Comparative analysis of accuracy of various neural networks and optimization algorithms in recognition of clothing items task

N A Andriyanov\textsuperscript{1,2}

\textsuperscript{1}JSC "RPC "Istok" named after Shokin", 2, Vokzalnaya st., Fryazino, Moscow Region, 141190, Russia
\textsuperscript{2}Ulyanovsk State Technical University, 32, Severny Venets st., Ulyanovsk, 432027, Russia

E-mail: nikita-and-nov@mail.ru

Abstract. The study conducted the investigation of the accuracy of neural networks trained with various parameters of the neural network model for the standard MNIST data set containing 70,000 images of clothing items belonging to 10 classes. In particular, the influence of the number of neurons and the number of layers on the recognition accuracy was studied, and dependences on indicators such as the number of training epochs and the optimization algorithm were obtained. In addition, a brief overview of existing solutions was performed to select optimization algorithms. According to the research results, it was found that the best estimates on the database used were provided by the ADAM optimization algorithm with the largest number of training epochs and the most complex network structure. At the same time, the least accuracy was provided by the optimization method based on simple gradient descent. Best results were obtained for a multi-layer network with ADAM optimization and k-fold cross-validation.

1. Introduction

The last few years are characterized by intensive research in the field of image processing. To solve the problems of describing and presenting images, as well as filtering them, the use of mathematical models of random fields has proven itself well \cite{1-4}. However, the use of such mathematical models to solve the recognition problem is very difficult, and can only be advisable if combined with other recognition methods. At the same time, neural network algorithms \cite{5,6}, which in many cases provide high accuracy, are especially popular in recognition problems. Unfortunately, neural networks are still insufficiently studied, but the general tendency of their development is such that neural networks are useful for a specific task. Moreover, the accuracy of solution depends on the data set being processed. In addition to this important parameter, the weights of the trained neural network also depend on the training model itself.

2. Optimization algorithms in neural networks

Let us consider some of the parameters. The simplest and most understandable parameter is the number of training epochs. It is also possible to understand the network structure, in the simplest case described by the number of layers and the number of neurons on each layer. One of the most important
parameters for training is the optimization method. Its meaning is to adjust the weight of the model according to some criterion so that it gives the best result. In this paper, for comparison, 3 following optimization methods were chosen:

1) Simple gradient descent (SGD): an iterative method for optimizing the objective function with suitable smoothness properties (for example, differentiable or subdifferentiable). This can be considered as a stochastic approximation of the gradient descent optimization, since it replaces the actual gradient (calculated over the entire data set) with its estimate (calculated from a randomly selected subset of the data) [7].

2) RMSProp method is also a method in which the learning rate is adapted to each of the parameters. The idea is to divide the training speed for the weight by the moving average of the last gradients for that weight [8].

3) Adaptive Moment Estimation (ADAM) is updated RMSProp Optimizer. This optimization algorithm uses moving averages of both gradients and second moments of gradients [9]. ADAM is an adaptive learning speed optimization algorithm designed specifically for training deep neural networks.

There are other optimization methods, however, in this paper, taking into account the computational and time costs of training neural networks, it is appropriate to restrict to the three presented. Based on the brief historical development of the above methods, the SGD algorithm is expected to lead to worse results, the RMSProp algorithm is expected to lead to average results, and ADAM is expected to lead to the best results.

3. Dataset description and network architecture
The study will be performed on the standard MNIST data set, which contains items of clothing belonging to the following 10 classes: T-shirt / top; trousers; pullover; dress; coat; sandals; shirt; sneakers; bag; ankle boots.

In total, 70,000 images of 28x28 pixels are taken from the database. For all neural networks, the initial data set for training and testing was divided in the ratio of 60,000 to 10,000. The training and testing of networks was performed in Python using Keras module and the GPU NVIDIA RTX-2070 SUPER. Figure 1 shows a neural network model in a general form. Here N1 is the number of outputs of the first hidden layer (the number of neurons), N2 is the number of outputs of the second hidden layer, if it is presented.

Figure 1. The basic model of the neural network
In Figure 1 on hidden layers the “Relu” activation function is used, which is a standard threshold function, and on the output layer a slightly more complex threshold “Softmax” function is used. Its application is justified by the fact that in this problem the classification is not binary.

4. Accuracy investigation results
Models were fitted using different optimization methods and a different number of training epochs. Tables 1-3 show recognition accuracy values for the SGD, RMSProp, and ADAM algorithms, respectively.

| Table 1. Recognition using SGD |
|-----------------------------|
| CNN Neurons\Epochs | 1 | 5 | 10 | 20 |
|---------------------|---|---|---|---|
| 16                  | 0.7717 | 0.8365 | 0.8457 | 0.855 |
| 64                  | 0.8 | 0.8379 | 0.8544 | 0.8674 |
| 128                 | 0.7971 | 0.8426 | 0.8521 | 0.8635 |
| 2 layers with 128   | 0.81 | 0.838 | 0.8653 | **0.8725** |

| Table 2. Recognition using RMSProp |
|-----------------------------|
| CNN Neurons\Epochs | 1 | 5 | 10 | 20 |
|---------------------|---|---|---|---|
| 16                  | 0.8286 | 0.8579 | 0.8605 | 0.8583 |
| 64                  | 0.8326 | 0.871 | 0.8671 | 0.871 |
| 128                 | 0.8556 | 0.858 | 0.8706 | 0.8721 |
| 2 layers with 128   | 0.8451 | 0.858 | **0.8799** | 0.875 |

| Table 3. Recognition using ADAM |
|-----------------------------|
| CNN Neurons\Epochs | 1 | 5 | 10 | 20 |
|---------------------|---|---|---|---|
| 16                  | 0.8245 | 0.848 | 0.8579 | 0.863 |
| 64                  | 0.8437 | 0.873 | 0.8801 | 0.8787 |
| 128                 | 0.8364 | 0.8713 | 0.8827 | 0.8875 |
| 2 layers with 128   | 0.8443 | 0.872 | 0.882 | **0.8906** |

An analysis of the presented results allows us to conclude that the neural network trained using ADAM optimization provides the best recognition accuracy. You can also highlight some trends of increasing accuracy such as an increase in the number of training epochs (not in all cases), an increase in the number of neurons and layers (not in all cases). For example, for the RMSProp algorithm, the best result was shown by a two-layer network trained over 10 epochs. Thus, the theoretical expectation that the ADAM optimizer will have better grades coincided with the results of the experiment.
Figure 2 shows an example of data processing based on a single-layer network consisting of 16 neurons and trained for 1 epoch with the RMSProp optimization algorithm.

It can be seen that the neural network determines the probability of belonging to each class and gives out the class for which this probability has a maximum value. To the right of each image there are the probability data, which can be examined in more detail as Figure 3 shows using the example of the first image.

Figure 3. The probability distribution of belonging to a particular class

Analysis of figure 3 shows that the probability for ankle boots and sneakers is approximately equal; however, the first is still some advantage. Another option for the network was sandals, which scored about 2 times less than the first two categories.
Finally, let us check how much it is possible to improve the results if, on the basis of Tables 1-3, the best ways to increase accuracy will be selected. For this, a neural network with ADAM optimization was trained during 100 epochs. Figure 4 shows the network model.

For the network shown in Figure 4, the probability of correct recognition was $p = 0.8923$. This result is only 0.17% higher than that of the previously best trained network.

It is possible to improve the network performance by applying a non-standard cross-validation method. An example of such a method is k-fold cross-validation [10]. Figure 5 shows the learning process of such a network with ADAM optimization during 25 epochs with 64 neurons in two convolutional layers and 100 neurons in a fully connected layer. The average probability of correct recognition in the test sample is $\bar{p} = 0.907$ and the standard deviation of this probability is $\sigma_p = 0.00326$. However, for the case $k = 5$, five learning operations during 25 epochs were performed sequentially. Orange lines indicate loss and accuracy for validation data and blue lines indicate loss and accuracy for train data. Axes X show the number of epochs.

Figure 4. Improved Neural Network

Figure 5. K-fold cross-validation learning process
An analysis of the curves shows that retraining of such a network is fast enough. At the same time, it is possible to increase the accuracy by almost 1.5% with the best of the previously considered models. However, this requires an increase in computational costs by a factor of $k$.

5. Conclusion
Thus, the comparative analysis of the various optimization algorithms such as SGD, RMSProp and ADAM was discussed. The latter provided the best results. In addition, a tendency has been established to increase the probability of correct recognition with an increase in the number of learning epochs, as well as the number of layers and neurons in the network. In addition, the use of k-fold cross validation also improves model accuracy. In general, it is possible to conclude that the training of neural networks remains an interesting scientific task.

6. Acknowledgments
The reported study was funded by the Russian Foundation for Basic Research (RFBR) and Ulyanovsk Region, Project No. 19-47-730011

References
[1] Lindgren F, Rue H, Lindström J 2011 Statistical Methodology. Series B 73(4) 423-498
[2] Dietterich T G, Hao G, Ashenfelter A 2008 Journal of Machine Learning Research 9 2113-2139
[3] Andriyanov N A, Dementiev V E, Vasiliev K K 2019 Pattern Recognition and Image Analysis (Advances in Mathematical Theory and Applications) 29 10-20 DOI: 10.1134/S1054661819010048
[4] Vasil’ev K K, Andriyanov N A 2020 Computer Vision in Advanced Control Systems-5. Intelligent Systems Reference Library 175. 11-52 DOI:https://doi.org/10.1007/978-3-030-33795-7_2
[5] Schmidhuber J. 2015 Neural Networks 61 85–117
[6] Ciresan D, Ueli M, Jonathan M, Gambardella M, and Schmidhuber J 2011 Proceedings of the Twenty-Second International Joint Conference on Artificial Intelligence 2 1237–1242
[7] Bottou L, Bousquet O 2012 Optimization for Machine Learning. Cambridge: MIT Press. 351–368
[8] Duchi J, Hazan E, Singer Y 2011 Journal of Machine Learning Research 12 2121–2159
[9] Diederik P and Ba J L 2014 arXiv:1412.6980v9
[10] Jung Y, Hu J 2015 Journal of Nonparametric Statistics 27 1-13 DOI: 10.1080/10485252.2015.1010532