Dean&Francis

ISSN 2959-6157

Classification of Fashion-MNIST Dataset by Using Lenet-5

Jiahao Guo *

Department of Computer Science, Shanghai University, Shanghai, China

*Corresponding author: 13681664265@shu.edu.cn

Abstract:

This study explores the application of LeNet-5 convolutional neural network (CNN) on the Fashion-MNIST dataset to analyze its effectiveness in complex tasks using image recognition. The Fashion-MNIST dataset, comprising 70,000 28x28 pixel clothing images across 10 distinct categories, is constituted to offer a challenging benchmark for image recognition research. In the experiment, the data is first normalized and the labels are preprocessed using the one-hot encoding technique to adjust to the model's input specifications. The LeNet-5 network combines several fully connected, pooling, and convolutional layers before using the Softmax layer to produce the probability distribution for each category. Through a series of experiments, it is found that LeNet-5 achieves 95% training accuracy and 90% validation accuracy on the Fashion-MNIST dataset, which not only confirms its effectiveness in modern image recognition tasks, but also demonstrates its good generalization ability. The study's findings provide new insights for the future application of traditional CNN in the field of deep learning and provide new perspectives for future applications in more complex image recognition tasks. In addition, this study also provides the possibility for the application of LeNet-5 in scenarios with limited resources or high real-time requirements, providing valuable experience and reference for subsequent research.

Keywords: LeNet-5; Fashion-MNIST; Image Recognition; Convolutional Neural Networks.

1. Introduction

In the field of computer vision, image recognition has become an extremely active research topic since deep learning technologies are developing so quickly. Among many image recognition tasks, handwritten digit recognition has been receiving extensive attention from researchers due to its wide application prospects and research value. The traditional MNIST dataset, as a benchmark test in this field, has been widely studied and achieved remarkable results [1-4]. However, with the deepening of research, there is a growing demand for more complex and diverse datasets in both academia and industry. In this context, the Fashion-MNIST dataset came into being. It not only contains the same number and size of images as MNIST, but also has richer and more diverse image content, covering 10 different clothing categories, providing new challenges for image recognition research, and has also attracted widespread attention from researchers [5-8].

LeNet-5, one of the early representatives of CNNs, was proposed by LeCun et al [9]. in 1998 and achieved breakthrough results in handwritten digit recognition tasks [10-12]. The development of subsequent deep learning models has been greatly influenced by its simple yet effective network structure. Although LeNet-5 was originally designed for recognizing handwritten digits, its powerful feature extraction capability and generalization performance make it equally potentially applicable to other image recognition tasks. Therefore, the aim of this paper is to explore the effectiveness of LeNet-5 when applied to the Fashion-MNIST dataset, with the intention of providing new solutions for complex image recognition tasks.

2. Methods

This article explores a study on the Fashion-MNIST dataset's handwritten number identification capability using LeNet-5 CNN, one of the early CNN architectures that has greatly affected the modern deep learning models's development. The study presents a detailed introduction to the LeNet-5 model's structure and implementation method, and evaluates its recognition performance through training and testing on the Fashion-MNIST dataset. The LeNet-5 network structure is detailedly described below and is primarily composed of the following components: a pooling layer S2, a fully connected layer F6, an output layer, a convolutional layer C1, a convolutional layer C3, a convolutional layer C5, and an input layer.

Mathematically, a single layer convolution operation can be expressed as:

$$O_{i,j,k} = b_k + \sum_m \sum_n (I_{m,n} * K_{i,k})_{m,n} \quad (1)$$

Where $O_{i,j,k}$ represents the k-th filter's response at position (i, j) on the output feature map O is indicated. $I_{m,n}$ is the value of the pixel on the input image I at position (m,n). $K_{i,k}$ is the weight corresponding to the k-th filter at position (i, j). b_k is the k-th filter's bias term. $*$ denotes the convolution operation, which is essentially a pointwise multiplication between a local area of the input image and the filter.

The operation of pooling is defined as:

$$P_{i,j,k} = \max_{m,n \in window} ?\left(O_{m+i,n+j,k}{}'\right) \quad (2)$$

Where $P_{i,j,k}$ represents the k-th filter's response on the pooled feature map P at position (i, j). $O_{m+i,n+j,k}{}'$ is the value after the activation function has been applied at position (m,n) within the pooling window on the original feature map.

The main input data of the input layer is a handwritten digital image of 28x28 pixels, the image is in grayscale i.e. each pixel value is between 0 and 255. Convolutional layer C1 contains 6 convolutional kernels, each kernel is of size 5x5 pixels and padding = 2. The size of the output feature map is 28x28. A 2x2 pooling window with a stride of 2 is utilized by the pooling layer S2 for average pooling. After pooling, the feature map's dimensions are lowered to 14x14. There are 16 convolutional kernels in convolutional layer C3, and each convolutional kernel is 5x5 pixels in size. The size of the output feature map is 10x10. Pooling layer S4 uses a pooling window of 2x2 with a step size of 2. After pooling, the feature map's dimensions are further reduced to 5x5. Convolutional layer C5, which contains 120 convolutional kernels, has each kernel sized 5x5 pixels, with a step size of 2 and padding=0. The output feature map's size is 1x1, i.e., each feature map is a single value. This layer spreads its output into a vector of 120 dimensions. Layer C5 sends the 120-dimensional vector to the 84-neuron completely linked layer F6. In the output layer, there are 10 neurons that correspond to the categories of numbers ranging from 0 to 9. The input is converted into a probability distribution with each category corresponding to a probability value by using the Softmax activation function.

Since the Fashion-MNIST dataset contains 10 different categories, each representing a type of clothing, this experiment uses solo thermal coding for data preprocessing because solo thermal coding converts integer target labels into a binary matrix.

3. Results

Fashion-MNIST is an alternative to the traditional MNIST handwritten digit recognition dataset for fashion product images. There are 70,000 28x28 pixel grayscale images in this dataset, with each of the 10 categories of clothing containing 7000 samples. Fig. 1 shows some sample images from the dataset.

ISSN 2959-6157

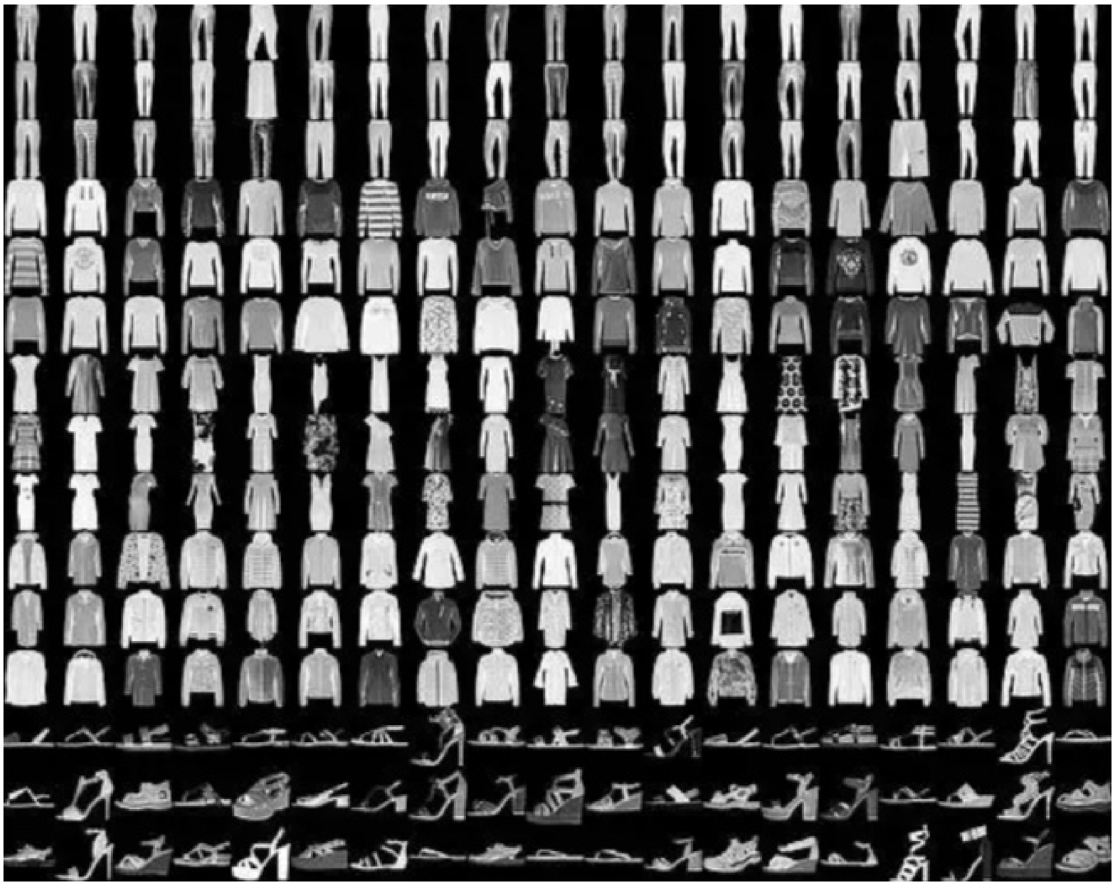

Fig. 1 Example Images in Fashion-MNIST Dataset

A baseline dataset for beginners in computer vision and machine learning is intended to be served by the Fashion-MNIST dataset, which is suitable for testing and training different systems for processing images. The validation set in this paper's trials consists of 20% of the original dataset.

In this paper, the handwritten digit recognition task on the MNIST dataset is handled by the classical LeNet CNN model. According to experimental results, the model performs exceptionally well both during training and validation.

The model's loss function values drop as it is being trained, as Fig. 2 shows. Specifically, the loss of training and the loss of validation both gradually decrease from higher initial values until they stabilize. The loss of training decreases slightly faster than the loss of validation, indicating that, on the training data, the model fits better. It is worth noting that the decreasing trend of the validation loss remains stable despite the gap between the two, which indicates the model's great generalization ability and that it does not suffer from overfitting.

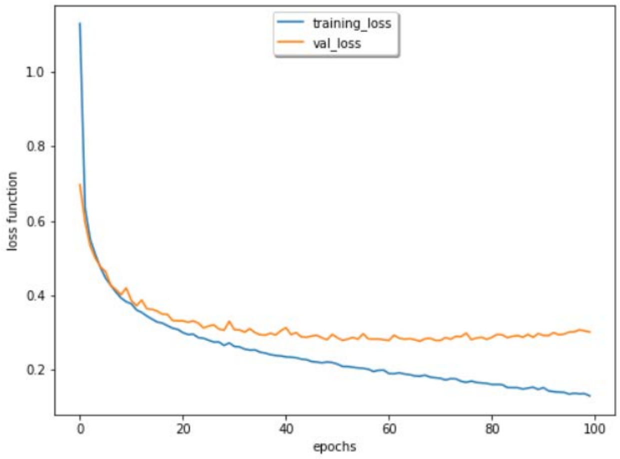

Fig. 2 Epoch-wise Training and Validation Loss Convergence

Fig. 3 shows the change in the model's accuracy on training set and validation set. It shows that training's accuracy and validation's accuracy both increase with the increase in the number of training cycles. The accuracy of training eventually stabilizes at around 95%, while the accuracy of validation also stays above 90%. This result shows that when using training data, the model performs well, and it demonstrates strong generalization ability on other unseen data too.

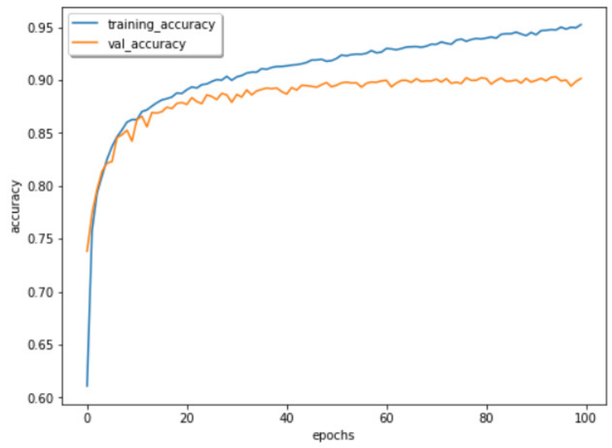

Fig. 3 Epoch-wise Training and Validation Accuracy Progression

The model evaluation results in a composite score on the testing set. The model's performance on the testing set further demonstrates its supremacy. The loss of testing and the accuracy of testing are 0.2 and 95%, respectively, and these measurements reach the field's existing state-of-the-art.

4. Summary

This paper comprehensively evaluates the performance of the classical LeNet-5 CNN in complex image recognition tasks through its application to the Fashion-MNIST dataset. The results show that LeNet-5 exhibits excellent recognition ability on the Fashion-MNIST dataset, and the loss function's continuous decrease during training and validation, as well as the significant improvement of the training accuracy and validation accuracy, validate the model's excellent learning ability and generalization capability. In addition, the test set further confirms the model's effectiveness and reliability in practical applications through its excellent performance.

This study not only provides an effective solution for image recognition on the Fashion-MNIST dataset, but also provides a strong demonstration of the potential of LeNet-5 to be applied in modern image recognition tasks. Although LeNet-5 is a relatively old model, its performance in this study shows that it can still be comparable to modern models in certain tasks. This finding opens up the possibility of future applications of LeNet-5 in scenarios with limited resources or high real-time requirements.

In conclusion, this study's findings expand the field in which LeNet-5 can be applied while also offering invaluable knowledge and resources for future investigations into increasingly intricate picture identification problems. Future work can further explore the combination of LeNet-5 with other modern deep learning techniques and the application effects on larger and more complex datasets, with a view to promoting the further development of image recognition technology.

References

[1] El-Sawy A, El-Bakry H, Loey M. CNN for handwritten arabic digits recognition based on LeNet-5. Proceedings of the International Conference on Advanced Intelligent Systems and Informatics 2016 2. Springer International Publishing, 2017: 566-575.

[2] Wang Y, Li F, Sun H, et al. Improvement of MNIST image recognition based on CNN. IOP Conference Series: Earth and Environmental Science. IOP Publishing, 2020, 428(1): 012097.

[3] Kumar A, Singh I, Singh B. Classification of MNIST Dataset Using Different CNN Architecture. Intelligent Circuits and Systems for SDG 3–Good Health and well-being, 2024: 273-280.

[4] Kadam S S, Adamuthe A C, Patil A B. CNN model for image classification on MNIST and fashion-MNIST dataset. Journal of scientific research, 2020, 64(2): 374-384.

[5] Nocentini O, Kim J, Bashir M Z, et al. Image classification using multiple convolutional neural networks on the fashion-MNIST dataset. Sensors, 2022, 22(23): 9544.

[6] Bin Tareaf R, Korga A M, Wefers S, et al. Applying XAI in Fashion-MNIST: Evaluating Hybrid Deep Neural Networks and Capsule Networks. Proceedings of the 2024 16th International Conference on Machine Learning and Computing. 2024: 100-107.

[7] Haji L M, Mustafa O M, Abdullah S A, et al. Enhanced Convolutional Neural Network for Fashion Classification. Engineering, Technology & Applied Science Research, 2024, 14(5): 16534-16538.

[8] LEITHARDT V. Classifying garments from fashion-MNIST dataset through CNNs. Advances in Science, Technology and Engineering Systems Journal, 2021, 6(1): 989-994.

[9] LeCun Y, Bottou L, Bengio Y, et al. Gradient-based learning applied to document recognition[J]. Proceedings of the IEEE, 1998, 86(11): 2278-2324.

[10] Zhang J, Yu X, Lei X, et al. A novel deep LeNet-5 convolutional neural network model for image recognition. Computer Science and Information Systems, 2022, 19(3): 1463-1480.

[11] Wei G, Li G, Zhao J, et al. Development of a LeNet-5 gas identification CNN structure for electronic noses. Sensors, 2019, 19(1): 217.

[12] Kayed M, Anter A, Mohamed H. Classification of garments from fashion MNIST dataset using CNN LeNet-5 architecture. 2020 international conference on innovative trends in communication and computer engineering (ITCE). IEEE, 2020: 238-243.