2019 2nd International Conference on Power Energy, Environment and Intelligent Control (PEEIC)
G. L. Bajaj Inst. of Technology and Management Greater Noida, U. P., India, Oct 18-19, 2019

Handwritten Digit Recognition Using Machine Learning : A Review

[1]Anchit Shrivastava
anchit1904.cse@chitkara.edu.in

[2]Isha Jaggi
isha1913.cse18@chitkara.edu.in

[3]Sheifali Gupta
sheifali.gupta@chitkara.edu.in

[4]Deepali Gupta
deepali.gupta@chitkara.edu.in

[1, 2, 3, 4]Chitkara University Institute of Engineering and Technology, Chitkara University, Punjab, India

Abstract— **The task for handwritten digit recognition has been troublesome due to various variations in writing styles. Therefore, we have tried to create a base for future researches in the area so that the researchers can overcome the existing problems. The existing methods and techniques for handwritten digit recognition were reviewed and understood to analyze the most suitable and best method for digit recognition. A number of 60,000 images were used as training sets of images with pixel size of 28x28. The images/training sets were matched with original image. It was found out after complete analysis and review that classifier ensemble system has the least error rate of just 0.32%. In this paper, review of different methods handwritten digit recognition were observed and analyzed.**

Keywords—CNN, Handwritten digit recognition, MNIST, SVM.

I. Introduction

Handwritten digit recognition has not only professional and commercial applications, but also has practical application in our daily life and can be of great help to the visually impaired. It also helps us to solve complex problems easily thus making our lives easier [1].

Many algorithms have been developed for hand written digit recognition. But due to infinite variation in writing styles they are still not up to mark. Poor contrast, image text vagueness, disrupted text stroke, unwanted objects, deformation, disoriented patterns and also interclass and intraclass similarity also cause misclassification in handwritten numeral recognition system [2]. Handwriting styles even differ when the same person writes a digit twice at two different places. In numeral recognition, the digits are written on a paper and are converted into digital images using a classifier, SVM is the most famous classification technique. To solve the stated problem there are mainly three approaches statistical, multilayer neural network and deformable ones. This paper focuses on digit recognition using multilayer neural network. It mainly comprises of three phases that is pre-processing phase, feature extraction and then the final phase of classification.

The pre-processing phases comprises of capturing the image in a spatio-luminance representation and then reducing the noise etc. Basically in the feature extraction phase the image is divided into multiple parts and even the finest details and features are extracted from the image. Then the system is trained on extracted features using different techniques like multilayer perceptron and SVM[3]. From this, a trained model is designed that is used for classifying the test image. For this, details features are withdrawn from the test image and these features are fed to the trained model. Then applying many techniques and procedures such as Multilayer Perceptron(MLP) and Support Vector Machines(SVM) etc. [4]. This phase mainly focusses on training the system by extracting even the finest details from the image. Then comes the classifying stage in which all the details extracted from the image is compared by the trained data set and the image with which most of its characteristics matches is finally given the tag.

A CNN [1] is a deep learning method that has been extensively used in developing applications for computer vision, natural language processing, data mining, computer games and handwritten recognition[5]. LeNet5 is the base architecture of CNN[6]. LeNet-5 was one of the first multilayer neural network where the concept of Convolutional neural network (CNN) was used in 1990 [7]. As the topic was researched further, more layers were added making the system more accurate but also started to increase the computational time[8]. The computational time can be reduced to some extent by using better GPU's. But even then achieving 100% accuracy rate is practically impossible. Although in the past few years CNN has performed well but still human beings have better ability.

Mainly, the database for western roman numerals classification using deep learning is Modified National Institute of Standards and Technology database (MNIST) which comprises of about 60,000 images for training the system and about 14,000 images for testing it [9] Then, finally the trained system can be used for practical purposes such as recognizing numerals written on bank cheques, sorting of mails etc. [1].

In this paper we compare various techniques and algorithms developed by scientists over the years to recognize handwritten digits, and tried to conclude and find the technique with least error rate and also having reasonable computational time.

II. Database Used

Modified National Institute of Standards and Technology (MNIST) is a database which is freely available for handwritten digits and is standard for machine learning algorithms. It is similar to TIDigit which is a database of speech created by Texas Instruments, which tasks in speech recognition[9].

MNIST database is considered as modified form of NIST Database. In MNIST there are 60,000 images which are used for training the system, and for validation purpose. These black and white digits are distributed uniformly and centered in a image which is of fixed size with 28*28 pixels. The dimension of every sample image vector is 28*28=784. MNIST Database provides simple statics classification tasks for researchers to help them to analyze machine learning and pattern recognition techniques. It also helps them save avoidable efforts on data pre-processing and formatting. Neural network classifiers perform better than the other classifiers, specifically Convolutional Networks structure performs outstandingly. According to the record performance the error rate is only about 0.27% which means that there is only about 27 errors in the full 10,000 test set which is attained by a committee of convolutional nets (with distortions). Even without "committee" , single large neural network also gives a error rate as low as 0.35%. The use of distortions, mainly elastic the training data plays an important role in achieving very little error rates, without these distortions, the error percentage is increased from 0.35% to 0.53%.

In most researches, the existing training data from the MNIST database were used in learning the classifiers whereas in other experiments, the training set was expanded with artificially distorted versions in addition to the original training samples. The distortions include shifting, scaling, desk Skewing, deslanting, smudging, and compressing. These distortions are specified in the "Pre-processing" column .

III. Methodology

A. Feature Extraction for Handwrittern Digit Recognition

In the literature, different types of features or characteristics are extracted for the classification of handwritten numeral digits. Different algorithms used for feature extraction have different types of error rate. Errors made by each separate algorithm does not overlap, therefore we can combine all types of features extraction methods which will lead to a perfect recognition rate. This method helps to reject the ambiguous digits recognition and improve the recognition rate of misclassified digits that can be recognized by humans[7]. Here different types of features extracted in literature are given below[10].

1. *Structural Characteristics*- In this, the algorithm extracts histograms profiles and then convert it to a single feature vector. The input image is resized in a 32x32 matrix. Radial histograms are calculated by counting the number of pixels in 72 directions at 5 degree intervals.
2. *Modified Edge maps*- The input image is divided to small parts of 25x25 each. Sobel operators are used to obtain mainly four edge maps; horizontal, vertical and the two diagonals. These four maps are divided to 25 further images of 5x5 pixels each. The features obtained are then merged to make a single feature vector having 125 features.
3. *Image projections*- It mainly extracts diagonal and radial projections. For this, the image is divided into four parts like quadrants; top, bottom, left and right. This is done to remove rotational uniformity which is entirely not required. Radial projections are attained by centering the image pixels by grouping it with its radius. These normalized features are then converted to a single vector which has 128 features.
4. *Multi zoning*- In this the percentage of black pixels is used as a feature from several images subdivided. To obtain better results many different configurations are used. In this 13 different configuration features are used (3 by 1, 1 by 3, 2 by 3, 3 by 2, 3 by 3, 1 by 4, 4 by 1, 4 by 4, 6 by 1, 1 by 6, 6 by 2, 2 by 6 and 6 by 6).
5. *Concavities Measurement*- First the image is converted to a 18x15 size matrix, then it is divided to six zones, each having 13-d feature vector. For any white pixel the algorithm searches for any black pixel in all the directions. The feature vector of each zone are formed into a single vector having 78 features.
6. *Gradient Features*- It calculates the gradient elements in a grayscale image. The reason for the use of grayscale image is that it has more information. First, the input image is modified to a pseudo-grayscale using Medial Axial Transformation(MAT). To generate amplitude and phases Sobel operators are used. The image is branched to 16 sub-images and for each sub image, the number of pixels in all the eight directions is counted as a feature. The size of feature vector is 128.

Table 1: Ensemble system compared with different classifiers[10].

Method	Distortion	Error
Boosted Le-Net 4	Affline	0.70%
TFE-Support Vector Machine		0.44%
PNCN	Skewing	0.44%
Cascade Ensemble Classifier	-	0.41%
Convolution Neural Network	Elastic	0.40%
Large Convolution Network + Unsupported pre-Training		0.39%
Ensemble System	-	**0.32%**

When a three layer multilayer perceptron model is trained using resilient backpropagation algorithm on all the six types of feature, it gives much better results as compared to the case if individual features are extracted. Table 1 compares the ensemble system with various other different classifiers. Also, table 2 [10] represents the comparison of error rate for individual feature extraction as well as combined features extraction and Fig. 1 depicts the graph showing mean error rates of methods for extracting features.

B. Classification using CNN

CNN Convolution layer and the subsampling layer can have various different layers[11]. The down sampling layer is also known as pooling layer [12]. The image is divided into small segments of small areas, and a value is calculated for each area. Then the calculated values are rearranged in sequence to form a new image [5]. This process is similar to fuzzy filter, which can increase the robustness of image feature withdrawl [13]. The pooling method used in this paper is average pooling [14].

In CNN, the computational time of the network training increases with the increase of the convolution layers. The recognition results will be limited because of the simpler structure of CNN. The minimal convolutional neural network consists of [10]

1. The input layer
2. The convolutional layer
3. The Rectified linear unit (ReLU) layer
4. The Pooling layer
5. The feature mapping Representation Layer
6. The SoftMax Layer
7. The output layer

Table 2: Error Rate for different types of feature extraction And combined features extraction

Digit	Structural Features (%)	Edge (%)	Projection (%)	Concatives (%)	Zoning (%)	Gradient (%)	Combined Features (%)
0	1.22	2.14	1.83	3.87	1.17	2.04	0.10
1	0.88	1.85	1.58	1.67	2.05	1.32	0.08
2	3.97	4.74	4.74	4.34	3.77	4.84	0.19
3	3.86	5.24	5.24	8.32	3.16	5.54	0.39
4	2.75	7.85	3.67	7.02	2.95	3.06	0.30
5	4.37	5.27	6.39	4.44	3.03	3.70	0.67
6	2.19	3.34	2.82	3.65	2.92	2.61	0.20
7	3.11	6.23	4.57	5.62	4.38	4.96	0.48
8	4.00	6.46	6.26	10.36	4.10	6.46	0.20
9	4.40	9.42	6.15	7.58	4.86	7.34	0.59
Mean	3.05	5.22	4.28	5.69	3.12	4.17	0.32

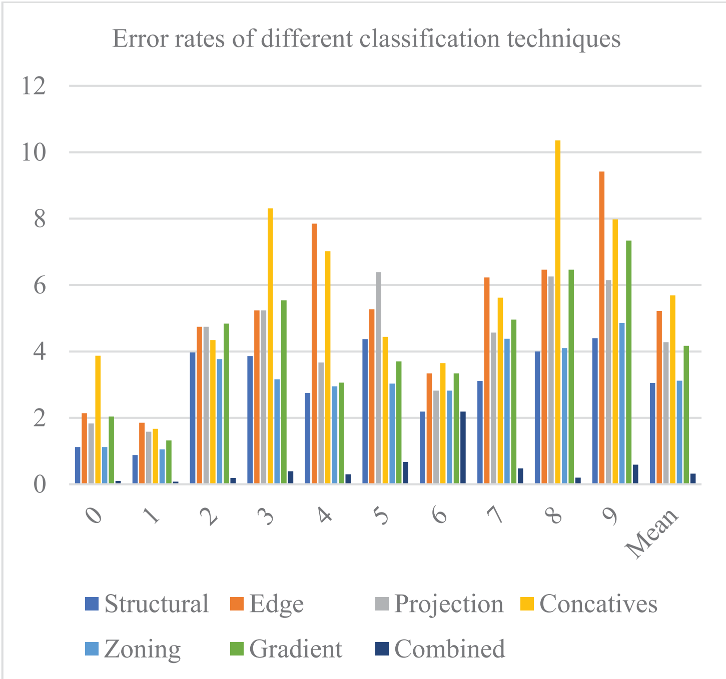

Fig 1: Graph showing mean error rates of individual feature extraction methods.

C. Classification using Pre-trained CNN and SVM

Here a system is proposed for handwritten digit recognition where a pretrained CNN is used in combination with the SVM as a classifier. For this Alex-net is used as a CNN which is pre-trained for large scale image set database having 1000 object divisions and 1.2 million training images. It takes the input image of size 227x227x3 [15].

Hence, the features of the training images are extracted using Alex-net. After that SVM is used as a classifier which is trainable for the extracted features. A multiclass SVM is trained using CNN features. Different augmentation methods like cosine translation, elastic distortion, skew and rotation are used to improve the recognition accuracy. Table 3 represents the error rate for various augmentation methods used in the combination of Alex-net and SVM.

Table 3: Error rate compared for different augmentation methods

Augmentation type	Ratio of augmentation	Patterns generated	Training patterns	Mean Error Rate	Best Error Rate
None	-	-	54200	1.336 %	1.03%
Cos Translation	1:30	1680510	504150	1.238 %	0.93%
Elastic Distortion	1:10	596310	1.248	1.248 %	0.95%
Skew	1:12	704730	507405	1.382 %	1.15%
Rotation	1:10	596300	506863	1.378 %	1.15%

Here, the best results are in case of cosine translation augmentation method.

D. CNN and Gabor Filters (GCNN)

In this to produce a large degree of invariance and geometric transformations, use of identical weight vectors are forced to be used by different receptive fields. The resulting convolutional and sub-sampling layers are varied to give a "bi-pyramidal" effect, i.e. as we move to the next layer, the abundance of depiction is increased while the spatial resolution is decreased [16].

Gabor filters are used for multiresolution examination. They illustrate the image in various level of frequencies, thus extracting various features on the basis of each filter and its frequency. Now, basically a GCNN is a six layer algorithm along with an input layer. The input layer consists of 28*28 sensory nodes, which accepts an image which is normalized in size.

In this experiment, The Gabor filter layer has a function equivalent to the Convolution layer, acting as a feature extractor. This layer consists of 12 sublayers, which are Gabor filter response to 2 different frequencies and 6 orientations, with (sigma)= 2.4. The size of each sublayer is 28x28.

Then there are two subsampling layers, which consists of neurons having receptive field size of 2x2. Then it consists of two convolutional layers, having convolutional mask used of 5x5, decreasing the rows and columns by 4 units. The output layer works as a perceptron. The layer is fully connected to the previous convolution layer, which consists of 84 neurons.

Figure 2 shows the GCNN architecture with Gabor filters compared with other similar classifiers [16].

Fig 2: GCNN model compared with other similar classifiers

IV. Conclusion

In this paper, various methods for handwritten numeral recognition based on MNIST database are compared. Over the years various scientists have proposed new methods for handwritten digit recognition which have helped in making our lives easier.

In this paper with increase the number of layers of the system, the accuracy as well as the computational time increases. Still after considerable much research and development the systems are not able to compete with the human intelligence. After comparing various classifiers, we find that the classifier ensemble proposed by Rafel et al[10] has the most accuracy but has high computational time and the best possible method for handwritten numeral recognition is 6-layer NN with least error rate of 0.35 refrence mizukami et al [17].

Acknowledgment

This research was sponsored by Chitkara University, Punjab. We acknowledge our mentor Kulbhushan Sharma who provided insight and expertise that greatly helped the research, for suggestions that greatly improved this manuscript.

We would also like to show our courtesy to Dr. Rajnish Sharma, Chitkara University for sharing his pearls of wisdom with us during the course of research, although any errors are our own and should not tarnish his reputation.

References

[1]. Xiaofeng Han and Yan Li (2015), "The Application of Convolution Neural Networks in Handwritten Numeral Recognition" in International Journal of Database Theory and Application, Vol. 8, No. 3, pp. 367-376.

[2]. Yoshihiro Shima, Meisei, Yumi Nakashima, Michio Yasuda, Meisei (2017), "Pattern Augmentation for Handwritten Digit Classification based on Combination of Pre-trained CNN and SVM", 6th international Conference on informatics, Electronics and vision (ICIEV) and 7th International Symposium n Computational medical and health technology (ISCMHT).

[3]. Xiao-Xiao Niu n , Ching Y. Suen (April 2012), "A novel hybrid CNN–SVM classifier for recognizing handwritten digits", Elsevier, Vol. 45, Issue 4, Pages 1318-1325.

[4]. Caiyun Ma, Hong Zhang (2015), "Effective Handwritten Digit Recognition Based on Multi-feature Extraction and Deep Analysis", 12th International Conference on Fuzzy Systems and Knowledge Discovery (FSKD), pp. 297-301.

[5]. Matthew Y.W. Teow Artificial Intelligence Lab (21 October 2017), "Understanding Convolutional Neural Networks Using A Minimal Model for Handwritten Digit Recognition", 2017 IEEE 2nd International Conference on Automatic Control and Intelligent Systems (I2CACIS 2017), Kota Kinabalu, Sabah, Malaysia, pp. 167-172.

[6]. Dan ClaudiuCires¸an, Ueli Meier, Luca Maria Gambardella, Jurgen Schmidhuber (March 2010), ¨Deep, Big, Simple Neural Nets for Handwritten Digit Recognition", arXiv, pp. 1-14.

[7]. Y. Le Cun, L. D. Jackel, B. Boser, J. S. Denker, H. P. Graf, I. Guyon, D. Henderson, R. E. Howard, W. Hub, "Handwritten Digit Recognition : Applications of Neural Network Chips and Automatic Learning" NATO ASI series F: Computer and system sciences, Vol. 68, pp. 41-46.

[8]. T Siva Ajay (July 2017), "Handwritten Digit Recognition Using Convolutional Neural Networks" International Research Journal of Engineering and Technology (IRJET), Vol. 04, Issue 07, pp. 2971-2976.

[9]. Li Deng (November 2012), "The MNIST Database of Handwritten Digit Images for Machine Learning Research", Best of the web series, IEEE signal processing magazine, pp. 141-142.

[10]. Rafael M. O. Cruz, George D. C. Cavalcanti and Tsang Ing Ren (2010), "Handwritten Digit Recognition Using Multiple Feature Extraction Techniques and Classifier Ensemble." 17th Internatonal conference on systems, signals and image processing (IWSSIP), pp. 215-218.

[11]. Hongjian Zhan, ShujingLyu, Yue Lu Shanghai (August 2018), "Handwritten Digit String Recognition using Convolutional Neural Network", 24th International Conference on Pattern Recognition (ICPR), pp. 3729-3734.

[12]. Alessandro Giusti, Dan C. Cires ̧an, Jonathan Masci, Luca M. Gambardella, Jurgen Schmidhuber (2010), "Fast image scanning with deep max-pooling convolutional neural networks" Published in 17th IEEE conference on image processing, pp. 4034-4038.

[13]. Vladimir Golovko, MikhnoEgor, AliaksandrBrich, and AnatoliySachenko (October 2016), "A Shallow Convolutional Neural Network for Accurate Handwritten Digits Classification" 13th international conference, PRIP, Minsk, Belarus, pp. 77-85.

[14]. Seiichi Uchida, Shota Ide, Brian Kenji Iwana, Anna Zhu (2016), "A Further Step to Perfect Accuracy by Training CNN with Larger Data", 15th International Conference on Frontiers in Handwriting Recognition, 405-410.

[15]. Yoshihiro Shima, Meisei, Yumi Nakashima, Michio Yasuda, Meisei (2017), "Pattern Augmentation for Handwritten Digit Classification based on Combination of Pre-trained CNN and SVM", 6th international Conference on informatics, Electronics and vision (ICIEV) and 7th International Symposium n Computational medical and health technology (ISCMHT).

[16]. A. Calderon, S. Roa and J. Victorino Grupo Simbiosis Bogota, Colombi (2003), "Handwritten Digit Recognition using Convolutional Neural Networks and Gabor filters" Proceedings of the International Congress on Computational Intelligence (CIIC), pp. 1-8.

[17]. Yoshiki Mizukami, Taiji Sato, Kanya Tanaka (2000), "Handwritten digit recognition by hierarchical displacement extraction with gradual prototype elimination" Published in Proceedings 15th International Conference on Pattern Recognition (ICPR), pp. 339-342.