RECOGNITON OF NUMERICS USING OPENCV AND DEEP LEARNING WITH HIGH ACCURACY RATE

Dr. KARTHIK V[1], SOUNDAPPAN. A[2], SIBI CHAKRAVARTHY. S[3], SHIVA DHARSAAN. S[4]

[1]Associate Professor
Department of Electronics and Communication Engineering
Sri Krishna College of Engineering & Technology, Coimbatore
karthikv1711@gmail.com
[2]U.G Scholar
Department of Electronics and Communication Engineering
Sri Krishna College of Engineering and Technology, Coimbatore
19euec149@skcet.ac.in

[3]U.G Scholar
Department of Electronics and Communication Engineering
Sri Krishna College of Engineering and Technology, Coimbatore
19euec142@skcet.ac.in

[4]U.G Scholar
Department of Electronics and Communication Engineering
Sri Krishna College of Engineering and Technology, Coimbatore
19euec136@skcet.ac.in

Abstract

With the goal of making machines more intelligent, developers are immersed in machine learning and deep learning techniques. Handwritten digit recognition is the ability of computers to recognize human handwritten digits. A solution to this problem is handwritten digit recognition, which uses an image of a digit and recognizes the digit present in the image. It introduces an offline recognition method for handwritten digits based on convolutional neural network. This system uses the MINST dataset as a training sample and image preprocessing using the OpenCv toolkit. It uses LeanNet-5 on a convolutional neural network to extract handwritten digital image features by iterative convolutional pooling, and pulls the result into a one-dimensional vector.

Finally, the Softmax regression model is used to find the highest likelihood point to determine the outcome of the handwritten digit. We have used Python as its coding language.

Key Words : ConvolutionalNeural Network, MINST, OpenCv, LeNet- 5, Softmax Regression Model, Python, etc.

1. INTRODUCTION:

Good progress has been made in looking at the model in recent years. The results of the strike were mainly seen in the area of signature honors. This rapid growth is a combination of advances that include the proliferation of powerful, inexpensive computers, the development of new algorithms that use these computers, and the discovery of large sets of characters that can be used for training and testing. We have created the Algorithm class. In this paper we compare the optimal beauty of each algorithm. In addition to accuracy, we look at metrics like training time, processing time, and recall.

METHODOLOGY:

1.1 Convolutional Neural Network(CNN):

A convolutional neural network (CNN) is actually a multilayer perception network (MLP) first used in the 1980s. CNN's computer is inspired by the human brain. People see or point to things based on appearances. People teach their children to see things by showing them hundreds of photos. It helps the child to discover or predict things that he has never seen before. CNN works in a similar way and is popular for visual image analysis. Popular CNN formats include GoogLeNet (22 layers), AlexNet (8 layers), VGG (16-19 Wings), and ResNet (152 layersCNN integrates classification and feature extraction steps and requires minimal input processing and implementation effort. CNN can extract advanced and automatic features from images. Furthermore, CNN can provide high recognition accuracy even if only a small amount of training data is available.

Construction details and previous properties and features no longer need to be stored. The use of topological information obtained from the inputs is a major advantage of using the CNN model in providing excellent recognition results. The CNN pattern recognition results are independent of the permutation and translation of the input images. In contrast, in-depth information about the inputs is not used in MLP models. Also, with a complicating problem, MLP is not found to be optimal and is insufficient to obtain high-resolution images due to the complete correlation between locations, also known as the famous “curse of greatness”. In recent years, the CNN model has been widely used to extract digital signatures from the MNIST database. Some researchers have reported accuracy as high as 98% or 99% with manual digital authentication. An integration model is designed using a combination of several CNN models. The MNIST digit is authenticated with a reported accuracy of 99.73%. Subsequently, the "committee of 7 numbers" was expanded into the "committee of 35 numbers" and improved visual accuracy was reported as 99.77% of the same MNIST database [10]. Niu and Suen reported an amazing 99.81% accuracy in combining support vector (SVM) capabilities to reduce structural risk and CNN model capabilities to extract deeper features of the MNIST digital identity experiment. Fold-in directional feature maps have been studied using CNNs for Chinese handwriting recognition. [12] Recently, the work of Alvear-Sandoval et al. We achieved a 0.19% error rate for MNIST by creating a set of different deep neural networks (DNNs). However, upon closer inspection, we found that high image accuracy in the MNIST dataset is only detected by the construction method. The combined method helps improve section accuracy, but it also helps drive up the cost of complex testing and real-world application computing costs.

The aim of the proposed work is to achieve comparable accuracy using a pure CNN structure and to thoroughly study the learning parameters in constructing the CNN architecture to obtain MNIST digital recognition. Another aim is to investigate the role of various parameters and to properly tune the important parameters in improving the structure of CNN architecture.

The main contributions of this work are therefore on two fronts. First, a complete overhaul of various parameters such as number of layers, stride size, kernel size, padding, and dilution was performed for CNN design in manual handwritten digitization to improve performance.

Second, the efficient use of learning parameters resulted in good visual performance on the MNIST database. The MNIST website is used in this project due to the availability of results published by various distributors. The website is popular and widely used as a benchmark for benchmarking studies in various digital sign language experiments in various regional and international languages. The proposed work innovation involves a thorough investigation of all design parameters of CNNs to achieve the highest accuracy among MNIST's digital recognition peers. The recognition accuracy submitted to this work using a pure CNN model is much higher than recognition data reported by peer-reviewed researchers. Collaborative technology use by peer researchers includes additional computers cost and difficulty of advanced testing. Therefore, the proposed pure CNN model outperforms the cohesion of techniques offered by both reviewers with accurate recognition and computational complexity

1.2 Open Cv Library:

OpenCV is also known as Intel's proprietary open-source computer vision library for computer display, machine learning, and image editing. It plays a key role in real-time performance and is critical in today's systems. It can process photos and videos to identify objects, faces, and handwriting.

1.3 The NIST Test:

After the competition, many competitors achieved <1% error on the validation set taken from the training data, but found that their performance on the test data was very poor.

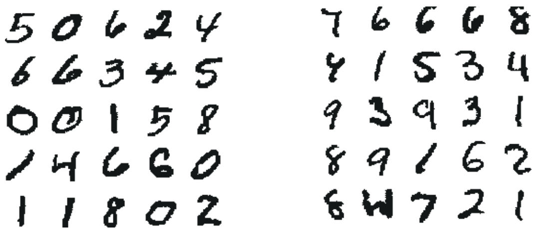


Fig. 1. a) Images from NIST training set, and NIST test set.

NIST noted that the training and test sets are representative of different geographies. The training set consisted of letters written by US enrollment staff, and the test set was collected from letters written by uncooperative high school students.

Examples of these training and testing sets are shown in Figure 1. Note that the test images contains very complex patterns. Variations in this distribution are quite possible in a real app, but it's wise (and common) to monitor .The best test results can usually be expected if the supervisor is limited to the types of data that may be encountered in transmission.

1.1 The MNIST:

For the reasons described above, we have shared NIST data to provide extensive training and test groups that share the same distribution. We now explain how our new database was created. The original NIST preview contains 58,527 digital images written by 500 previous authors. In contrast to a training set, where blocks of data from each author appear sequentially, the data in the NIST test set is confusing.

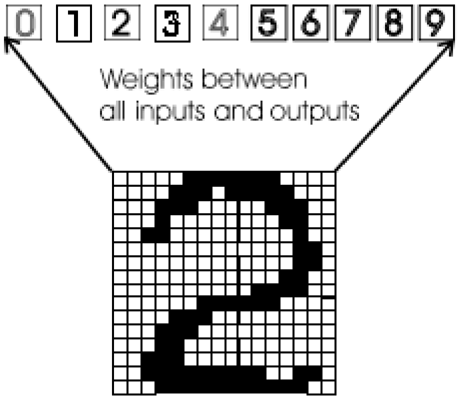


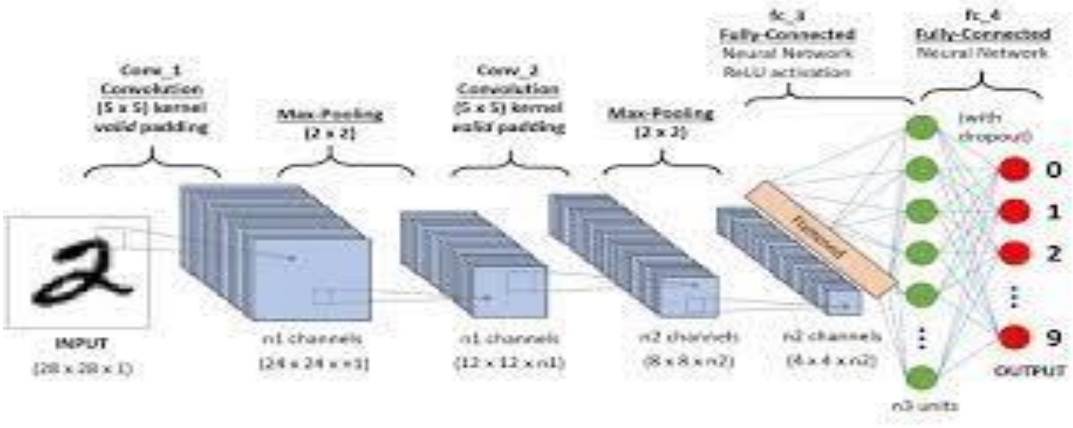
Fig. 2. Linear Classier. Each input unit pixel value contributes to a weighted sum for each output unit. The output unit with the largest sum indicates the class of the input digit.

The author ID is available for the test set and we used this information to process the authors. Then we divided this NIST test into two parts: letters by 250 authors entering our new training cohort. The remaining 250 authors were placed in our test group. So we had two sets with roughly 30,000 examples each. The new training set is complete with enough examples from the old NIST training set, starting with Model 0, to create a complete set of 60,000 training patterns.

Likewise, the new test group was complete with legacy training patterns starting with pattern #35,000 to make a complete set of 60,000 test patterns. In the experiments described here, we only used the first 10,000 images, but we used 60,000 training samples.

2. ARCHITECTURE:

The basic convolutional neural network consists of three components, namely the convolutional layer, the integration layer, and the extraction layer. The integration layer is sometimes optional. The typical architecture of a three-layer convolutional network is well chosen for separating handwritten images as shown in Fig. 1. It consists of an input layer, multiple hidden layers (convolutional redundancy, normalization, aggregation), and a complete connection and extraction layer. Neurons in one layer connect to other neurons in the next, making measurement easier in high-resolution images. Low sampling or blending performance can be used to reduce the input size. In the CNN model, the input image is a set of small areas called "reception fields".



The mathematical function of convolution is implemented in the input layer, which simulates the answer in the next layer. The response is essentially a visual stimulus. The detailed description is as follows:

2.1 Input layer:

The input data is loaded and stored in the input layer. This layer defines the length, width, and number of channels (RGB details) for the input image.

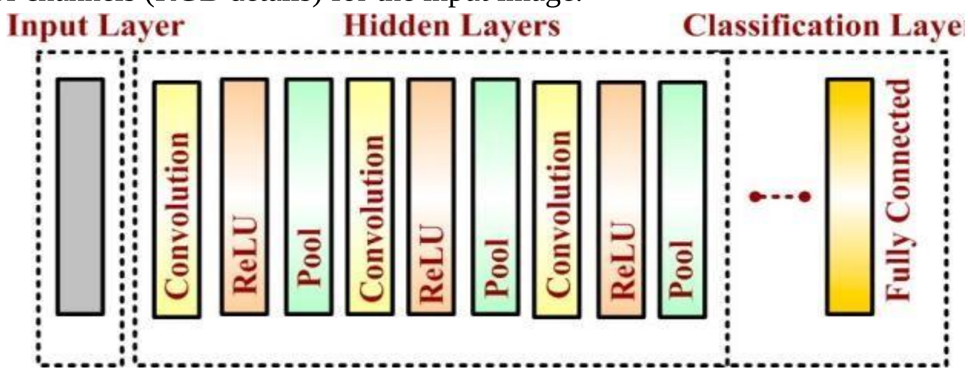


Figure 1. Typical convolutional neural network architecture.

2.2 Hidden Layer:

Hidden layers are the backbone of the CNN architecture. They perform a feature extraction process in which a series of mixing, mixing, and activation is used. Distinctive features of handwritten numbers are found in this section.

2.3 Convolutional Layer:

The convolutional layer is the first layer placed above the insertion image. Used to extract image elements. The n × n input neurons of the input layer are integrated with an m × m filter and by retrieving (n - m + 1) × (n - m + 1) as output. The present is nonlinear using the neural activation function. The main contributors to the convolutional layer are the field of acceptance, step by step, minification and folding, as described in the next section. CNN computations inspired by the visual cortex in animals.

The cortex is the part of the brain that processes information transmitted from the retina. It processes visual information and is hidden in small input areas. Similarly, the receiving field is computed on a CNN, which is a small area for capturing images that can affect a specific network area. It is also one of the most important limitations of the CNN architecture and is useful for setting other CNN parameters. It is about the same size as the nucleus and works in the same way as the vision of the human eye to produce sharp central vision. The receiving field is affected by fingerprint, integrity, kernel size, and CNN depth.

The terms receive field (r), active receive field (ERF), and projective field (PF) are used to calculate subnetwork areas. The location of the first image affecting neuron function is determined using the ERF, while the PF is the number of neurons in which the neuron exerts its influence, as shown in Fig. 2. The 5 × 5 filter size and the activation map are described in Fig. 3. Stride is another parameter used in building a CNN architecture.

Defined as a step-by-step filter. Stamped value 1 indicates the movement of the pixel by moving the filter in pixels. A large rowing size indicates a small overlap between cells. The performance of the kernel and step in the convolution layer is shown in Figure 4.

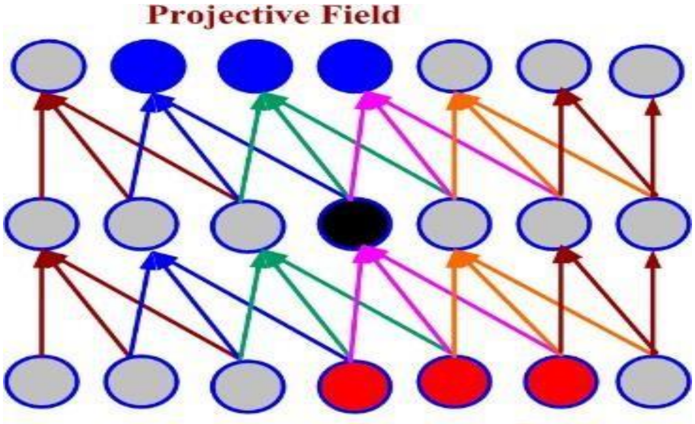


Figure 2. Receptive field and projective field

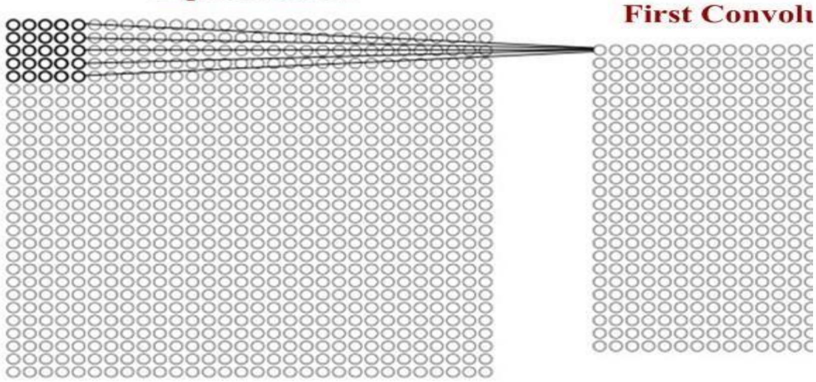


Figure 3. Visualization of filter of size 5× 5 with activation map. (Input neuron 28 × 28 and convolutional layer 24 ×24).

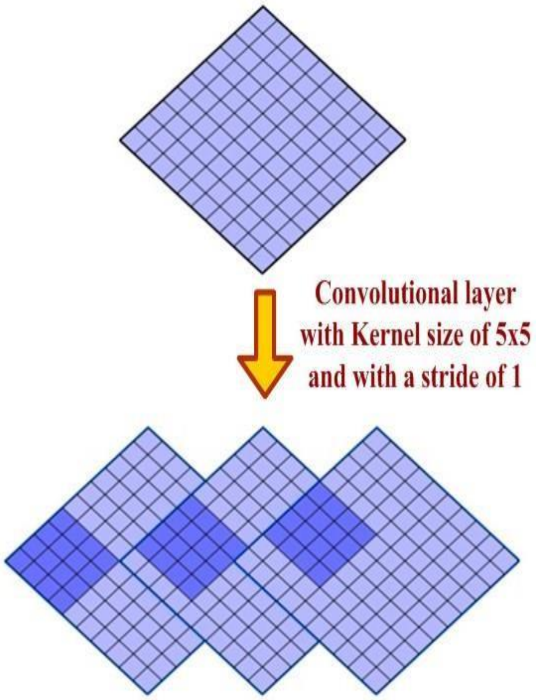


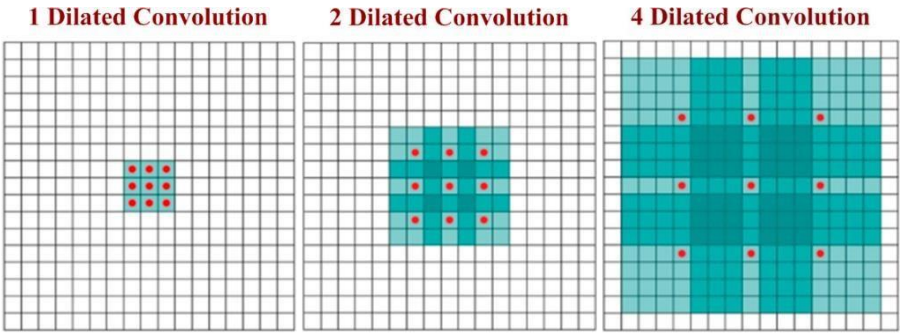
Figure 4. Representation of kernel and stride in a convolutional layer

The concept of padding was introduced into the CNN build to gain more accuracy. Padding was introduced to control the reduction of convolutional layer emissions.
The output in the convolutional layer is a feature map, smaller than the inserted image. The output feature map contains a lot of detail about the center pixels, so you lose most of the information available in the corners. Zero rows and columns are added to the borders of the image to prevent the feature map from shrinking.
Figures (1) and (2) describe the relationship between the size of the feature map, the kernel size and stride while calculating the size of the output element map.

$$Wnx = Wn - 1x - FnxSnx + 1 \quad (1)$$
$$Wny = Wn - 1y - FnySny + 1 \quad (2)$$

where (Wnx, Wny) represents the output size map size, (Snx, Sny) is the stride size, and (Fx, Fy) is the kernel size. Here 'n' is used to describe the layer index.

Reduction is another important parameter for the formation of a CNN that has a direct impact on the receiving sector. A reduction can increase the CNN's display area (FOV) without changing the feature map. Figure 5 clearly shows that the expansion rates can significantly increase the CNN receiving segment. Excessive cuts can overload computers and thus can degrade the system by increasing processing time. So it must be chosen wisely. The relationship between filter, weight and input is shown in equations (3) and(4)below



(b) (c)

Figure 5. Dilated convolution: (**a**) receptive field of 3 × 3 using 1-dilated convolution; (**b**) receptive field of 7 × 7 using 2-dilated convolution; (**c**) receptive field of 15 × 15 using 4-dilated convolution.

2.4 Hidden Layer:

The integration layer is added between the two resolution layers to reduce the size of the installation and thus reduce the complexity of the computer. Installing pools allows selected values to be transferred to the next layer while leaving unnecessary values behind. The composite layer also helps in character selection and over-control. The integration work is done independently. It works by extracting only one output value in the regions under the tiles for input images. The most common types of max-pooling and avg-pooling operations (where max and avg represent maxima and average, respectively). Max-pooling work is often done in modern systems, because it takes high prices from each region, storing high data. This leads to faster integration and better performance . The max-pooling function of converting 4 × 4 output to 2 × 2 output by stride 2 output is shown in Figure 6. The maximum number is taken for each combined output (size 2 × 2) resulting in a total reduction of 2 × 22.

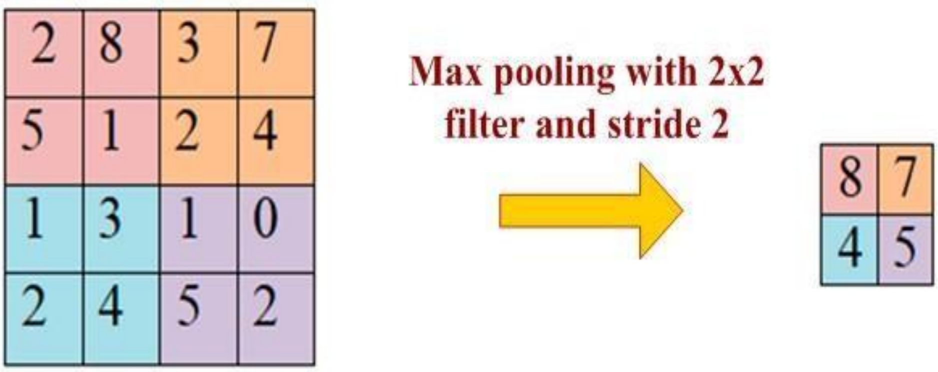


Figure 6. Max pooling with filter 2 × 2 and stride size.

2.5 Activation Layer:

Like the construction of a typical neural network, the structure of CNN also contains the function of making a presentation that is not compatible with another system. Sigmoid function, fixed modified unit (ReLu) and Softmax are some of the popular options among the various application functions that are most exploited in in-depth learning models. It has been shown that sigmoid performance may slow down the CNN model due to the loss of information present in the input data. The activation function used in this current function is a noncompliant line line (ReLu) unit, with the result of 0 input less than 0 and a green output otherwise.

Other advantages of ReLu's performance function are its similarity to the human emotional system, and the ability to perform rapid training of large networks.

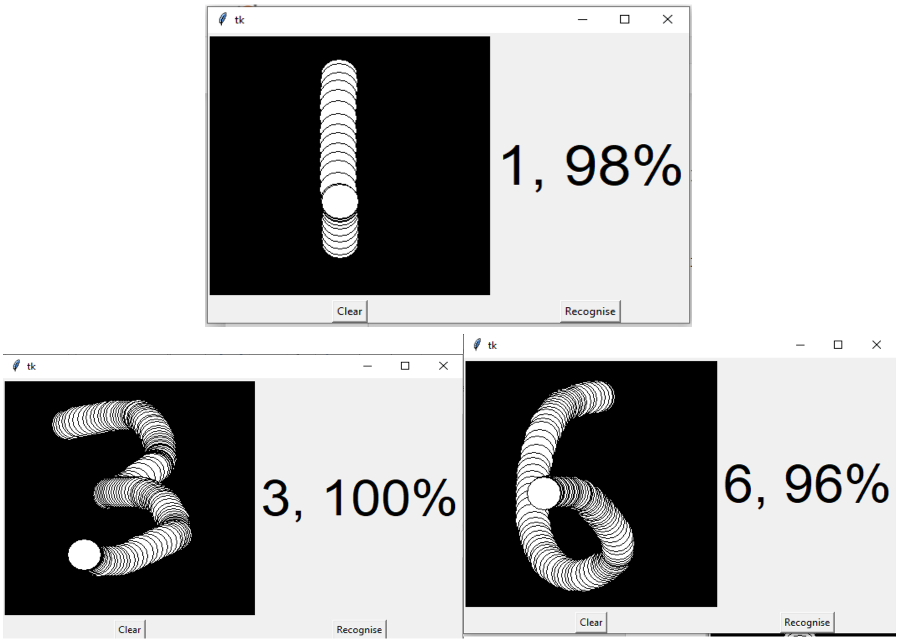
2.6 Classification Layer:

Layer to separate the last layer in the CNN architecture. It is a fully integrated feed supply network, which is accepted as a separator. Neurons in the fully connected layers are connected to all the neurons of the previous layer. This layer lists the predicted classes by pointing to the input image, which is done by combining all the features learned in the previous layers. The number of output classes depends on the number of classes available in the targeted database. In the present work, the partition layer uses the 'softmax' function to separate the generated elements of the input image obtained from the previous layer into different categories depending on the training details.

3. EXPERIMENTAL SETUP:

The handwriting digit recognition process consists of the following steps:

1. Finding or collecting handwritten digital photographs of MNIST.
2. Divide input images into training and image testing.
3. Use the pre-processing process on both the training database and the test database.
4. Make data standard from 0 to 1.
5. Divide the training database into groups of appropriate size.
6. Training the CNN model and its diversity using labeled information.
7. Using a separate segmentation model.
8. Analyse the accuracy of detection and processing time of all types.



These images show the results from the experimental setup

5. APPLICATIONS:

Handwritten digit recognition is the ability of a computer to recognize the human handwritten digits from different sources like images, papers, touch screens, etc, and classify them into 10 predefined classes
(0-9). This has been a topic of boundless- research in the field of deep learning. The use of digit recognition includes

- Postal planning
- Processing of bank checks
- Data form entry
- License plate recognition
- Mark sheet Evaluation etc.

6. CONCLUSION:

In this work, with the aim of improving the performance of digital recognition manually, we explored a variety of convolutional neural networks to avoid complex pre-processing, costly removal of features and complex way of integrating traditional system. With extensive testing using the MNIST database, the current function suggests the role of various hyper parameters. We also confirmed that a good adjustment of hyper parameters is important in improving the performance of CNN architectures. We obtained a 99.89% discovery rate with the Adam optimizer of the MNIST database, the best of all previously reported results. The effect of increasing the number of convolutional layers in the CNN structure in the performance of handwritten digital**.**

7. REFERENCES

[1] Yang Shuying, Image recognition and project practice. Beijing: Publishing House of Electronics Industry, 2014.

[2] Basu S., Das N., Sarkar R., Kundu M., Nasipuri M., Basu D.K. (2009) Recognition of Numeric Postal Codes from Multi-script Postal Address Blocks. In: Chaudhury S., Mitra S., Murthy C.A., Sastry P.S., Pal S.K. (eds) Pattern Recognition and Machine Intelligence. PReMI 2009. Lecture Notes in Computer Science, vol 5909. Springer, Berlin, Heidelberg.

[3] Impedovo S, Pirlo G, Modugno R, Ferante A, "Zoning Methods for Hand- Written Character Recognition - An Overview," International Conference on Frontiers in Handwriting Recognition,

[4] Huailin Zhao, Yawei Hou, Shifang Xu, Congdao Han, Masanori Sugisaka, "Research on a Method of Character Recognition for Self- learning Errors," The 2017 International Conference on Artifical Life and Robotics, Japan, 2017, pp.190-193.

[5] Deng L, Hinton G, Kingsbury B, "New types of deep neural network learning for speech recognition and related applications : an overview," Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing. Vancouver BC: IEEE, 2013, pp. 8599- 8603.

[6] Zhang C, Zhang Z Y, “Improving multiview face detection with multi- task deep convolutional neural networks,” Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Steamboat Springs, CO:IEEE, 2014. pp. 1036-1041.

[7] Wu Meixian, Zhang Xueliang, Wen Shuhua, Guo Qin, “BP neural network dual learning rate adaptive learning algorithm,” Modern Manufacturing Engineering, no. 10, pp. 29-32, 2005.

[8] Yi Chaoren, Deng Yanni, “Multichannel convolution neural network image recognition method,” Journal of Henan University of Science & Technology(Natural Science), no. 3, pp. 41-44, 2017.

[9] M. Haghighat, S. Zonouz, M. Abdel-Mottaleb, "CloudID: Trustworthy cloud-based and cross-enterprise biometric identification," Expert Systems with Applications, vol. 42, no. 21, pp. 7905- 7916, 2015.

[10] Wan L, Zeiler M, Zhang S, LeCun Y, Fergus R, “Regularization of neural networks using dropconnect,” Proc of the 30th International Conference on Machine Learning, 2013, pp. 1058-1066.

[11] Jarrett K, Kavukcuoglu K, Ranzato M, LeCun Y, “What is the best multi-stage architecture for object recognition?” Proc of IEEE 12th International Conference on Computer Vision, 2009, pp. 2146-2153.

[12] Chen Haoxiang, Cai Jianming, Liu Kengran, Lin Qiushuang, Zhang Wenling, Zhou Tao, “Handwritten digital depth feature learning and recognition,” Computer Technology and Development, vol. 26, no. 7, pp. 19-23, July 2016.