International Journal of Scientific Research in Computer Science, Engineering and Information Technology
© 2020 IJSRCSEIT | Volume 6 | Issue 3 | ISSN : 2456-3307
DOI : https://doi.org/10.32628/CSEIT206396

Handwritten Digit Recognition using Image Preprocessing and CNN

Anshul Dubey, Ashley Lazarus, Dharmendra Mangal
Computer Science Department, MediCaps University, Indore, Madhya Pradesh, India

ABSTRACT

Handwritten digit recognition, is a technique of identifying and enlisting the recognized digit, that uses neural networks, deep learning and machine learning. The applications and demand of handwritten digit recognition systems such as zip code recognition, car number plate recognition, robotics, banks, mobile applications and numerous more, are soaring every day. It can be done through numerous approaches, but convolutional neural network is considered one of the best methods. The special neural network uses multilayer architecture for identification and classification. Although the accuracy factor can be increased, based on image preprocessing, in this paper we discuss how the accuracy of the system can be increased for better handwritten digit recognition, using convolutional neural networks, image preprocessing; binarization, resizing, rotation. The accuracy rate obtained is 99.33%.

Keywords : Digit Recognition, Image Processing, CNN.

I. INTRODUCTION

Digit recognition is an application of a computer system which makes it able to recognize the handwritten inputs like digits from various kinds of sources like images, papers, emails, letters and more. This ability of a computer is very important as it can reduce manpower in areas like postal address interpretation, bank check processing, signature verification, etc. This topic has been for a lot of years a big topic of research. A lot of classification techniques using Machine Learning have been developed and used for this like K-Nearest Neighbors, SVM Classifier, Random Forest Classifier, etc. but these methods although have enough accuracy, are not enough for the real-world applications. One example of this is, if you send a letter with postal code as "452311" and the system detects and recognizes it as "452811" then it will not be able to deliver the letter due to incorrect postal code [1]. Thus, accuracy in these applications is very critical but these techniques do not provide the required accuracy due to very little knowledge about the topology of a task. Here comes the use of Deep Learning. In the past decade, deep learning has become the hot tool for Image Processing, object detection, handwritten digit and character recognition etc. A lot of machine learning tools have been developed like scikit-learn, SciPy-image and pybrains, Keras, Theano, TensorFlow by Google, TFLearn etc. for Deep Learning. These tools make the applications robust and therefore more accurate, also the Artificial Neural Networks can almost mimic the human brain and are a key ingredient in image processing field. We applied CNN model, using Keras library in python, for handwritten digit recognition, with improves image preprocessing. Implementing a better and accurate approach to perceive and foresee manually written digits from 0 to 9 from images [2]. Previously many problems have been faced in this research area one of them which we will be solving is the preprocessing part, so as to make it perfect for input. Improving the image for better classification and identification, with methods like binarization, rotation, resizing. This can possibly overcome problem of current pre-processing issues and result us in a non-distorted, clear image so that we can deploy this highly accurate algorithm in real world application.

II. METHODS AND MATERIAL

A. CNN and Previous Work

Convolutional neural network is a special type of neural network, which are computing systems that can perform specified tasks, by learning algorithms and examples. These are inspired by the neurons and neural network in human beings, in artificial neural network there are nodes and connections between these nodes make a connected neura network.

Convolutional neural networks or CNN dominate in image, video processing and recognition, are artificial intelligent systems that apply deep learning techniques for classification, processing and recognition tasks. In CNN, there are one or more convolutional layers, which get the input values from the previous, multiple layers, that are similar. The input values share common weights and are similar [3]. Convolutional neural networks are extensively used for image recognition systems such as digit recognition. Several projects have been made on digit recognition using numerous algorithms and models like CNN, k- nearest neighbor, random forest classifier, support vector machine etc. Research has shown CNN has the best results out all other methods [4]. Combinations of the several algorithms such as KNN-SVM, or with extra filters such as Gabor filters [5]. The results from CNN, can be further increased, if the inputs provided to the neural network are distinctive, clear and easy to process, in the image preprocessing stage.

B. MNIST Dataset and selecting algorithms

The MNIST database or Modified National Institute of Standards and Technology database, is a huge database of gray scaled images of handwritten digits from 0 to 9. It's broadly used in machine learning, image processing and training. We obtained and visualize the dataset; it has 70,000 total images, out of which 60,000 are training and 10000 for testing. The images are represented in square size of 28x28 pixels. All the images are written differently, for instance the digit 1 is written in slightly different manner than other in terms of size, alignment and same for the other numbers, as in the figure 1.

Figure 1. Handwritten digits from MNIST dataset

The data obtained can be classified using a few algorithms: kNN, SVM, RFC, CNN. kNN is k-nearest neighbor algorithm, that performs classification and regression, according to the character data in the neighborhood in the feature space. kNN showed accuracy of 96.20% [6]. SVM or support vector machine is also a classification algorithm which uses hyperplane for classification. SVM showed accuracy of 97.3% [7]. CNN or convolutional neural network, has been proven to be a method of high accuracy as compared to others. CNN has achieved 98.10% accuracy [8]. The results from CNN can be expanded, if the image preprocessing is performed with precise rotation, resizing and binarization.

C. Image preprocessing

A higher accuracy can be achieved for any image recognition, depending upon the preprocessing process [9]. Once the training of the model is done we are going to test our model on the test images from the MNIST dataset but before testing our model we first have to make the input image ready so that we can convert an RGB bigger sized image to a grey scale 28*28 pixel image because the image in the dataset are in the form of grey scale and having a size of 28*28 pixel for doing so we will use some of the algorithms provided by OpenCV [10]. One thing we should keep in mind isthat we can also test our model on images of dataset itself and in doing so we don't need any kind of image pre- processing [11].

1). Grey-scaling of RGB image: Grey-scaling of an image is a step by which an RGB image is converted into a black and white image. This process is very important for Binarization because after rescaling of the image, only shades of grey remains in the image, Binarization of such image is now more efficient than before [12].

1). Binarization: Binarization of an image includes converting it into an image which only has pure black and pure white pixel values in it. During Binarization of a grey-scale image, pixels with intensity lower thanhalf of the full intensity value gets a zero-value converting them into black ones. And the remaining pixels get a full intensity value converting it into white pixels. Binarization is easy to acquire, i.e., simple digital cameras can be used together with very simple frame stores, or low-cost scanners, or thresholding may be applied to grey-level images. Binarization needs low storage, no more than 1 bit/pixel, often this can be reduced as such images are very amenable to compression [13]. Processing is simple, i.e., algorithms are in most cases much simpler than those applied to grey-level images.

2). Rotation: For usage in the real world out there our project should also be able to read handwritten digits from various angles and of different orientation. Thus, rotation of the digits extracted from the image becomes very necessary. Rotation is the changing of the orientation of the digit in the pre-processing step for making the recognition part easier for out project [14]. We are going to train our model to make it capable of rotating the digits by itself for more efficient recognition of handwritten digits. We are going to train our model with the help of a dataset called MU Hand Images ASL. The MU Hand Images ASL dataset comprise of 700 images of each digit collected by 5 individuals. Each one of them has multiple images which are taken by them from different positions of camera and lighting. There are 70 pictures of each type of digit [15]. For image processing, the images collected images are of random sizes. Now to feed our model, the images we putshould need to be of equal size. Thus, we are going to use image scaling technique to resize each image to 28*28 [16].

Preparation of new dataset, we are using uniform random rotation in range [0degree, 360degree] to create our rotates image. We have created a number of rotated images from each fixed angle image to increase the efficiency of our project. Therefore, the images of our dataset consist of 7000 images.

TABLE I DIGITS AND SAMPLES

Digit	No. of samples
0	700
1	700
2	700
3	700
4	700
5	700
6	700
7	700
8	700
9	700
Total	7000

3). Resizing: Resizing is the increasing or decreasing of pixels in an image without distorting the quality of it. Our model requires 28 * 28 pixels as the input image size. Anything more or less will face problems and thus will affect the accuracy of our project. Not all of our images are the exact size we need them to be thus resizing becomes necessary [17].

The Connected components is the data extracted from an image(2D), which are clusters of pixels with the same value, which are connected to each other through either 4-pixel, or 8-pixel connectivity. These CCs left in this process of resizing, are most probably hand written digits. Thus, they are filtered CCs which can now be resized to the standard size of 28*28 pixels. Conventional resizing of images causes in poor data, which is very difficult for recognition. For example, the digit '8' is very tough to recognize after its conventional resizing because of its large aspect ratio [18]. A very smart resizing algorithm is necessary for image to nicely adapt space available without causing harm to the image so that the resized image is smaller in size, but its aspect ratio is similar as that of the original image. In this algorithm, firstly the resizing of CCs is done dynamically to match the height of 60 pixels or to match width of it around 40 pixels so as to maintain the aspect ratio of the components. Next, the placement of CCs will be done in standard sized image of 28*28 pixels in the top left corner.

As the images in the dataset are of 28 * 28 pixels thus we need to resize the input image to 28*28 pixel.

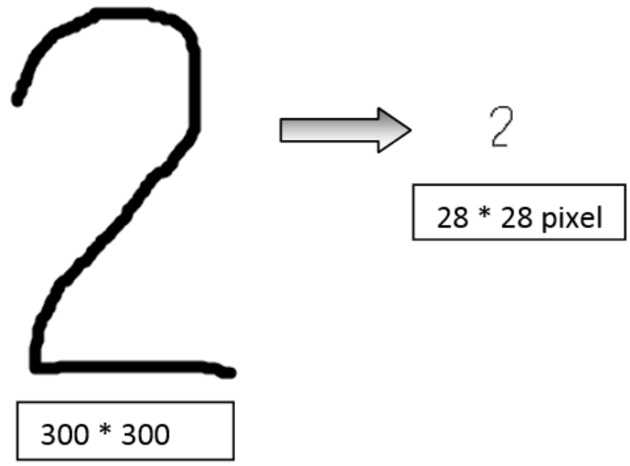

Figure 2. Resizing the image from 300*300 to 28*28 pixel

D. The CNN Model

Convolutional neural network is a multilayer network that has the essential convolutional layer or layers. This layer is vital for the image processing part. The other layers include ReLU, pooling and fully connected layer, which are as important as the convolution layer. CNN architecture is best for the identification, classification and extraction of any image. The model processes the whole image by dividing the image into small sections called features, these features undergo through a series of processes, in the layers, which then produces a result as an identified digit. The model consists four major layers: Convolution, ReLU, pooling and fully connected layer [19].

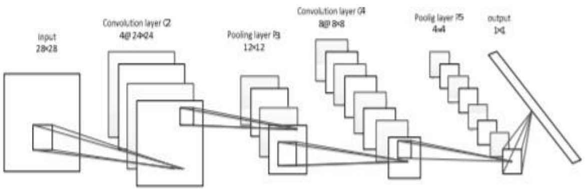

Figure 3. CNN Architecture

The convolution layer divides the image into small segments called features. The first layer has 20 convolution filters, which go over a sliding window of 5x5, on the image. The image of 28x28 pixels, matrix size. The layer produces the pixels with most intensity values, called convolutional feature map, to the next layer. ReLU or rectified linear unit layer, is an activation layer. It acts as an activation function for the values from the first layer. It filters the values and ensures that only the matching and required values pass through the layer. The intensities with higher value are of dark shade as compared to the intensities with lower value. The darkest pixel shade is black having an intensity value of 255 and white has 0. High and low intensity values are identified and passed further, to the pooling layer. This layer performs dimension reduction, it reduces the size of the sample image and makes a small image, by pooling the pixel values from ReLU layer. The processing of this layer is quick, as the sample size is reduced. The small images are again put through the layers from convolution to pooling, until a smallest and precise set of pixels is obtained, called a pooled feature map. This map is further feed into the fully connected layer. The layer is fully connected to all the other layers in the network. Classification of the data is performed by this layer, using SoftMax classifier. The classifier, from the input data, normalizes it and produces a probability distribution of the predicted 10 output classes. It returns a probability value for each of the 10 class, 0 to 9. The highest probability is selected for the final classification, for the network and the digit is recognized.

III. RESULTS AND DISCUSSION

The CNN model, took 1 hour 10 minutes for training, and produces an accuracy rate of 99.33%. The model took 15 epochs, for training 60,000 samples and validating 10,000 samples. The model works better for images of size 28x28 pixels, when the images aren't of that pixels, the model doesn't function proper. But with improved image preprocessing, the model showed growth. CNN is exceptional for image processing, kNN, SVM don't provide as high accuracy rate as of CNN. The graph below, depicts the comparison amidst the three algorithms, and the bar for CNN is quite high. The model was successful for better handwritten digit recognition.

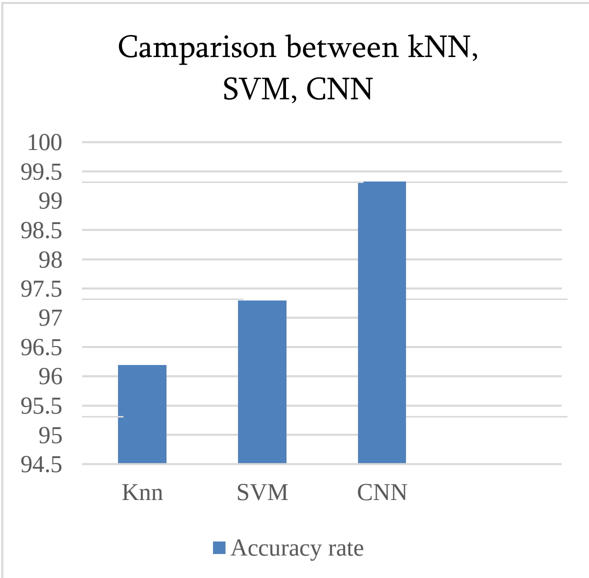

Figure 4. Comparison between kNN, RFC, CNN.

IV. CONCLUSION

In this paper, we applied CNN model for handwritten digit recognition, with refined image preprocessing. CNN is exceptional approach for image processing problems, and with better preprocessing, the results are even better. We applied binarization, resizing, rescaling, on the MNIST dataset, that has 70,000 images of handwritten digits, and then feeding the refined images as input for CNN model. CNN processed the images through the multiple layers performing classification, extraction and finally producing a recognized image of the digit. The model accomplished an accuracy rate of 99.33%. We also, compared the CNN model to other algorithms such as SVM, kNN and find out it is superior to the two. Although the CNN model requires time for training and testing of the data, the results it produces are commendable, which makes it convenient and pragmatic for future usage.

V. REFERENCES

[1]. G. Hinton, L. Deng, and D. Yu: Signal Processing Magazine, Vol. 29 (2012) No.6, p.82.

[2]. S. Ji, W. Xu, and M. Yang: Pattern Analysis and Machine Intelligence, Vol. 35 (2013) No.1, p.221.

[3]. Meiyin Wu and Li Chen, “Image Recognition Based on Deep Learning”, Chinese Automation Congress (CAC), IEEE, pp. 542-546, 2015.

[4]. Kajol Gupta, “Digit Recognition Using Convolution Neural Network”, Computer Vision and Pattern Recognition, 2020.

[5]. Xiao-Xiao Niu, Ching Y. Suen,” A novel hybrid CNN–SVM classifier for recognizing handwritten digits”, Pattern Recognition 45 (2012), pp. 1318–1325, 2012.

[6]. U Ravi Babu, Dr. Y Venkateswarlu, Aneel Kumar Chintha, “Handwritten Digit Recognition Using K- Nearest Neighbour Classifier”, 2014 World Congress on Computing and Communication Technologies, pp. 60- 65, 2014.

[7]. D. C. Cireşan, U. Meier, L. M. Gambardella, and J. Schmidhuber, “Deep, big, simple neural nets for handwritten digit recognition,” Neural Comput., vol. 22, no. 12, pp. 3207–3220, 2010

[8]. Tanya Makkar, Yogesh Kumar, Ashwani Kr Dubey, Alvaro Rocha, & Ayush Goyal, “Analogizing time complexity of KNN and CNN in recognizing handwritten digits” 2017 Fourth International Conference on Image Information Processing (ICIIP), pp. 182-187, 2017.

[9]. Yahia S. Halabi, “Modelling Adaptive Degraded Document Image Binarization and Optical Character System”, online access

[10]. X. Glorot and Y. Bengio, "Understanding the difficulty of training deep feedforward neural networks," in JMLR W&CP: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics (AISTATS 2010), vol. 9, May 2010, pp. 249-256.

[11]. LUONG CHI MAI,” Introduction to Computer Vision and Image Processing E-book, Department of Pattern Recognition and Knowledge Engineering”, Institute of Information Technology, Hanoi, Vietnam 12Amy A. Gooch, Sven C. Olsen, Jack Tumblin, Bruce Gooch, Color2Gray: Salience-Preserving Color Removal, ACM SIGGRAPH, 2005, pp. 634 – 639.

[12]. B. Gatos, I. Pratikakis, and S. J. Perantonis, “Improved document image binarization by using a combination of multiple binarization techniques and adapted edge information,” 19th Int. Conf. Pattern Recognit., pp. 1– 4, Dec. 2008.

[13]. Adam, S., Ogier, J., Cariou, C., Mullot, R., Labiche, J., and Gardes, J. (2000). Symbol and character recognition: application to engineering drawings. IJDAR, 3(2):89–101.

[14]. Deseilligny, M. P., Mena, H. L., and Stamonb, G. (1995). Character string recognition on maps, a rotation invariant recognition method. Pattern Recognition Letters, 16(12):1297–1310

[15]. Pal, U., Sinha, S., and Chaudhuri, B. B. (2003). Multioriented English text line identification. In Proceedings of the 13th Scandinavian conference on Image analysis, pages 1146–1153

[16]. Wang Y S, Tai C L, Sorkine O, Lee T Y. Optimized scale and-stretch for image resizing. ACM Trans. Graph., 2008, 27(5), Article No. 118.

[17]. Huang Q X, Mech R, Carr N. Optimizing structure preserving embedded deformation for resizing images and vector art. Computer Graphics Forum, 2009, 28(7): 1887-1896.

[18]. Mahmoud M. Abu Ghosh, Ashraf Y. Maghari, "A Comparative Study on Handwriting Digit Recognition Using Neural Networks", 2017 International Conference on Promising Electronic Technologies, pp. 77-81, 2017.

Cite this article as :

Anshul Dubey, Ashley Lazarus, Dharmendra Mangal, "Handwritten Digit Recognition using Image Preprocessing and CNN", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 6, Issue 3, pp.896-902, May-June-2020. Available at
doi : https://doi.org/10.32628/CSEIT206396
Journal URL : http://ijsrcseit.com/CSEIT206396