

Character Recognition from Handwritten Image Using Convolutional Neural Networks



Ranjan Jana and Siddhartha Bhattacharyya

Abstract Character recognition from handwritten images is of great interest in the pattern recognition research community for their good application in many areas. To implement the system, it requires two steps, viz., feature extraction followed by character recognition based on any classification algorithm. Convolutional neural network (CNN) is an excellent feature extractor and classifier. The performance of a CNN for a particular application depends on the parameters used in the network. In this article, a CNN is implemented for the MNIST dataset with appropriate parameters for training and testing the system. The system provides accuracy up to 98.85%, which is better with respect to others. It also takes very low amount of time for training the system.

Keywords Handwritten character recognition · Optical character recognition · Convolutional neural networks · Deep neural networks · MNIST dataset

1 Introduction

Character recognition from printed text images or handwritten image documents is very important in the domain of optical character recognition (OCR). Some of the research areas include automatic number plate recognition, automatic postal address checking from envelopes, processing of bank cheques to name a few, as illustrated in

R. Jana (✉)

Department of Information Technology, RCC Institute of Information Technology, Kolkata, West Bengal, India

e-mail: ranjan.rcciit@gmail.com

S. Bhattacharyya

Department of Computer Application, RCC Institute of Information Technology, Kolkata, West Bengal, India

Faculty of Electrical Engineering and Computer Science, VSB Technical University of Ostrava, Ostrava, Czechia

e-mail: siddhartha.bhattacharyya@vsb.cz

© Springer Nature Singapore Pte Ltd. 2019

S. Bhattacharyya et al. (eds.), *Recent Trends in Signal and Image Processing*,

Advances in Intelligent Systems and Computing 922,

https://doi.org/10.1007/978-981-13-6783-0_3

[1]. Extraction of text from real images is always a challenging proposition in many applications [2]. Recognizing text from real images becomes very complicated due to a wide range of variation in textures, backgrounds, font size and shading. The three basic steps of character recognition are segmentation, feature extraction, and feature classification. In the domain of computer vision, the Multi-layer perceptron (MLP) has been a revolution. However, the performance of MLP fully depends on the performance of feature selection algorithms [3]. After the invention of Deep Neural Network (DNN), it is proved that, DNN is an excellent feature extractor as well as a good classifier [4]. However, it takes huge amount of time for training the network due to large numbers of nonlinear hidden layers and connections. Convolutional neural network (CNN) has been discovered to solve various problems in computer vision by using lesser amount of hidden layers compared to DNN [5]. So, CNN is used to extract the position-invariant features in a reasonable amount of time for its simple structure. CNN takes relatively few parameters and it is very easy to train the system. CNN is able to map between input dataset to output dataset with temporal sub-sampling to offer a degree of rotation, distortion, and shift invariance [6]. So, CNN is used in this article to implement the system to recognize character from handwritten image.

2 Related Work

This section surveys related works on character recognition systems. Hanmandlu and Murthy have implemented a character recognition system that used different priorities for different features based on the accuracies of individual features [7]. The recognition system has been implemented using a fuzzy model. The average recognition accuracy was 98.4% for the numeric digits in English and 95% for the numeric digits in Hindi. A recurrent neural network (RNN) using Hidden Markov Model has been used to discover patterns in a script to find the sequence of characters [8]. The model has been implemented by Graves and Schmidhuber and it classified the IFN/ENIT database of handwritten Arabic words with 91% accuracy. For recognition of handwritten English characters, Pal and Singh have implemented the Multi-layer Perceptron [9]. The features have been extracted and analyzed by comparing its features for character recognition using boundary tracing and its fourier descriptors. It has been also analyzed to determine the number of hidden layers required to attain high accuracy. It has been reported with 94% accuracy of Handwritten English characters with very less amount of training time. Neves et al. have implemented a Support Vector Machine (SVM) based offline handwritten character recognition, which gave better accuracy compared to the Multi-layer perceptron classifier for the NIST SD19 standard dataset [10]. Although, MLP is suitable for segmentation of nonlinear separable classes, but it can easily trapped into a local minima. An implementation of deep neural network models has been established by Younis and Alkhateeb so that it can extract the important features without unnecessary pre-processing [11]. Such, a deep neural network has demonstrated an accuracy of 98.46% on MNIST dataset for

the handwritten OCR problem. A multilayer CNN using Keras with Theano has been implemented by Dutt and Dutt in [12] with accuracy of 98.70% on MNIST dataset. It provided better accuracy compared to SVM, K-Nearest Neighbor and Random Forest Classifier. A comparative study of three neural network approaches has been provided by Ghosh and Maghari [13]. The results showed that Deep Neural Network was the most efficient algorithm with accuracy of 98.08%. However, it was noted that each neural network has an error rate because of their similarity in digit shape for the digit tuples (1 and 7), (3 and 8), (6 and 9), (8 and 9).

3 Overview of CNN Architecture

A CNN is a special type of artificial neural network, which consists of one input and one output layer with multiple hidden layers. The hidden layers are convolutional layers, pooling layers, and fully connected layers [6]. The network consists of repetitive convolutional and pooling layers. Finally, it ends with one or more fully connected layers.

3.1 Convolutional Layer

A convolutional layer applies sliding filters vertically and horizontally through the input image. This layer learns the features of all the regions of input image while scanning. It computes a scalar product of values of the filter with the values of image regions and adds a bias for each region. A Rectified Linear Unit applies element wise activation function, viz., $\max(0, x)$, Tanh, Sigmoid: $1/(1 + e^{-x})$ to the output of this layer for thresholding.

3.2 Pooling Layer

Pooling layer is generally used after one or more convolutional layers to shrink the volume of the data to make the network computationally faster. It restricts the networks from overfitting as well as provides the network into translation invariance. Max pooling and average pooling are generally used to implement pooling. It applies sliding filters vertically and horizontally through the input image to get max value or average value for each region of the input data.

3.3 Fully Connected Layer

After all the convolutional and pooling steps, the networks generally use fully connected layers with separate neurons for each pixels like a standard neural network. The last fully connected layer contains n numbers of neurons, where n is the number of predicted classes. For digit classification problem, it should be 10 neurons for 10 classes (digit 0–9) and for English character classification problem, it should be 26 neurons for 26 classes (character a to z).

4 Proposed Methodology

Character recognition from handwritten images has been applied for extraction of text. To implement a character recognition system, one initially needs features extraction technique supported by a classification algorithm for recognition of characters based on the features. Several feature extraction and classification algorithms have been used for this purpose before the advent of deep-learning. Deep-learning has proved that, no separate algorithm is required for feature extraction and feature classification. Deep-learning is an excellent performer in the field of computer vision for both feature extraction and feature classification. DNN architecture consists of many non-linear hidden layers with enormous number of connections and parameters. So, it is very difficult to train the network with a small set of training samples. CNN is the solution, which takes relatively few set of parameters for training the system. So, CNN is capable to map the input dataset to output dataset correctly by changing the trainable parameters as well as the number of hidden layers. Therefore, CNN architecture is proposed for character recognition from the images of handwritten digits. For performance verification of the system, the standard normalized MNIST dataset has been used.

4.1 Database Used for Training and Testing

The MNIST dataset is a subset of NIST database [14]. The MNIST dataset is a collection of 70,000 images of handwritten digits. The dataset is divided into 60,000 images for training and 10,000 images for testing [14]. All images have resolution of 28×28 and the pixel values are in the range of 0–255 (gray-scale). The background of digit is represented by 0 gray value (black) and a digit is represented by 255 gray value (white) as shown in Fig. 1.

The MNIST database is a collection of training and test set image files as well as training and test set label files. The pixel values are organized in row-wise for training and test set image files. So, the training set image file consists of 60,000 rows and 784 columns (28×28) and the testing set image file consists of 10,000 rows and

Fig. 1 Examples of MNIST data set



784 columns. In the training and test label files, the labels values are 0–9. So, the training label file consists of 60,000 rows and 10 columns (0–9) and the testing file consists of 10,000 rows and 10 columns.

4.2 Design of CNN Architecture for MNIST Dataset

The performance of a CNN for a particular application depends on the parameters used in the network. So, the CNN architecture with convolutional layers is implemented for MNIST digit recognition as shown in Fig. 2.

At first, 32 filters of window size 5×5 with a ReLU activation function for nonlinearity are used in first convolutional layer. After that, a pooling layer is used to perform a down-sampling operation using a pool size 2×2 with stride by 2. As a result, the image volume is reduced. After that, 64 filters with window size 7×7 with a ReLU activation function for nonlinearity is used in another convolutional layer. Then, another pooling layer is used to perform a down-sampling operation using a pool size 2×2 with stride by 2. After that, a fully connected layer is used with 1024 output nodes. Finally, another fully connected layer with 10 output nodes is used to get network results for ten digits (0–9).

5 Experimented Result

Experiments have been conducted in Intel Xeon 2.2 GHZ processor with 128 GB RAM using python programming. The experimented result of the proposed method is detailed in this section. A comparative study of proposed CNN with other state-of-the-art works is shown in Table 1.

The proposed CNN gives 98.85% accuracy, which is better with respect to others. The accuracy with required training time for different training steps of the proposed CNN and CNN in KERAS are presented in Table 2. The results shows that proposed CNN architecture takes less amount of time for better accuracy with respect to CNN in KERAS as shown in Fig. 3.

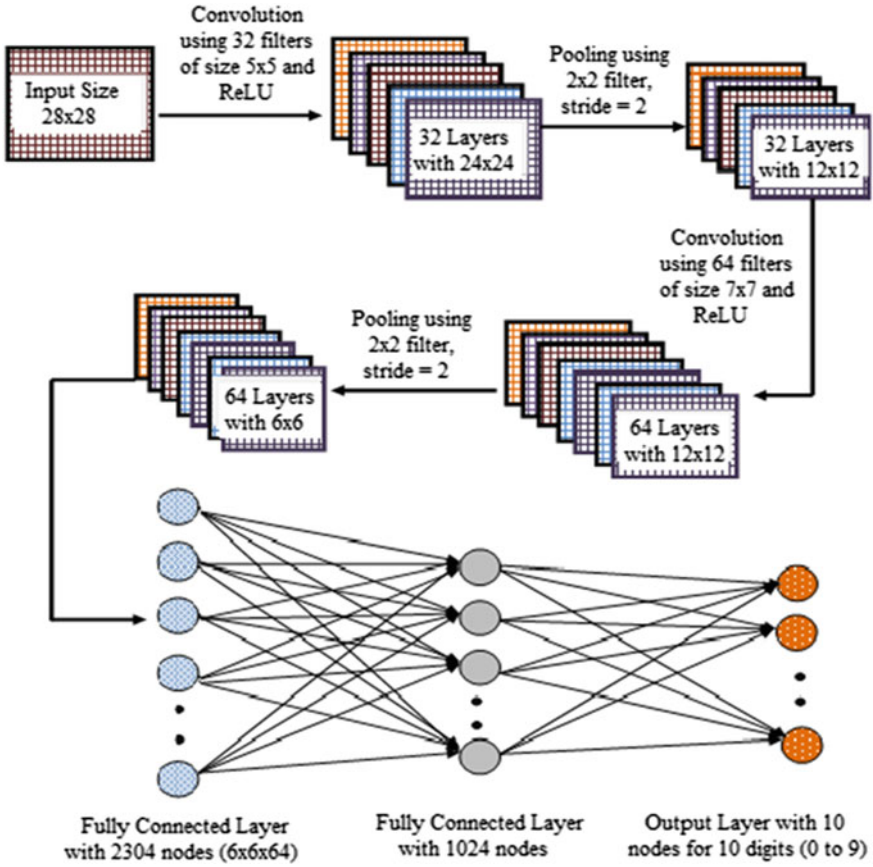


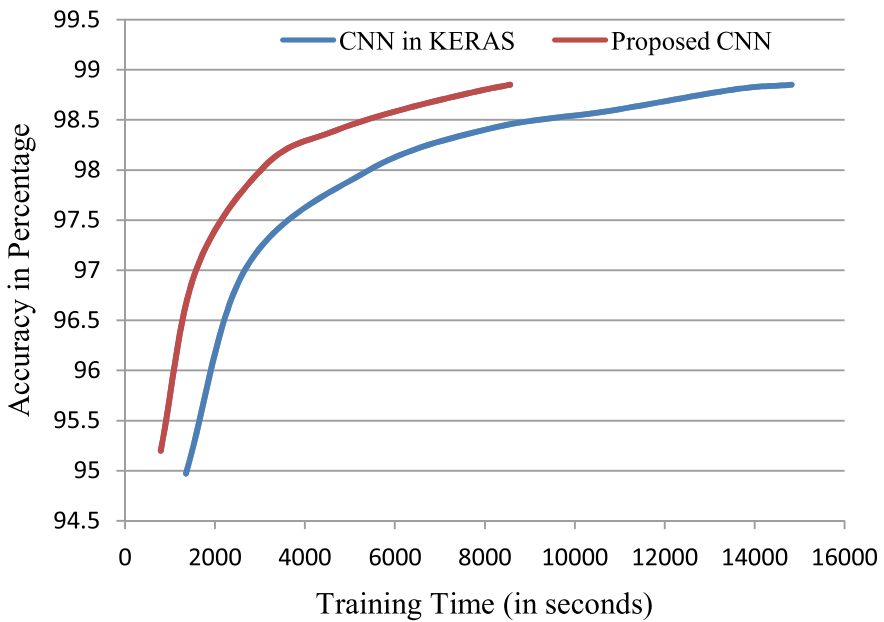
Fig. 2 The proposed CNN architecture

Table 1 Accuracy of different CNN architectures

Author	Method	Accuracy (%)
Younis and Alkhateeb [11]	CNN	98.46
Dutt and Dutt [12]	CNN (Keras + Theano)	98.70
Ghosh and Maghari [13]	DNN	98.08
Proposed method	CNN	98.85

Table 2 Performance of proposed CNN and CNN in KERAS

Training steps	CNN in KERAS		Proposed CNN	
	Accuracy (%)	Training time (s)	Accuracy (%)	Training time (s)
10,000	94.97	1353	95.20	797
20,000	97.03	2701	96.99	1573
40,000	97.99	5394	98.05	3136
60,000	98.41	8087	98.39	4674
80,000	98.59	10,785	98.61	6232
100,000	98.80	13,508	98.78	7793
110,000	98.85	14,827	98.85	8569

**Fig. 3** Accuracy versus training time

6 Conclusion

In this paper, an implementation of handwritten digit recognition using CNN is implemented. The proposed CNN architecture is designed with appropriate parameters for good accuracy of the MNIST digit classification. The time required to train the system is also considered. The CNN architecture is designed with 32 filters with window size 5×5 for the first convolutional layer and 64 filters with window size 7×7 for the second convolutional layer. The experimented results proved that, the proposed CNN architecture is the best in term of accuracy and time for the MNIST

data set as compared to others. It is worth mentioning here that more filters can be used for better accuracy at the cost of higher training time. For further improvement of the accuracy, the system needs to do more training, which requires a huge amount of time. The parallelism technique of GPU machines can be used for better accuracy with extensive training in a less amount of training time.

Acknowledgements The authors are grateful to IT department, RCC Institute of Information Technology (RCCIIT), Kolkata for providing the requisite infrastructure for the research work. The authors are also grateful to Prof. (Dr.) Ajoy Kumar Roy, Chairman, RCCIIT for his constant inspiration.

References

1. Arica N, Vural FTY (2001) An overview of character recognition focused on offline handwriting. *IEEE Trans Syst Man Cybern—Part C: Appl Rev* 31(2):216–233
2. Epshtein B, Oyek E, Wexler Y (2010) Detecting text in natural scenes with strokes width transform. In: *IEEE conference on computer vision and pattern recognition*, pp 1–8
3. Yang JB, Shen KQ, Ong CJ, Li XP (2009) Feature selection for MLP neural network: the use of random permutation of probabilistic outputs. *IEEE Trans Neural Netw* 20(12):1911–1922
4. Lee H, Grosse R, Ranganath R, Ng AY (2009) Convolutional deep belief networks for scalable unsupervised learning of hierarchical representations. In: *26th Annual international conference on machine learning*. ACM, pp 609–616
5. LeCun Y, Bottou L, Bengio Y, Haffner P (1998) Gradient-based learning applied to document recognition. *Proc IEEE* 86(11):2278–2324
6. Krizhevsky A, Sutskever I, Hinton GE (2012) Imagenet classification with deep convolutional neural networks. In: *Advances in neural information processing systems*, pp 1097–1105
7. Hanmandlu M, Murthy OVR (2007) Fuzzy model based recognition of handwritten numerals. *Pattern Recogn* 40(6):1840–1854
8. Graves A, Schmidhuber J (2009) Offline handwriting recognition with multidimensional recurrent neural networks. In: *Advances in neural information processing systems NIPS'22*, vol 22. MIT Press, Vancouver, pp 545–552
9. Pal A, Singh D (2010) Handwritten english character recognition using neural network. *Int J Comput Sci Commun* 1(2):141–144
10. Neves RFP, Filho ANGL, Mello CAB, Zanchettin C (2011) A SVM based off-line handwritten digit recognizer. In: *International conference on systems, man and cybernetics, IEEE Xplore, Brazil, 9–12, October 2011*. pp 510–515
11. Younis KS, Alkhateeb, AA (2017) A new implementation of deep neural networks for optical character recognition and face recognition. In: *Proceedings of the new trends in information technology, Jordan, April 2017*. pp. 157–162
12. Dutt A, Dutt A (2017) Handwritten digit recognition using deep learning. *Int J Adv Res Comput Eng Technol* 6(7):990–997
13. Ghosh MMA, Maghari AY (2017) A comparative study on handwriting digit recognition using neural networks. In: *International conference on promising electronic technologies*, pp 77–81
14. <http://yann.lecun.com/exdb/mnist/>