International Journal of Research in Engineering and Science (IJRES)
ISSN (Online): 2320-9364, ISSN (Print): 2320-9356
www.ijres.org Volume 11 Issue 5 ǁ May 2023 ǁ PP. 482-487

Handwritten Digit Recognition using Machine Learning : A Review

Aman Chandra[1], Vineet Sharan[2], Satish Patil[3], Varun Kumar[4], Students,
Department of Information Science and Engineering,
Dayananda Sagar Academy of Technology and Management, Bengaluru, India.
And Ms. Spandana SG, Assistant Professor, DSATM, Bengaluru, India.

Abstract—*Handwritten digit recognition is a common problem in the field of machine learning and image processing that involves training a model to classify handwritten digits based on images of those digits. One dataset frequently used for this task is the MNIST (Modified National Institute of Standards and Technology) dataset, which contains 70,000 images of handwritten digits from 0 to 9. These images are divided into a training set of 60,000 images and a test set of 10,000 images. When using the MNIST dataset to train a model for handwritten digit recognition, the model is presented with an image of a handwritten digit and must output the correct digit label. This process is repeated for each image in the training set, and the model's performance is evaluated using the test set. The goal is to train the model to accurately classify new, unseen images of handwritten digits. Using machine learning techniques such as neural networks and the MNIST dataset can be very effective for handwritten digit recognition, with the right training and appropriate model selection, these approaches can achieve high levels of accuracy in classifying handwritten digits.*

Keywords— *MNIST (Modified National Institute of Standards and Technology), CNN (Convolutional Neural Network), SVM (Support Vector Machine), ANN (Artificial Neural Network)*

Date of Submission: 07-05-2023 Date of acceptance: 18-05-2023

I. INTRODUCTION

The recent rapid development of electronic information, the inputs to computers have been becoming more common but handwritten information passed between people have been an irreplaceable way of transfer information. Like a link combining handwritten characters and computer keying, handwritten recognition has received increasing attention for its practicability. In the past decade, machine learning and pattern recognition have extended many highly intelligent handwriting recognition classifications, including artificial neural networks (ANN), support vector machine (SVM), modified quadratic discriminant function (MQDF) and hidden Markov model , etc.

Convolutional neural network have the most advantage in image processing than all the other neural networks, therefore this paper designs a handwritten digit recognition system based on convolutional neural networks. Handwritten digit recognition can be classified into two categories: online and offline recognition. Online recognition involves the computer recognizing handwritten characters as they are being written on a handwriting device, based on the order and sequence of the strokes. This technology has been well-studied and is relatively mature. Offline recognition, on the other hand, involves the computer recognizing characters that have already been written on paper. This process relies on image recognition, which has less information to work with and therefore presents greater challenges. The system described in this paper is designed for offline recognition of handwritten digits.

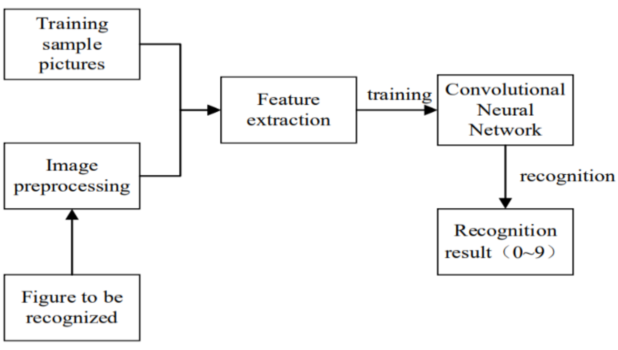

Fig.1 The flowchart of handwritten digit recognition system

This paper presents an identification system that consists of two main components: a data source module and a digital identification module. The data source module is responsible for obtaining handwritten digit images and extracting features from them. The digital identification module, on the other hand, uses a convolutional neural network (CNN) to perform the actual recognition of the handwritten digits.

II. DATA SOURCE MODULE

A. MNIST Dataset

The MNIST dataset is provided by the National Institute of Standards and Technology (NIST) and consists of handwritten digits from 250 different individuals, with half being high school students and the other half being Census Bureau staff. It is divided into a training set and a test set, both of which contain a equal mix of digits written by the different groups. The training set is made up of 70000 samples taken from the MINST library's SD1 and SD3 sections, while the test set consists of 10000 samples, with 5000 coming from each of the SD1 and SD3 sections.

The MINST dataset includes standardized images of handwritten digits that are all 28 x 28 pixels in size. This identification system uses the MNIST dataset to train its recognition model by extracting features from the images and performing the necessary training. Once the model has been trained, it can be used to identify handwritten digits by inputting them for recognition.

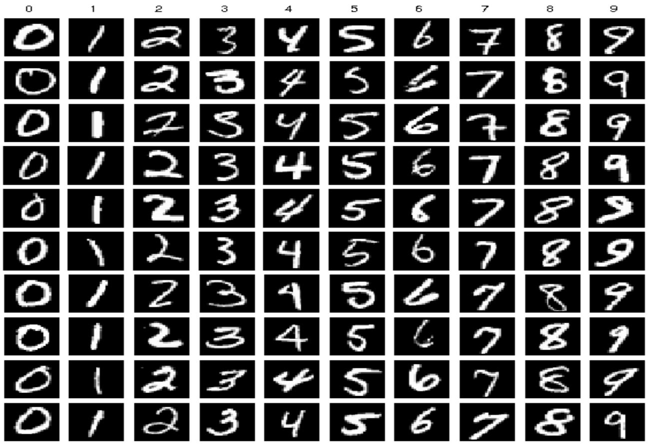

Fig.2 MNIST dataset

B. Opencv Toolkit

Opencv (Open Source Computer Vision Library) is a free and open-source library that includes a wide range of computer vision algorithms. In this paper, the system employs the Opencv toolkit to perform pre processing and feature extraction on the digits that need to be recognized. Specifically, Opencv is used to resize the handwritten images to the same 28 x 28 pixel size as the MNIST dataset, as well as to perform various image processing tasks such as linear and non-linear filtering, image transformations (e.g. resizing, rotation, perspective), colour space conversions, and histogram analysis.

This system utilizes the OpenCV toolkit to process handwritten digit images by first converting them to grayscale, performing a grayscale histogram analysis, and then applying thresholding and binarization techniques to extract the outline features of the images.

III. DIGITAL IDENTIFICATION MODULE

A. Deep Learning

In 2020, Jinze Li introduced the concept of deep learning for the first time in the paper. Deep learning is a type of machine learning that is based on artificial neural networks (ANN) and has the ability to adapt well to local minima during the training process. ANNs are composed of multiple perceptrons, which are also known as multi-layer perceptron (MLP) neural networks. The perceptron model is shown in Figure 3.

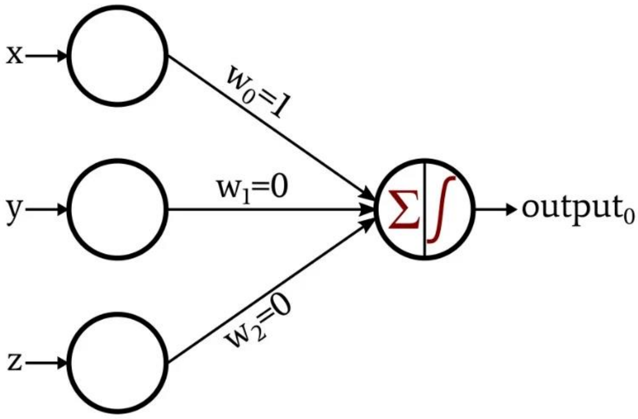

Fig.3 Perceptron Model

The output of the above perceptron is the formula (1):

$$\begin{aligned} y &= 1 \quad if \sum_{i=1}^{n} w_i * x_i \geq \theta \\ &= 0 \quad if \sum_{i=1}^{n} w_i * x_i < \theta \end{aligned} \qquad \boxed{1}$$

The perceptron is a type of artificial neural network that was developed by Frank Posenblatt in the 1960s. It takes in multiple binary inputs (x1, x2, ... xn) and produces a binary output. The output of the perceptron is determined by whether the sum of the weighted inputs is above or below the threshold value. This is shown in formula (1). The structure diagram of the ANN is shown in Figure 4:

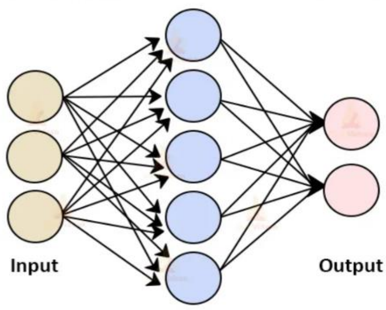

Fig. 4 Structure of Artificial Neural Network (ANN)

Artificial neural networks (ANNs) are inspired by the structure and function of biological neural networks, but are simplified and optimized for practical use through a learning process based on mathematical statistics. ANNs are capable of making decisions and judgments in a manner similar to human thinking and have many characteristics of biological systems, such as robustness, high parallelism, nonlinearity, fault tolerance, and good learning ability. In many cases, ANNs are more reliable than traditional logical reasoning methods.

Deep learning networks are made up of multiple artificial neural networks (ANNs). These networks typically have three layers: an input layer, a hidden layer, and an output layer. The structure of a deep learning network is illustrated in Figure 5.

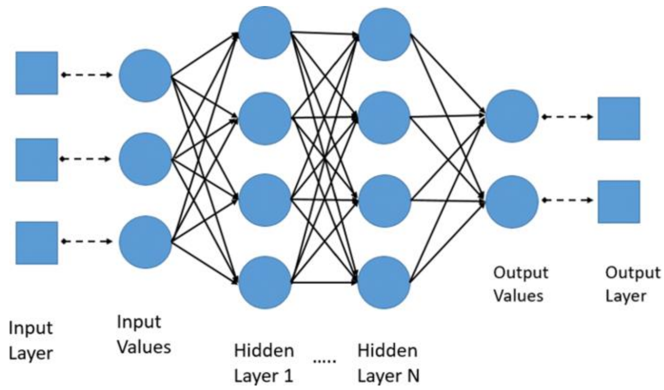

Fig.5 Structure diagram of deep learning network

B) Principle of Convolutional Neural Network

1) Convolutional neural network structure

The structure of the convolutional neural network is shown in Figure 6:

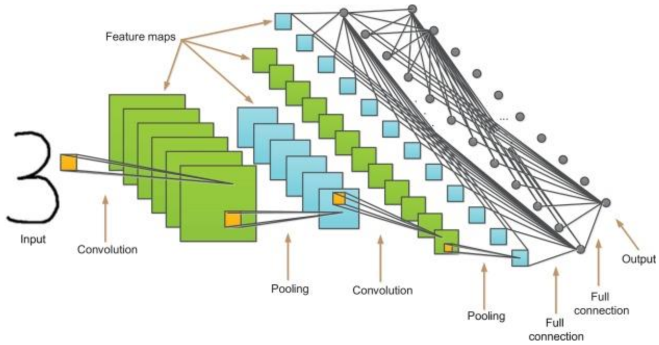

Fig.6. Structure of the convolutional neural network

A convolutional neural network is made up of two main types of layers: convolutional layers (C) and pooling layers (S). The image is input into the network through the input layer and then processed by the convolutional layer C, which uses three trainable digital filters to create three feature maps. These feature maps represent the features that the layer has learned. The output of C is then passed through the pooling layer P, which reduces the spatial dimensions of the data. This process is repeated with the next convolutional layer and the next pooling layer, after this the features are flattened and fed into the rest of the neural network.

2) Convolutional layer

The flow chart of the convolution layer operation is shown in Figure 7:

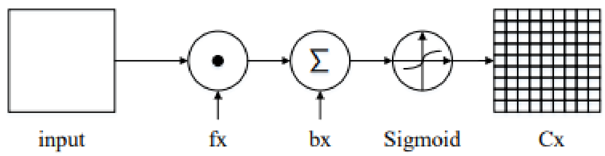

Fig.7. The flow chart of the convolution layer

In Figure 7, the digital filter fx and the bias voltage bx are shown. In order to apply the convolution formula to two-dimensional images, which are represented as numerical arrays, we must use a two-dimensional version of the formula.

$$s(i,j)=(X*W)(i,j)=\sum_m \sum_n x(i+m,j+n)\omega(m,n) \tag{2}$$

Equation (2) shows that the values in the final matrix are obtained by multiplying elements of the input image with elements of the convolution kernel matrix and then adding the results. The examples is shown in figure 8.

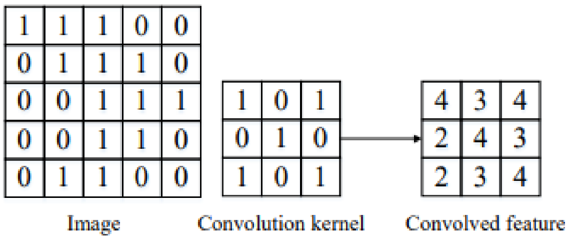

Fig. 8 Image Convolutional example

3) Pooling Layer

Pooling operations are simpler than convolution operations. The purpose of pooling is to compress each sub-matrix in the input tensor by reducing it to a single value. This is typically done by taking the maximum or average value of the sub-matrix. In this paper, the average value method is used for higher accuracy in image processing.

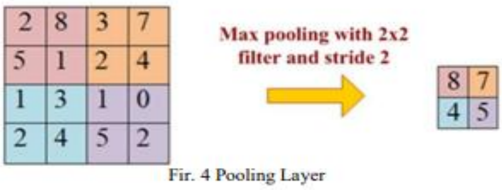

Fig. 9 Pooling Layer

The purpose of the pooling operation in a convolutional neural network is to reduce the dimensionality of the data by mapping the average of a 2x2 area of pixels in the convolutional layer C to a single point in the pooling layer S. This helps to reduce the size of the matrix.

C. Convolutional Neural Network Based on LeNet-5

The convolutional neural network used in this system is based on the LeNet-5 model [8-9], which was developed by Yann LeCun in 1998 for recognition of handwritten digits. At the time, it was widely used by US banks to process check transactions. LeNet-5 is a pioneering example of a convolutional neural network and remains one of the most well-known models in the field.

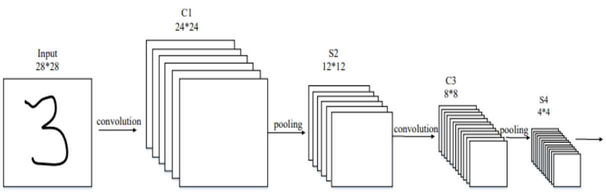

Fig. 11 CNN model based on LeNet-5

Figure 11 shows that LeNet-5 is composed of two convolutional layers, two pooling layers, and one fully connected layer. The operations of C1, S2, C3, and S4 are similar to those in Figure 6. The number of feature maps produced by the pooling layers is fixed, while the number of feature maps in the convolutional layers is not dependent on the previous layer. For example, the C1 layer uses six trainable digital filters to create six feature maps, and the C3 layer combines all 12 feature maps from the S2 layer to create its own set of 12 feature maps. The fully connected layer takes the output of S4 and flattens it into a 1x10 one-dimensional vector.

D. Classification of Recognition Results

This system is designed to identify ten discrete values(0-9) and is therefore dealing with a multi-classification problem where the sum of the probabilities of all the possible outcomes is 1. In this case, the system uses a Softmax regression model, which is a generalization of the logistic regression model for multi-classification problems. The Softmax regression model adds a layer of function mapping on top of the continuous values, which have the property of linear superposition. This mapping is then used to transform the continuous values into discrete values between 0 and 1.

Softmax regression model can better solve the multi-classification problem. In the Softmax regression model, the class label y can take different K values. The probability of category i in Softmax regression is:

$$s_i = \frac{e^{v_i}}{\sum_j e^{v_j}} \qquad \boxed{3}$$

The Softmax function in this system maps the probabilities of the output values (0-9) to the interval between 0 and 1, with the sum of these values being 1. When selecting the output node in the output layer, the node with the highest probability is chosen as the prediction target.

IV. CONCLUSION

This paper presents a handwritten digit recognition system based on a convolutional neural network that uses deep learning and the MNIST data set for training. The system uses the LeNet-5 model to perform repeated convolution and pooling operations to recognize handwritten digits in images through the Softmax regression model. Handwriting recognition is important because it allows machines to understand and process handwritten input, which can be useful in fields like finance, accounting, and education. However, current handwriting recognition technologies often require manual review and have room for improvement in terms of accuracy and speed. Improving the recognition accuracy and reducing the recognition delay are ongoing challenges in the field.

REFERENCES

[1]. Burel, G., Pottier, I., & Catros, J. Y. (1992, June). Recognition of handwritten digits by image processing and neural network. In Neural Networks, 1992. IJCNN, International Joint Conference on (Vol. 3, pp. 666-671) IEEE.

[2]. Salvador España-Boquera, Maria J. C. B., Jorge G. M. and Francisco Z. M., “Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models”, IEEE Transactions on Pattern Analysis and Machine Intelligence, Vol. 33, No. 4, April 2014.

[3]. Long, M.; Yan, Z. Detecting iris liveness with batch normalized convolutional neural network. Compute, Mater. Contin. 2019, 58, 493–504.

[4]. G. E. Hiton, S. Osindero, Y. Teh, “A fast learning algorithm for deep belief nets”, Neural computation, vol. 18, no. 7, pp. 1527-1554, 2006.

[5]. G. H. Rosa, J. P. Papa (2019). Soft-Tempering Deep Belief Networks Parameters Through Genetic Programming. Journal of Artificial Intelligence and Systems, 1, 43–59.