

# Handwritten Digit and Text Recognition Based On Convolutional Neural Network Approach

<sup>1</sup>Aruna A, <sup>2</sup>Vivekanandan S J, <sup>3</sup>Dr Sivasubramanian S

<sup>1</sup>P. G. Scholar, Department of Computer Science and Engineering, Dhanalakshmi College of Engineering , Chennai, Tamil Nadu, India

<sup>2</sup>Assistant Professor, Department of Computer Science and Engineering, Dhanalakshmi College of Engineering , Chennai, Tamil Nadu, India

<sup>3</sup>Associate Professor, Department of Computer Science and Engineering, Dhanalakshmi College of Engineering , Chennai, Tamil Nadu, India

<sup>1</sup>arukanishka94@gmail.com, <sup>2</sup>vivekanandan.sj@dce.edu.in, <sup>3</sup>sivasubramanian.s@dce.edu.in

**Abstract:** The goal of this undertaking is to foster a powerful penmanship acknowledgment methods utilizing ideas of Machine learning and PC vision. An expansion of MNIST digits dataset called the Emnist dataset has been utilized. It contains 62 classes with 0-9 digits and A-Z characters in both capitalized and lowercase. To recognize transcribed content and convert it into computerized structure utilizing Convolutional Neural Network and Support Vector Machine, shortened as CNN and SVM, for text arrangement and identification, has been made. Before that we pre-prepared the dataset and applied different channels over it. Our framework will perceive the content precisely utilizing tensorflow libraries.

**Keywords:** Hand Written Text, CNN, SVM and MNIST.

## 1. INTRODUCTION

Handwritten Text Recognition is an innovation that is truly necessary in this world starting today. Before appropriate execution of this innovation we have depended on composing writings with our own hands which can bring about mistakes. It's hard to store and access actual information with productivity. Physical work is needed to keep up with legitimate association of the information. Since forever, there has been extreme loss of information due to the conventional technique for putting away information. Current innovation is allowing individuals to store the information over machines, where the capacity, association and getting to of information is moderately simpler. Embracing the utilization of Handwritten Text Recognition programming, it's simpler to store and access information that was customarily put away. Besides, it gives greater security to the information. One such illustration of Handwritten Text Recognition programming is the Google Lens. The point of our venture is to perceive the penmanship utilizing ideas of AI. We moved toward our concern utilizing CNN and SVM as they give better precision over such assignments.

## 2. RELATED WORK

Since seemingly forever, human used to compose their considerations as letter, records and so on; to pass on them to other people. However, since the advancement of PC innovation the configuration of transcribed content changed quickly to PC created computerized text thus individuals feel a need of such strategy that can change the manually written content to advanced content as it makes the preparing of

such information exceptionally quick and simple. Many investigators attempted to foster such framework in past. However, there is as yet a need of substantially more examination in this field. Numerous acknowledgment reads have been made for disconnected and online manually written characters of significant dialects utilized around the world: like English, Chinese, Indian scripts like Devanagari, Malayalam yet they all endure with a type of downside: like low change speed, low precision, higher bogus identification rate and horrible showing with uproarious information and so forth In this manner, acknowledgment investigations of transcribed person picture tests actually stay significant in light of their tremendous application possibilities. This paper proposes a proficient methodology towards the advancement of manually written content acknowledgment frameworks. Convolutional Neural Network (CNN) is used in this Paper utilizing neural organization approach. Two diverse learning calculations are used in this paper. CNN calculation and SVM calculation.

### 3. METHODOLOGY

This methodology is applied to achieve the objectives of the study. This method is about CNN algorithm that is based on neural network approach. In this paper the algorithm improves the performance. In this section the proposed methodology is given below in the detail.

#### ***Module Description***

- Segmentation Module
- Training Module

#### ***Segmentation Module***

Segmentation module is vital for this undertaking as its yield will be the contribution of the other module.

- ❖ Read the picture
- ❖ Detecting the letters
- ❖ Bounding and trimming
- ❖ Resizing

#### ***❖ Read the picture***

We have numerous ML libraries like Pillow, openCV and so forth for performing procedure on pictures. Here openCV is utilized to peruse and control pictures. A picture is perused and afterward put away in numerous duplicates for performing various activities.

In the wake of perusing the picture is plotted in its shape to ensure it is perused consummately. That picture contains letters that should be pictures each edited into 28 X 28 pictures before the finish of division model.

#### ***❖ Detecting the letters***

Object identification is a Computer vision strategy that detects certain segments from in a picture or a video. It utilizes Machine Learning and Deep Learning Algorithms to yield great outcomes. Identifying the letters is same as recognizing objects. We need to apply some standard channels to the information picture for accomplishing this assignment.

#### ***Stage 1: Convert a BGR picture to Greyscale picture.***

A picture with 3 channels is a BGR picture however a Grayscale picture comprises of a solitary channel. A channel is a third measurement of a picture.

#### ***Stage 2: Applying Gaussian haze to the Greyscale picture.***

This progression is done to eliminate any commotion and aggravations in the picture. In the event that the picture is obscured the shading intensities can be perceived effectively from the picture. The obscuring is in fact called Gaussian Blur in the Computer Vision.

***Stage 3: Otsu Thresholding A standard step for object discovery***

It is estimation of the proportion of spread for the pixel levels each side of the limit.

***Stage 4: Finding and Drawing forms***

The `findcontours()` and `drawcontours()` are the techniques utilized for finding and drawing shapes which are by and large boundaries of an identified article in a picture. The `drawcontours` need picture, identified shapes, shading and measurements of the boundary as the boundaries.

***❖ Bounding and Trimming***

This piece of code is added prior to editing to add spaces between letters. In the rundown that contains bouncing subtleties, we add a space string if the distance between the relating x coordinates is  $> 50$  pixel esteems addressing they should be isolated by a space. And afterward Bounding and Cropping for non-spaces and putting away them in a list `img`. The components in the rundown of distinguished articles might be the bounding rect upsides of a recognized item from the picture or a space string. The identified item is viewed as a letter just if its stature is more prominent than 20 pixels. (Our presumption)

Then for those which are viewed as letters we use bounding rect qualities to edit the letter from another duplicate of the first picture put away in another variable and annex every one of the trimmed pictures which are in numpy cluster structure into a python list.

***❖ Resizing***

For each edited picture in the rundown, we resize them to 28 X 28 pixels. We do as such on the grounds that the yield pictures from this module which will be the contribution of the other module should be pictures of size 28 X 28 pixels as the preparation information of that module is of that format. Each resized pictures should be changed over to Grayscale to coordinate with necessities of preparing model. And afterward sent as contribution to the preparation model.

***Training Module***

A model is prepared by utilizing past information and AI algorithms. It gains from the past information by highlight extraction and examples. In this undertaking Convolutional Neural Networks are utilized. We split the information into preparing and testing in the proportion of 80:20. The dataframe with credits, the dataframe with just name section, `train_size` or `test_size` and shuffled are significant parameter of `train_test_split`. Shuffled will be True naturally which shuffles the information prior to splitting. This strategy need not generally return same yield.

Scikit-learn library is utilized to change the type of information. We need to change over the trait esteems to the buoy datatype and their names into the categorial structure to prepare them. The characteristic qualities are changed over to coasting point numbers going from 0 to 1 which prior were 0 to 255. A pixel worth of 0 will be 0.0000, a pixel worth of 1 will be 1.0000, and a pixel worth of 128 will be 0.5000.

Categorial structure brings about a rundown of size rises to number of all potential names wherein an occasion with lable worth I will have  $I = 1$  and different components 0. Indeed, even these qualities should be in float. The categorial structure will appear as though this. Now we can continue to prepare our information by utilizing standard Machine Learning Algorithms.

A succession of covered up layers are made for certain hubs in every one of them. The previously covered up layer is 2Dimensional Convolution layer with kernal size 5 X 5 and 32 hubs. The enactment work utilized is relu. And afterward max pooled with 2 X 2 and over fitting is diminished utilizing dropout of 0.3. Then, at that point two additional layers are added with 128 and n hubs separately, where n is the quantity of potential yields (here  $n = 36$ ). Now we assemble the model by categorical cross entropy misfortune work, adam enhancer with a measurement of

exactness.

### Convolutional Neural Network Architecture

An essential convolutional neural organization includes three parts, specifically, the convolutional layer, the pooling layer and the output layer. The pooling layer is discretionary now and again. The commonplace convolutional neural organization design with three convolutional layers is very much adjusted for the order of written by hand pictures as displayed in Figure 1. It comprises of the information layer, different secret layers (redundancies of convolutional, standardization, pooling) and a completely associated and a yield layer.

Neurons in a single layer interface with a portion of the neurons present in the following layer, making the scaling simpler for the higher goal pictures. The activity of pooling or sub-testing can be utilized to diminish the components of the info. In a CNN model, the info picture is considered as an assortment of little sub-locales called the "open fields". A numerical activity of the convolution is applied on the info layer, which imitates the reaction to the following layer. The reaction is fundamentally a visual boost. The point by point depiction is as per the following.

#### Input layer

The information is stacked and put away in the info layer. This layer depicts the stature, width and number of channels (RGB data) of the info picture.

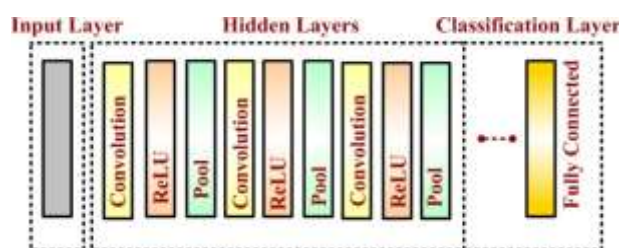


Fig. 1 Convolutional Neural Network Architecture

#### Hidden layer

The hidden layers are the foundation of CNN design. They play out a component extraction measure where a progression of convolution, pooling and initiation capacities are utilized. The discernable highlights of transcribed digits are identified at this stage.

#### Convolutional layer

The convolutional layer is the principal layer put over the information picture. It is utilized for removing the highlights of a picture. The  $n \times n$  input neurons of the info layer are tangled with a  $m \times m$  channel and consequently convey  $(n - m + 1) \times (n - m + 1)$  as yield. It presents non-linearity through a neural actuation work. The fundamental supporters of the convolutional layer are open field, stride, dilation and padding. The idea of cushioning is acquainted in CNN design with get more precision. Cushioning is acquainted with control the contracting of the yield of the convolutional layer.

The yield from the convolutional layer is an element map, which is more modest than the info picture. The yield include map contains more data on center pixels and consequently loses heaps of data present on corners. The lines and the segments of zeros are added to the boundary of a picture to forestall contracting of the component map.

#### Pooling layer

A pooling layer is added between two convolutional layers to lessen the info dimensionality and thus to diminish the computational intricacy. Pooling permits the chose qualities to be passed to the following layer while abandoning the pointless qualities. The pooling layer additionally helps in

include determination and in controlling overfitting. The pooling activity is done freely. It works by extracting just one yield esteem from the tiled non-covering sub-locales of the info pictures. The normal kinds of pooling activities are max-pooling and avg-pooling (where max and avg address maxima and normal, individually). The maximum pooling activity is for the most part positive in present day applications, since it takes the greatest qualities from each sub-locale, keeping greatest data. This prompts quicker combination and better speculation.

### *Activation layer*

Very much like ordinary neural organization design, CNN engineering likewise contains the actuation capacity to present the non-linearity in the framework. The sigmoid capacity, redressed straight unit (ReLU) and Softmax are some popular decisions among different actuation capacities abused broadly in profound learning models. It has been seen that the sigmoid actuation capacity may debilitate the CNN model as a result of the deficiency of data present in the info information. The enactment work utilized in the current work is the non-straight amended direct unit (ReLU) work, which has yield 0 for input under 0 and crude yield in any case. A few benefits of the ReLU initiation work are its likeness with the human nerve framework, straightforwardness being used and capacity to perform quicker preparing for bigger organizations.

### *Classification layer*

The classification layer is the last layer in CNN engineering. It is a completely associated feed forward network, fundamentally received as a classifier. The neurons in the completely associated layers are associated with every one of the neurons of the past layer. This layer computes anticipated classes by recognizing the information picture, which is finished by consolidating every one of the highlights learned by past layers. The quantity of yield classes relies upon the quantity of classes present in the objective dataset. In the current work, the characterization layer utilizes the 'softmax' initiation work for grouping the produced highlights of the info picture got from the past layer into different classes dependent on the preparation information.



Fig. 2 Flow chart of methodology

## 4. RESULTS

In this part, CNN has been applied on the MNIST dataset to notice the variety of exactnesses for manually written digits and text. The exactnesses are acquired utilizing tensorflow in python. Preparing and approval exactness for 15 distinct ages were noticed trading the secret layers for different mixes of convolution and secret layers by taking the clump size 100 for every one of the cases. Figure 3, 4, 5 and 6 shows the exhibition of CNN for various blends of convolution and hidden layers.

The outcomes are very useful for a particularly straightforward model with CPU preparing and less preparing time. In spite of the fact that there are a few digits and text which are not a decent penmanship, our model will actually want to order them accurately. For instance, our model orders the accompanying picture as 'little, are, ab9 and 123. This model recognizes right acknowledgment of awful hand composing.



Fig. 3 Sample data

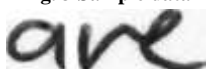


Fig. 4 Sample data



Fig. 5 Sample data

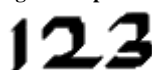


Fig. 6 Sample data

Testing precision 97% suggests that the model is prepared well for expectation. Preparing set size influences the exactness and precision increments as the quantity of information increments. The more information in the preparation set, the more modest the effect of preparing mistake and test blunder, and eventually the precision can be improved.

### *Mnist dataset*

Adjusted National Institute of Standards and Technology (MNIST) is a huge arrangement of PC vision dataset which is widely utilized for preparing and testing various frameworks. It was made from the two uncommon datasets of National Institute of Standards and Technology (NIST) which holds parallel pictures of written by hand digits. The preparation set contains manually written digits from 250 individuals, among them half preparing dataset was representatives from the Census Bureau and its remainder was from secondary school understudies [26]. In any case, it is normal credited as the first datasets among other datasets to demonstrate the viability of the neural organizations.

The information base contains 60,000 pictures utilized for preparing too as not many of them can be utilized for cross-approval purposes and 10,000 pictures utilized for testing. Every one of the digits are grayscale and situated in a fixed size where the force lies at the focal point of the picture with  $28 \times 28$  pixels. Since every one of the pictures are  $28 \times 28$  pixels, it frames an exhibit which can be leveled into  $28 \times 28 = 784$  dimensional vector. Every segment of the vector is a twofold worth which depicts the force of the pixel.



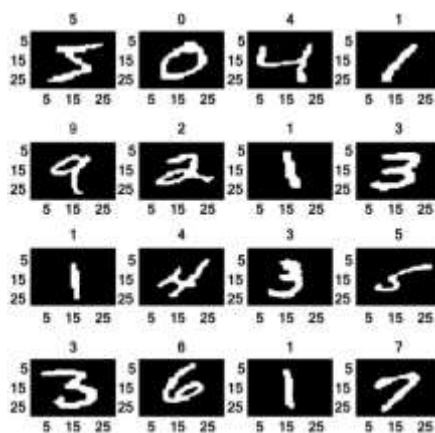


Fig. 7 Sample MNIST handwritten digit images

## 5. CONCLUSION

Convolutional Neural Network and SVM gains from the dataset and works on model by lessening the quantity of boundaries and subsequently gives impressive exactness. Further convolution and mystery layers can make the outcomes further solid. Manually written digit acknowledgment is a great trial issue for finding out about neural organizations and it gives a fantastic method to further develop more complex profound learning procedures. It tends to be extended in the future to character acknowledgment, penmanship progressively, perusing automated bank check numbers, signature confirmation, translation of postal addresses, and so on.

In future we are wanting to stretch out this examination to a bigger degree where diverse inserting models can be considered on huge assortment of the datasets. This work additionally reached out to the person acknowledgment for different dialects and for cursive penmanship with more exactness. What's to come is totally founded on innovation so nobody will utilize the paper and pen for composing. In that situation they will use to compose on touch cushions so the inbuilt programming which can naturally distinguishes and perceive the content which they have written in touch cushions and convert into computerized text so the looking and comprehension is especially improved. In future the work might be stretched out on multilingual content acknowledgment framework with more exactness and more productivity.

In future, various structures of CNN, in particular, half breed CNN, viz., CNN-RNN and CNN-HMM models, and space explicit acknowledgment frameworks, can be examined. Transformative calculations can be investigated for enhancing CNN learning boundaries, to be specific, the quantity of layers, learning rate and part sizes of convolutional channels.

In this work, fully intent on working on the exhibition of transcribed digit acknowledgment, we assessed variations of a convolutional neural organization to stay away from complex pre-preparing, expensive component extraction and an unpredictable troupe (classifier mix) approach of a conventional acknowledgment framework. Through broad assessment utilizing a MNIST dataset, the current work proposes the job of different hyper-boundaries. We additionally confirmed that tweaking of hyper-boundaries is fundamental in working on the exhibition of CNN design.

## REFERENCES

- [1] Hand-Written and Machine-Printed Text Classification in Architecture, Engineering & Construction Documents, Supriya Das;Purnendu Banerjee;Bhagesh Seraogi;Himadri Majumder;Srinivas Mukkamala;Rahul Roy;Bidyut Baran Chaudhuri, *2018*
- [2] A Comparative Study of Various Techniques and Challenges for Hand Written Document Processing of Indian Script, Ashwin R. Dobariya;Maulik Tanna, *2019 6th International Conference on Computing for Sustainable Global Development*
- [3] Personalized Hand Writing Recognition Using Continued LSTM Training, Pragya Paramita Sahu;Vikrant Singh;Indra Kiran;Viswanath Veera;Thanda Abhinav;Ankit Vijay;Shankar M. Venkatesan, *2017 14th IAPR International Conference on Document Analysis and Recognition (ICDAR)*
- [4] Knowledge Transfer Using Neural Network Based Approach for Handwritten Text Recognition, Rathin Radhakrishnan Nair;Nishant Sankaran;Bharagava Urala Kota;Sergey Tulyakov;Srirangaraj Setlur;Venu Govindaraju, *2018 13th IAPR International Workshop on Document Analysis Systems (DAS)*
- [5] Crossing the lines: making optimal use of context in line-based Handwritten Text Recognition, J. Tanha;J.D. Does;K. Depuydt;J.A. Sánchez, *2015 13th International Conference on Document Analysis and Recognition (ICDAR)*.