A Study on Real-Time Handwritten Digit Recognition Leveraging a Modified Convolutional Neural Network

Jian Guang

Jincheng College, School of Computer and Software, Sichuan University, Chengdu, Sichuan 611731

Abstract: ***Handwritten digit recognition is a technique in which an identification program interprets and recognizes handwritten numeric digits and returns the corresponding recognition results. With the rapid advancement of deep learning, recognition technologies—including digit recognition, Chinese character recognition, and English character recognition —have undergone significant development. Digit recognition, in particular, has a wide range of application scenarios, such as enabling batch scoring in the education sector or facilitating automatic statement importation in the financial domain. This paper provides a detailed introduction to the fundamental concepts and key architectural components of the LeNet-5 network, and introduces the concept of dilated convolution as a means of network improvement. Using the well-established MNIST handwritten digit dataset, both the original and improved models are trained. Finally, a visual interface is employed to enable simultaneous real-time digit writing and recognition.***

Keywords: Deep Learning; MNIST; LeNet-5 Convolutional Divine Network; The empty convolution.

1. 1INTRODUCTION

The emergence of LeNet-5 Divine Networks can be said to be the beginning of convolutional Divine Networks. Until now, the basic structure of the input layer, the convolution layer, the incentive layer, the pooling layer, and the full connection layer was still used. Because the structure is simple and easy to use, the utility model is widely used by the Volkswagen. Digital identification is divided into canonical print digital identification and uncanonical handwriting digital identification. The former utilizes the immutable geometric shape of numbers, and can easily extract the feature value to achieve identification. But the shape of handwritten numbers varies from person to person. Take a five. Some people prefer to finish one pen, others prefer to finish two pen, so that the extracted features can identify whether it is a five or a three. So a method with a high recognition rate is needed. Convolutional Divine Network can be used to improve the accuracy of handwritten digit classification, so this paper uses mature LeNet-5 Divine Network to complete handwritten digit recognition. Li et al. [1] distinguish between substantive and strategic green innovation effects within the pilot policy promoting technology-finance integration. In the realm of intelligent systems, Bi et al. [2] design a financial risk control platform using big data and deep machine learning, while Bi et al. [3] examine the potential and challenges of AI like ChatGPT in financial forecasting. Shifting to customer and market analysis, Zhou [4] investigates hierarchical needs in US automotive feedback, and Wensi [5] discusses AI-enabled data visualization marketing for automated production lines. For transaction security, Ximeng and Yiming [6] apply offline conservative RL to balance fraud risk and customer friction, whereas Zhao et al. [7] optimize deep learning models for dynamic market behavior prediction. Further extending risk monitoring, Yang et al. [8] construct credit-related transaction risk maps using graph neural networks. In object detection and transfer learning, Wu et al. [9] focus on small-sample crack detection in concrete structures, Ren [10] adapts YOLOv8 for infrared-visible fusion, and Tian et al. [12] improve brain tumor segmentation with GSConv and ECA attention. Financial fraud detection also benefits from metaheuristics, as Shen et al. [11] apply the whale optimization algorithm, while Yi [13] designs real-time fair-exposure ad allocation for small businesses using contextual bandits. Lastly, Tang et al. [14] contribute to photonic device design via shallow-angle grating couplers for indium phosphide devices.

2. SUMMARY OF THE DATA SET

This paper uses the most commonly used handwritten digit dataset, MNIST, which is collected by the National Institute of Standards and Technology (NIST). With 60,000 training sets and 10,000 test sets labelled 0-9, the data came from 250 different people's handwritten numbers, and 50% of the data came out of high school students, given that students' words were the most variable. The size of the digital sample image is unified to 28 × 28, and is

in the center of the position. This is more conducive to doing digital recognition training and testing. The figure below shows.

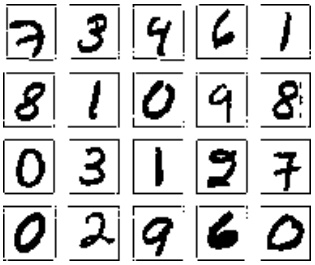

Figure 1: Partial sample of the MINIST dataset

3. INTRODUCTION TO CONVOLUTIONAL NEURAL NETWORKS

CNN (Convolutional Neural Network) network, also known as convolutional neural network. It is usually composed of convolutional layer, activation layer and pooling layer [2]. The input of a CNN network is usually a picture. The convolution layer extracts a feature matrix of the image from the picture. The model then transfers this feature matrix to the full connection layer as input. The full connection layer calculations are used to obtain a picture-to-label mapping value.

3.1 The convolutional layer principle

An important concept in CNN is convolution. The most important concept in convolution is the convolutional nucleus. In the field of imagery, convolutional nuclei are called discrete two-dimensional filters, which are equivalent to a second-order matrix, usually sizes 3 × 3, 5 × 5, etc.The convolution operation between the image and the convolution kernel is to slide the convolution kernel from the top left corner of the image by the stride, and at the same time do the inner product operation, put the obtained value into the new matrix, that is, the feature matrix [3]. As shown in figure 2. It is widely used in the field of image processing. Using different convolution kernels to the same image will result in different characteristic matrices.

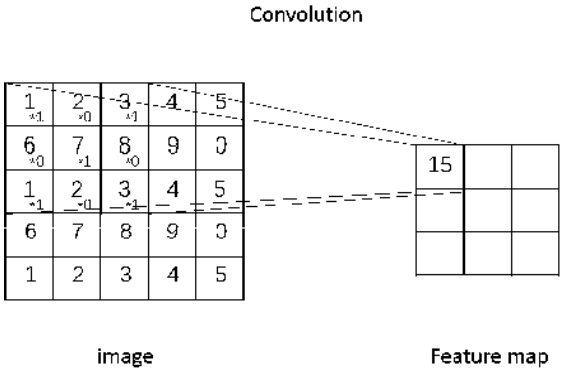

Figure 2: Example of Convolutional Operations

3.2 LeNet-5 Network

LeNet-5, one of the earliest convolutional neural networks, was developed by LeCun in the 1990s to solve the problem of number recognition on Bank Of America cheques [4].LeNet-5 network can be divided into 5 layers, which are convolutional pooling layer 1, convolutional pooling layer 2, fully connected layer 1, fully connected layer 2, and fully connected layer 3 [5]. Figure 3 shows:

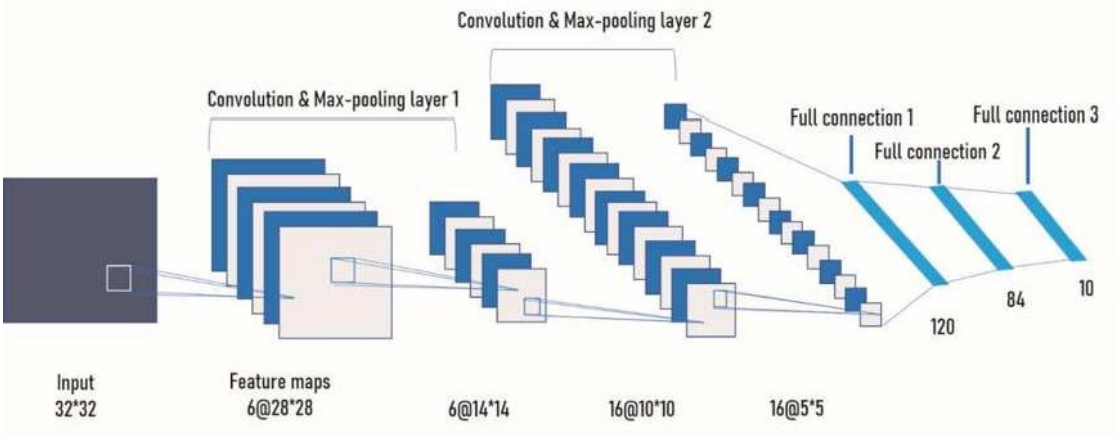

Figure 3: LeNet-5 Network Structure

There are a total of five layers in Figure 3, which do not include the input layer. The following describes the layers in detail:

The first layer is convolution pooling layer 1. This layer consists of one convolution and one pooling. The convolution layer uses 6 5 × 5 convolution cores, with a stride of 1, and sliding inner product on a 32 × 32 picture filled with 2 layers at the edge, to get 6 28 × 28 feature maps. The pooling layer used 2 × 2 nuclei, with step length 2 for character deviation (maximum pooling), resulting in 6 14 × 14 feature maps.

The second layer is convolution pooling layer 2. This layer will do the second convolution and pooling operation. This layer uses 16 convolutional nuclei of 5 × 5 with step length 1, convolutionated on 14 × 14 feature maps to form 16 10 × 10 feature maps. The pooling layer still uses 2 × 2 nuclei with a step length of 2 to be pooled, resulting in 16 5 × 5 feature maps.

The third layer is fully connected layer 1, flattening the results of convolutional pooling layer 2, which has 120 neurons, and fully linked 84 neurons of fully bound layer 2.

The fourth layer is fully connected layer 2, which has 84 neurons each connected to 120 neurons in the previous layer, so with a partial item, there are 84 × 120 + 1 = 10164 power value parameters.

The fifth layer is fully connected layer 3, mapping 84 neurons to 10 neurons to generate the corresponding probability of 0-9 for the 10 numbers. The maximum value is the result of recognition.

4. IMPLEMENTATION OF HANDWRITTEN NUMBERS IN REAL TIME

4.1 Training with LetNet-5

The loss function is used to measure the size of the difference between the predicted value and the true value. The LetNet-5 network uses a cross-entropy loss function (CrossEntropyLoss). This article uses LetNet-5 to train the MINIST training set of 60,000 data with batch _ size 64 and to validate 10,000 validation data with the same batch size. The resulting loss value curve and the accuracy curve are shown below. It can be seen that both the loss value and the accuracy value began to converge around the 10th epoch, and the final accuracy value stabilized at about 98.70%. Thus, the LeNet-5 network to solve the problem of efficiency can be said to be very good. But to keep improving, this article will make some modifications to the LeNet-5 network to get higher accuracy.

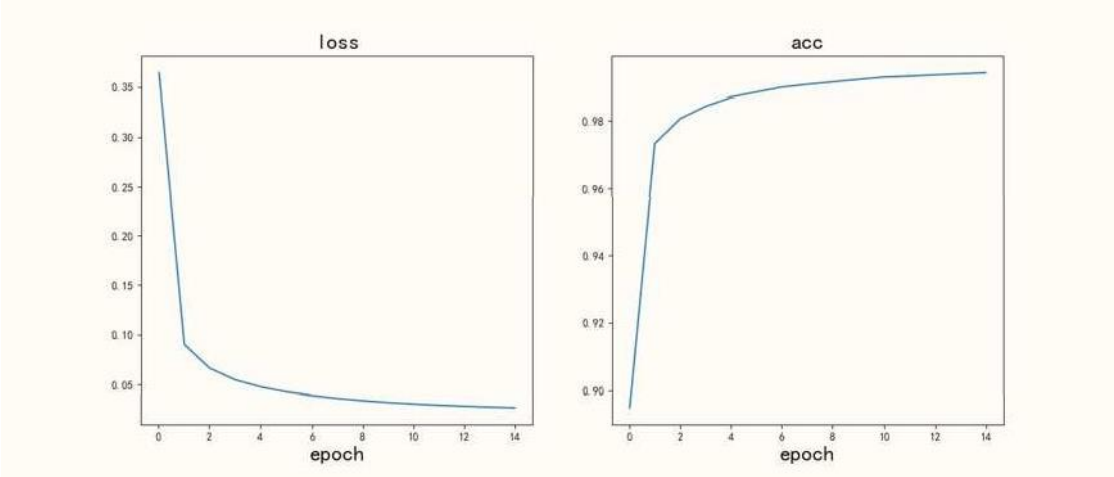

Figure 4: The loss value and the accurate value curve

4.2 Optimizing the LeNet-5 Network

As you can see from the MINIST data set above, most of the numbers are in the center of the picture, and the resulting feature matrix also contains a lot of background values, which can create a huge computational load if larger convolutional nuclei are chosen. Therefore, this paper adopts an important concept in the field of image restoration & mdash; An empty convolution.

Dilated convolution, also called dilated convolution. Its main function is to increase the field of sensation without changing the size and computation of the feature diagram [6].. Simply put, it is to add several empty ds between the elements of a standard convolutionary nucleus to support the nucleus, such as in this article, a 3 × 3 convolutionary with 4 empty d transformed into a 9 × 9 convolutionary. The calculation formula for dilated convolution is as follows:

$$n = k + (k-1) \times (d-1) \tag{1}$$

This paper applies a void convolution to the original first convolutional layer, applying a vacuous convolutional core of 9 × 9 to a convolution filled with 4 on a 28 × 28 input image to obtain a 28 × 28 feature diagram.

The resulting loss values and accuracy for the first 15 epochs are shown on the right side of the figure below. You can see an improvement in accuracy on the test set relative to the LeNet-5 network (shown on the left side of the graph below).

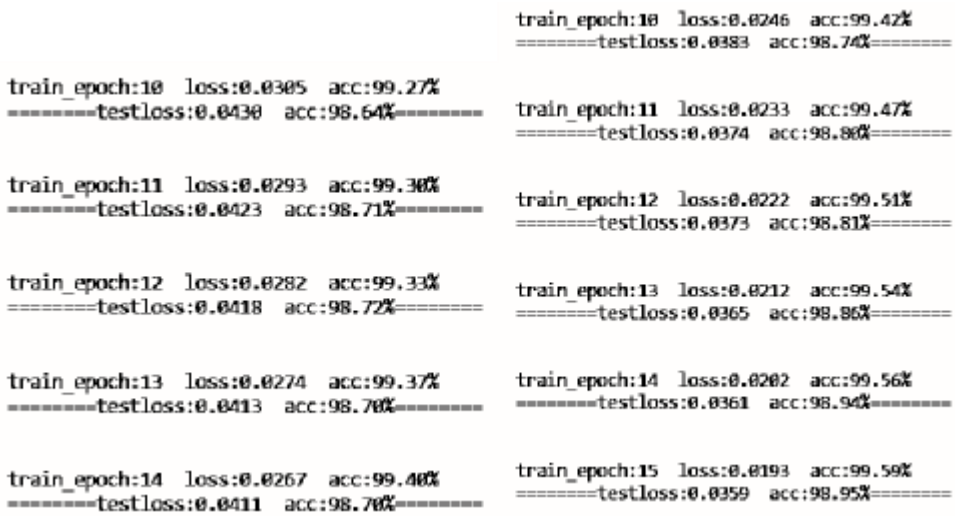

Figure 5: Contrast of results after optimization

4.3 Handwriting board implementation

This experiment set the display window to full black, not only for display purposes, but also to verify the performance of the model. Because both the training and validation sets are white with black characters underneath, they are specially set to full black. The logic of writing is analogous to the process of real writing, giving rise to three major mouse events: the left key presses, the left key releases, and the mouse moves while the left key is pressed. This is the end of the drawing function[7].

The screenshot recognition function requires that after converting the window to type, doing image enhancement and dimension transitions and transferring them to the model for training. The result will be 10 values, but this is not an ideal result. So we also need to use the Softmax function to map the values to probabilities [8]And find the maximum value.

5. EXPERIMENTAL RESULTS

The results are shown in the figure below. Be able to make predictions on standard written numbers and display results. By this time the main function of this experiment has been accomplished.

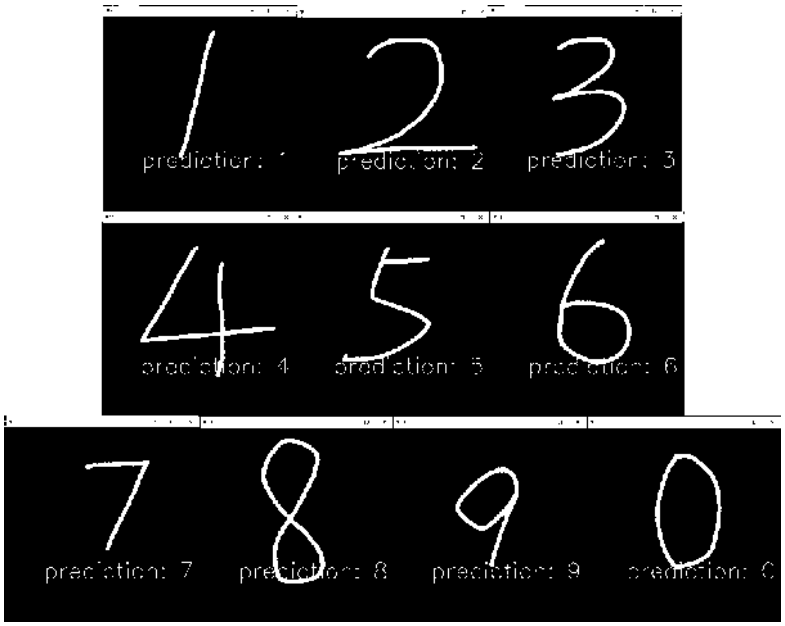

Figure 6: Figure of test results

The following figure is the result of using both the improved and pre-improved models for the same number, respectively. Obviously, for a number with similar writing trajectories, the improved network can identify results more accurately than the LeNet-5 network.

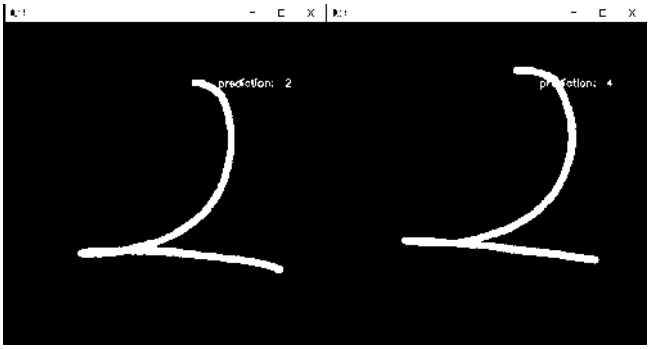

Figure 7: Contrast before and after the improvement (on the left is the improvement after and on the right)

6. CONCLUSION

LeNet-5 networks are popular with newbies for their simple and important features. With the introduction of hardware and some algorithms, some better network structures have emerged since then, such as VGG16, which deepens the network, NIN, which enhances convolutional module capabilities, and R-CNN, which migrates from classification to detection tasks. But these models still use the basic structure of the LeNet-5 network, only deeper and more complex, so learning about the LeNet-5-network will lay a solid foundation for learning about neural networks.

This paper completed and visualized the results of handwritten digital recognition experimentally, rather than relying on the tedious process of taking a photo of handwritten digits and uploading them to the web. This simplifies the traditional prediction process and provides an effective inspiration for future experimental ideas.

Experiments have found that when the numbers are not written correctly, the results are not the same as expected, the main reason is that the data set is not complete. Because everyone has everyone's writing habits, and the characteristics of writers may not be included in the MNIST dataset, Therefore, to improve accuracy, you can add your own written numbers to your training and test sets, but it is important to note that the data added must be large and formatted in a 28 * 28 size, otherwise the results will not be ideal or even errors.

The deficiency of this paper is that the LeNet-5 network and VGG16 network are not compared with the experiment, but the network structure is deepened by the improved model, which makes the experimental results general.

REFERENCES

[1] Li, L., Gan, Y., Bi, S., & Fu, H. (2024). Substantive or strategic? Unveiling the green innovation effects of pilot policy promoting the integration of technology and finance. International Review of Financial Analysis, 103781.
[2] Bi, S., Lian, Y., & Wang, Z. (2024). Research and Design of a Financial Intelligent Risk Control Platform Based on Big Data Analysis and Deep Machine Learning. arXiv preprint arXiv:2409.10331.
[3] Bi, S., Deng, T., & Xiao, J. (2024). The Role of AI in Financial Forecasting: ChatGPT's Potential and Challenges. arXiv preprint arXiv:2411.13562.
[4] Zhou, Z. (2026). Hierarchical Needs in US Automotive Customer Feedback and the Sentiment–Function Nexus. Journal of Industrial Engineering and Applied Science, 4(1), 27-33.
[5] Wensi, L. (2026). AI-Enabled Data Visualization Marketing for Automated Production Lines: Building Customer Trust and Improving Lead-to-Order Conversion. Academic Journal of Natural Science, 3(1), 8-13.
[6] Ximeng, Y., & Yiming, Z. (2026). Offline Conservative RL for Transaction Authorization: Smartly Balancing Fraud Risk and Customer Friction. Journal of Economic Theory and Business Management, 3(1), 1-9.
[7] Zhao, S., Lin, Y., Yang, X., Lu, Q., Xue, H., & Jiang, G. (2025). Optimization of Deep Learning Models for Dynamic Market Behavior Prediction. arXiv preprint arXiv:2511.19090.
[8] Yang, X., Zheng, X., & Lu, Q. (2025, October). Construction and early warning of multi-dimensional network credit-related transaction risk maps by integrating graph neural network (GNN). In Proceedings of the 2025 2nd International Conference on Digital Economy and Computer Science (pp. 919-923).
[9] Wu, J., Luo, L., & Liao, N. (2025). Small-Sample Object Detection of Surface Cracks in Concrete Structures of High-Rise Buildings via Multi-Level Transfer Learning. Innovation & Technology Advances, 3(2), 57–72. https://doi.org/10.61187/ita.v3i2.262
[10] Ren, Z. (2024). Adaptive Multi-Scale Fusion for Infrared and Visible Object Detection in YOLOv8. Journal of Theory and Practice of Engineering Science, 4(09), 28–34. https://doi.org/10.53469/jtpes.2024.04(09).04
[11] Shen, Zepeng, et al. "Research on Application of Whale Optimization Algorithm in Financial Payment Fraud Detection." 2025 4th International Conference on Artificial Intelligence, Internet and Digital Economy (ICAID). IEEE, 2025.
[12] Tian, Q., Wang, Z., & Cui, X. (2024). Improved Unet brain tumor image segmentation based on GSConv module and ECA attention mechanism. arXiv preprint arXiv:2409.13626.
[13] Yi, X. (2025, October). Real-Time Fair-Exposure Ad Allocation for SMBs and Underserved Creators via Contextual Bandits-with-Knapsacks. In Proceedings of the 2025 2nd International Conference on Digital Economy and Computer Science (pp. 1602-1607).

[14] Tang, Y., Kojima, K., Gotoda, M., Nishikawa, S., Hayashi, S., Koike-Akino, T., ... & Klamkin, J. (2020). Design and Optimization of Shallow-Angle Grating Coupler for Vertical Emission from Indium Phosphide Devices.