INTERNATIONAL JOURNAL of ENGINEERING TECHNOLOGIES-IJET
Koyuncu and Koyuncu, Vol.5, No.1, 2019

Handwritten Character Recognition by using Convolutional Deep Neural Network; Review

Baki Koyuncu*, Hakan Koyuncu**

*Electrical & Electronic Engineering Dept., Faculty of Engineering and Architecture, Istanbul Gelisim University

** Computer Engineering Dept., Faculty of Engineering and Architecture, Istanbul Gelisim University

‡ Baki Koyuncu;

bkoyuncu@gelisim.edu.tr

Received: 18.02.2019 Accepted: 23.03.2019

Abstract - Handwritten character recognition is an important domain of research with implementation in varied fields. Past and recent works in this field focus on diverse languages to utilize the character recognition in automated data-entry applications. Studies in Deep Neural Network recognize the individual characters in the form of images. The reliance of each recognition, which is provided by the neural network as part of the ranking result, is one of the things used to customize the implementation to the request of the client. Convolutional deep neural network model is reviewed to recognize the handwritten characters in this study. This model, initially, learned a useful set of admittance by using local receptive areas and densely connected network layers are employed for the discernment task.

Keywords Handwritten Character Recognition, Deep Neural Network (DNN), Deep Convolutional Neural Network (DCNN).

1. Introduction

Manually handwritten character recognition is an area of research in computer vision, image contrast and style recognition. An ordinary computer achieves an ability to distinguish the characters on paper records, photographs, touch screen gadgets from different sources and converts them into machine-encoded characters. Computer application of character recognition assists in optical character reception and helps to develop frameworks of character management [1]. Picture rating is one of the symbolic issues with computers in which input pictures ought to be sent to a mark from a steady arrangement of gathering dependent pictures. In optical character recognition, (OCR), a calculation is carried out on a dataset of realized characters with the end goal of how to group the characters incorporated into the test set [2].

Previously, arrangement of calculations has been produced for characterizing letters or digits. Many of the advanced calculations in this field are brought together by digit recognition [3]. At the beginning of OCR, format coordinating and basic calculations are utilized prominently. In these calculations, the models for acknowledgment issue is made by averaging a couple of tests of letters and digits. In a considerable measure of tests, these calculations were so easy to digest the distinctive types with everything being equal would produce poor outcomes for OCR issues.

Since late 80's with the end goal of deploying larger datasets and arrangement strategies, neural network systems were utilized prevalently for recognition issues [4]. A large portion of these frameworks these days employs machine learning techniques such as neural network systems for manually handwritten character recognition (HCR). Neural systems are learning methods connected to character recognition in machine learning. Their motivation is to copy the learning task that occurs in a creature or human neural network. Being a standout amongst the most ground-breaking learning models, neural networks are valuable in mechanization of missions where the goals of an individual takes too long or is not in exact nature. A neural system can be quick at finding results and may reveal associations between observed occasions of information that humans cannot see [5].

A neural system can be deployed such as a Deep Neural Network (DNN), which utilizes in excess of one concealed layer. The contrast between neural system and deep neural system is on the depth or the quantity of concealed layers used in the system. Deep Neural Network can be a feed forward neural system which has more than one hidden layers as shown in Fig. 1 [6].

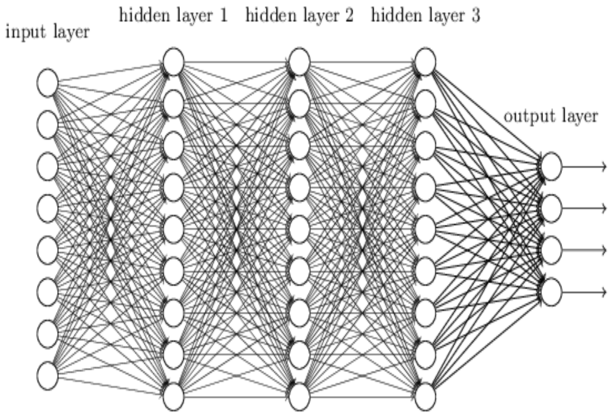

Fig. 1. Architecture of DNN [6].

DNN comprises of input layer, output layer and various intermediate layers. Hence, the quantity of associations and trainable components are very large. The deep neural system requires substantial gathering of examples to hinder over fitting and numerous regular signs have compositional arrangements [7]. In images, nearby arrangements of edges create frame themes, themes gather into parts and parts shape subjects. Comparable advances exist from sounds as well such as phonemes, syllables, words and sentences.

The pooling enables a summary of data to shift to next layers when data in the previous layer change in position and appearance. One class of Deep Neural System with generally smaller arrangement of components and simple to prepare is called Convolutional Neural System (CNN) [8][9]. CNN is an organic disclosure of multilayer perceptron (MLP). A multilayer neural network system was proposed by Fukushima, [10], and has employed manually written character recognition and other computer vision issues. LeCun [11], has utilized Convolutional Neural Network system to sort out the ImageNet dataset.

Current developments on CNN has been focused on computer vision issues such as picture division [12], picture inscribing [13] and picture grouping [14]. There has been a considerable measure of interest about manually written character digits [15] and recognition of characters in different dialects. This interest has grown due to the plausible perplexity and likeness of written characters by hand and generosity in classes.

In this study, hand written characters will be analyzed and character recognition is reviewed by utilizing, initially, DNN and later on CNN techniques. This paper organizes as follows. In Section I, the introduction and literature review is given. In Section II, related work is summarized. In Section III, CNN technique is explained and all the related models are described. Distinctive kinds of layers in CNN such as input layer, hidden layers output layer is explained in this section. In Section IV, general conclusion is given.

2. Alternative Techniques

Many researchers have developed systems for handwritten character recognition. Several important systems are mentioned in this work. Character recognition frameworks have been engineered utilizing different rationale [16]. The framework developed by some researches can be constructed by using hardware with very large scale integration circuitry (VLSI). The input character recognition of this framework is resistant to dynamic motion. Other researches utilized hamming error correcting codes from communication theory with neural network system in their framework. Another technique was developed to acknowledge the written hand characters in different dialects in its' Neural network System [17]. These frameworks generated accurate results but also made mistakes if the written hand characters are in extreme format. One of the researchers has even offered a strategy to relate the dependence between hand writers and their penmanship [18]. These studies have mostly utilized the Multi-layer feed forward neural network system in their methods.

3. Reviewed Techniques

Convolutional Neural Network Technique is basically neural network systems that utilize convolution instead of general network systems with similar number of layers. It has wide applications in fields like picture and video acknowledgment, characteristic dialect handling and recommended frameworks. From design perspective, CNN is a neural network system that utilizes no less than one convolution process in no less than one of its layers. The convolution procedure is carried out in convolution layer. There are three fundamental process that must exist in convolution layer. These are convolution, sub sampling/pooling, and actuation. A CNN incorporates a heap of convolution layer and a maximum pooling layer pursued by a complete actuation layer. The convolution layer is the most critical layer of system. It does the convolution task. The pooling layer comes after the convolution layer. This layer is essential on the grounds that if there should be an occurrence of bigger pictures, the quantity of trainable components can be exceptionally large. This expands the time taken to prepare a neural network system and this is not practical. The pooling layer is utilized to reduce the span of picture.

In this study, Modified National Institute of Standards and Technology (MNIST) database published by US department of commerce is deployed. This database contains a huge number of pictures of written hand characters. Reducing the size of the pictures reduces the general time taken to prepare the neural network system to operate. Fig. 2 shows the general view of layers present in CNN. The first layer is input layer, second layer is convolutional layer, intermediate layers are pooling layers (subsampling) and convolutional layers. Last two layers of the network are the fully connected layer and output layer.

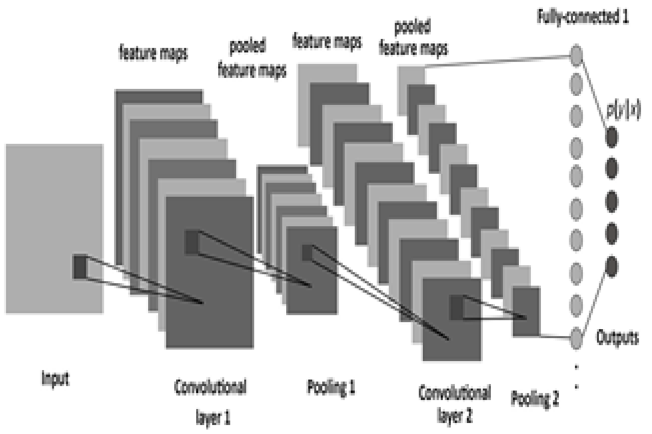

Fig. 2. Architecture of CNN [15]

A. Input Layer

The input layer is the layer of initial image which contributes to the operation at the front of the system architecture and the input is the character image as shown in Fig. 3. The input layer can be a gray scale or an RGB image. The input layer can have dimensions WxHxD where WxH is the width and height of the image and D is the depth of the image. Depth is 1 pixel for grayscale and 3 pixels for RGB images. Thus, the input layer for RGB image has dimensions 32x32x3 as seen in Fig. 3a and for Gray scale image has dimensions 32x32x1 as seen in Fig. 3b.

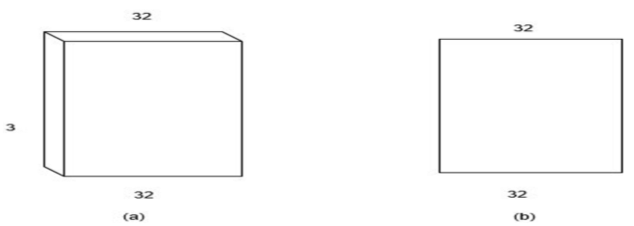

Fig. 3. Input images

B. Convolution Layer

This is the block structure after the input layer as seen in Fig. 2 where most of the computational work is done. The convolutional layer comprises of channels with learning capabilities and called components of this layer. Each channel is identified as a filter. This filter is a square matrix of spatial width and length in pixels with a depth. These channels, hence the filters, cover the full information volume. A model channel in convolutional layer can have a size of 5x5x3 pixels where 5 pixels' width, length and 3 pixels' depth. These channels are identified as shaded channels and the images used are RGB images. In this study, filters will have depth of 1 pixel and a size of 5x5x1 pixels and the character images employed are non-colored images.

During the forward pass of the neural network operation, each channel is sided widthwise and lengthwise with other channels creating a 2 dimensional information volume. Pixel intensity information of the character shapes are considered across the channels and other areas in the channels are shown as 0 pixels. As each channel is crossed through channel cross section with a width and length of the information volume, a 2-dimensional partial character outline is delivered from each channel which gives the response of that channel at each image local position.

Instinctively, the filter section is convolved over the entire image and the generated output after convolution are called initial maps as shown in Fig. 4.

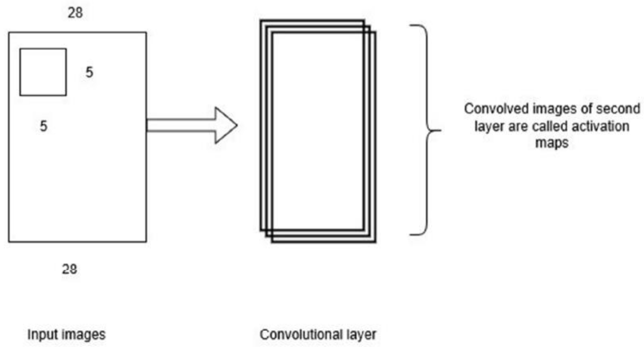

Fig. 4. Activation maps of second layer

These initial maps are called activation maps of 2nd convolutional layer. Size and number of filters depend on the experimental conditions There is no well-defined procedure to identify them. Initially, filters can contain small number of random values. But they are learnable parameters and their values will be updated with each learning stage of the network. The input layer accepts the image with dimensions WxHxD. Additional two hyper parameters such as Filter size (F) and stride (S) are deployed during convolution operation in order to generate an input for another layer with dimensions $W_1xH_1xD_1$.

W_1 and H_1 are given by equations (1) and (2). Depth D_1 remains same as D. In these equations P is called padding. It introduces new row and column of zeros on each side of image.

$$W_1 = (F - W + P)/(S+1) \qquad (1)$$

$$H_1 = (F - H + P)/S + 1) \qquad (2)$$

In this study, 32 filters were employed each with a size of 5x5x1, P=1, and S=0. Thus, the dimensions of second layer image become 32(28x28).

C. Pooling Layer

The location of Pooling layers is between convolutional layers in a convolutional architecture. Pooling layers reduce the quantity of components when the images are excessively large and also control overfitting. Additionally, local pooling called sub testing or down inspecting is introduced in this layer to eliminate the unused elements of each image and to keep the critical data. The Pooling Layer works freely on the information section at each depth and changes information at spatial dimensions. There are several types of spatial pooling. These are Maximum Pooling (MAX), Normal Pooling and Whole Pooling. MAX pooling is the most commonly used pooling. It has a pooling layer with channels of size 2x2 utilized with a stage of 2 down-inspecting local pooling. Each Maximum pooling task selects a maximum value from 4 numbers. The depth dimension stays constant. In this study, maximum filter size of 3x3, P=1 and S= 2 are deployed in first pooling layer. The output dimensions of this layer become 32(14x14). Different filters can also be employed in pooling layer.

D. Fully connected Layer

The layer identified as FC layer is the last layer of the neural network system. The image matrix arriving from pooling layer 2 with W and H dimensions are converted into 1 dimensional vector form and applied into network system. See Fig. 2. The introduction of this layer can thus improve the framework enlargement pursued by a tendency to balance. There can be multiple fully connected layers depending upon the application architecture. In this study, it is assumed that, there are 41 character classes; hence the output layer has 41 neurons. The fully connected layer has 256 neurons. The neuron number in this layer is chosen experimentally. Fig. 5 displays the complete architecture used in this study. Matlab Neural network toolbox is used in the experiments.

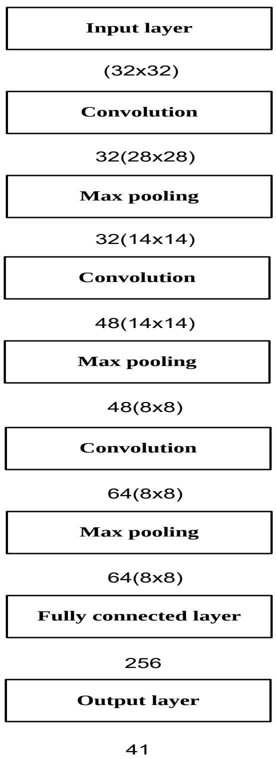

Fig. 5. CNN architecture used in this study.

4. Conclusion

A convolutional neural system is reviewed for readers' attention. This system is extremely dynamic for perceiving written hand characters. This work depends on the reception of characters at the input of (CNN). Contrast with other deep learning architectures, CNN has preferable execution in both images and big data. The aim to use deep learning was to take advantages of the power of CNN that are able to administer large dimensions of input and share their weights. The CNN architecture has thousands of elements and hyper-elements to tune. It is not clear why convolutional networks are prosperous when general back-propagation algorithms fail. It may simply be that convolutional networks work in hierarchy and solve complex framework by simpler ones.

Acknowledgements

The authors appreciated the assistance received from engineering faculty of Istanbul Gelisim University and MSc student Ali H. Abdulwahhab for his contributions in the study.

References

[1] R. Vaidya, D. Trivedi, S. Satra, M. Pimpale, "Handwritten Character Recognition Using Deep-Learning". Second International Conference on Inventive Communication and Computational Technologies (ICICCT), pp. 772-775, 2018

[2] G. S. Budhi and R. Adipranata, "Handwritten Javanese Character Recognition Using Several Artificial Neural Network Methods", J.ICT Res. Appl., vol. 8, no. 3, pp. 195–212, 2015

[3] A. Rajavelu, M.T. Musavi, and M.V. Shirvaikar, "A neural network approach to character recognition", *Neural Netw.*, vol. 2, no. 5, pp. 387– 393, 1989.

[4] S. Mori, C. Y. Suen, and K. Yamamoto, "Historical review of OCR research and development," Proc. IEEE, vol. 80, no. 7, pp. 1029 –1058, 1992

[5] J. Pradeep, E. Srinivasan and S. Himavathi. "Neural Network based Handwritten Character Recognition system without feature extraction", International Conference on Computer, Communication and Electrical Technology ICCCET 2011

[6] K. Gurney, "An introduction to neural networks", UCL Press, 1997

[7] Y. LeCun, Y. Bengio and G. Hinton, "Deep learning", Nature, Vol. 521, pp. 436-444, 2015

[8] Y. Liang, J. Wang, S. Zhou, Y. Gong, and N. Zheng, "Incorporating image priors with deep convolutional neural networks for image super resolution", Neurocomputing, Vol. 194, pp. 340- 347, 2016

[9] R. Nijhawan, H. Sharma, H. Sahni, and A. Batra, "A deep learning hybrid CNN framework approach for vegetation cover mapping using deep features", 13th International Conference on Signal Image Technology & Internet-Based Systems (SITIS), pp. 192-196, 2017

[10] K. Fukushima, "Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position", Biological Cybernetics., vol. 36, no. 4, pp. 193–202, 1980

[11] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet classification with deep convolutional neural networks," Advances in neural information processing systems, pp. 1097–1105, 2012

[12] C. Farabet, C. Couprie, L. Najman, and Y. LeCun, "Learning hierarchical features for scene labeling", IEEE Trans. Pattern Anal. Mach. Intel., Vol. 35, no. 8, pp. 1915–1929, 2013

[13] O. Vinyals, A. Toshev, S. Bengio, and D. Ethan, "Show and tell: A neural image caption generator", Proceedings of the IEEE Conference onComputer Vision and Pattern Recognition, pp. 3156–3164, 2015

[14] D. C. Ciresan, U. Meier, J. Masci, L. Maria Gambardella, and J. Schmidhuber, "Flexible, high performance convolutional neural networks for image classification", Proceedings in 22nd International Joint Conference on Artificial Intelligence, Vol. 22, pp. 1237-1242, 2011

[15] E. Kussul and T. Baidyk, "Improved method of handwritten digit recognition tested on MNIST database", Image Vis. Compute., vol. 22, no. 12, pp. 971–981, 2004

[16] W. Lu, Z. Li, B. Shi." Handwritten Digits Recognition with Neural Networks and Fuzzy Logic", IEEE International Conference on Neural Networks, Vol. 3, pp.1389-1392, 1995

[17] P. Banumathi, G. M. Nasira, "Handwritten Tamil Character Recognition using Artificial Neural Networks", International Conference on Process Automation, Control and Computing, 2011

[18] B. V. S. Murthy," Handwriting Recognition Using Supervised Neural Networks", International Joint Conference on Neural Networks, 1999