Cai, X.D. Liao, H.T. Huang, Q.J. Ding, Preliminary Study on Hand Gesture Recognition Based on Convolutional Neural Network, Computer system application, 24(04), 2015.

3. R. Tian, J.M. Bi, Q. Zhang, Y.X. Liu, "Research on Lane Occupancy Rate Forecasting Based on the Capsule Network", IEEE Access, 2020.

4. Abien Fred M. Agarap. "On breast cancer detection", Proceedings of the 2nd International Conference on Machine Learning and Soft Computing - ICMLSC '18, 2018.

5. Z.W. Jiang, Improvement of adaptive gradient descent method based on Neural Network. 10.27061/d.cnki.ghgdu.2021.001061.

6. Q.Y. Yuan, Research on training optimization method of deep neural network. 10.27151/d.cnki.ghnlu.2020.000020.

7. X. Wang, K. Li, C. Ning, F.C. Huang, Remote sensing image classification method based on deep convolution neural network and multi-core learning [J], Journal of electronics and information, 2019 (05).

8. P, VINCENT, H. LAROCHELLE, Y. BENGIO et al., Extracting and composing robust features with denoising autoencoders, Proc of the 25th International Conference on Machine Learning, 2008.

9. X.Y. Fan, W.G. Zhu, "Research on SAR Image Target Recognition Based on Convolutional Neural Network", Journal of Physics: Conference Series, 2019.

10. R. SALAKHUTDINOV, Learning deep generative models, Graduate Department of Computer Science, University of Toronto, 2009.

11. X.T. He, Z.C. Lian, Stability of spanning connected multilayer neural networks [J], Journal of Liaoning University of science and technology, 2008(02).

12. Chinese Software Developer Network, from: https://blog.csdn.net/u013733326/article/details/80086090.