Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Handwritten character recognition using convolutional neural network

To cite this article: I Khandokar *et al* 2021 *J. Phys.: Conf. Ser.* **1918** 042152

Handwritten character recognition using convolutional neural network

I Khandokar, Md M Hasan, F Ernawan, Md S Islam, M N Kabir*

Faculty of Computing, College of Computing and Applied Sciences, Universiti Malaysia Pahang, 26600 Pekan, Pahang Darul Makmur, Malaysia

*Corresponding author: nomanikabir@ump.edu.my

Abstract. Handwritten character recognition (HCR) is the detection of characters from images, documents and other sources and changes them in machine-readable shape for further processing. The accurate recognition of intricate-shaped compound handwritten characters is still a great challenge. Recent advances in convolutional neural network (CNN) have made great progress in HCR by learning discriminatory characteristics from large amounts of raw data. In this paper, CNN is implemented to recognize the characters from a test dataset. The main focus of this work is to investigate CNN capability to recognize the characters from the image dataset and the accuracy of recognition with training and testing. CNN recognizes the characters by considering the forms and contrasting the features that differentiate among characters. Our CNN implementation is experimented with the dataset NIST to obtain the accuracy of handwritten characters. Test result provides that an accuracy of 92.91% accuracy is obtained on 200 images with a training set of 1000 images from NIST.

1. Introduction

Handwritten character recognition (HCR) is a mechanism which enables to translate different types of documents into analysable, editable and searchable data. An ultimate aim of HCR is to emulate human reading capabilities in such a way that the machine can read, edit and interact with text as a human in short time. Identification of HCR has drawn great attention of numerous researchers over half a century, and many great achievements have been made in this field [1]. With current technical innovations, handwriting recognition carried out by vision sensors that can capture the position and movements of 3D fingers that write in the space has attracted remarkable interests. However, in the past years, a significant progress is made on HCR performance, but still now HCR is a challenging task due to the great diversity of handwriting style, the existence of many similar characters and large number of character categories [2].

Convolutional Neural Network (CNN) is a very well-known deep learning architecture motivated by the natural visual perception technique of human brain. Taking advantage of the recent exponential growth in the volume of annotated data and the rapid increases in the capabilities of the graphics processing units, the study on CNN has quickly arisen and obtained state-of-the-art performance on different tasks, e.g., image classification, text detection, pose estimation, object tracking, action detection, visual saliency detection, scene marking, speech and natural language processing. Although there are many variants of CNN architectures, their basic elements are very similar. It comprised of three types of layers, as convolutional, pooling, and fully-connected layers. It has brought great advances in the area of computer vision [3] and it has been introduced to the HCR to achieve excellent recognition performance [4]. Therefore, the researchers are using different CNN-based model for solving HCR problems [5].

Many researchers have been done research in HCR field but 100% accuracy cannot be achieved. Kato et al. in [2] presented a handwritten method for the identification of Chinese and Japanese characters. The approach used to recognize the characters from dataset ETL9B, where *directional element features* were extracted for each character and *asymmetric Mahalanobis distance* was used for fine classification. The accuracy reported in the paper is 99.42%. Chaudhuri and Adak [6] developed an approach for identification and deletion of strike-out texts in offline handwritten document images. They present a combined SVM classification and shortest path algorithm for identifying such texts and their obtained overall accuracy of the output was 94%, approximately. In [7], Gan. et al. used CNN algorithm to test Chinese handwritten character recognition in-air handwriting dataset IAHCC-UCAS2016. The authors used a unified algorithm to compress CNN algorithm considering channel pruning and network quantization. For channel pruning, they delete the least important feature maps and their related filters from the original network structure. The accuracy of their test is reported as 95.33%. Fardous and Afroge [8] and Das et al. [9] used CNN for recognizing isolated Bangla HCR. The two groups of researchers obtained their accuracy as 95.5% and 98.3%, respectively.

In this paper, our aim is to develop an efficient HCR system using CNN. To test the HCR system NIST database has been used. To assess the performance of CNN algorithm, we experimented with the dataset NIST and found the accuracy of handwritten characters. The handwritten characters in NIST are given as images. The images are split into training and testing sets. Training is carried out with various number of images; then testing is conducted to find the accuracy of the CNN.

This paper has five sections. Section 1 describes about the introduction, while Section 2 presents the background of the study. Section 3 demonstrates the methodology of the paper. Section 4 presents the test results. Finally, the conclusion is described in section 4.

2. Background Study

Recognition of handwritten characters has gained significant popularity in the field of pattern recognition and machine learning because of its use in various fields. Various techniques in handwriting recognition system have been proposed for character recognition. Sufficient studies and papers describe the techniques used to convert textual content from a paper document into readable machine form. Character recognition system may serve as a key factor in creating a paperless environment by digitization and processing of existing paper documents in the coming days [1].

CNN architecture mimics a human brain communication pattern of the neurons, boosted by visual cortex arrangement. Nerve cells just react to stimuli from visible spectra by observation and analysis. A range of such spectra fields overlaps to fill the whole display space. A CNN can capture the spatial dependencies inside an image to relate with the content of the image for recognition purpose. The CNN design provides a strong fit of the source image to find the distinguishing features in order to be classified. The weights, parameters and the biases involved with the transformations from original image to feature vector to know better about the nature of the image are found during the training stage [4]. Figure 1 shows the process of CNN that consists of the fundamental parts – convolution, pooling, fully-connected neural network - which is described as follows:

2.1. Convolution

Convolution is commonly used only for image manipulation, such as camera smoothing, feature enhancement and sharpening. Convolution not only used in CNNs, it is still a core feature in several other algorithms for machine learning. It is a process where a single matrix of numbers (known a kernel or filter), pass this over an image to reshape the image with the filter. Each selected pixel value of the image submatrix needs to be multiplied with the corresponding pixel value from the kernel and then sum up the result to formulate a single pixel value of the filtered image. This micro-scale operation is continued for the whole image [5]. Conventionally, convolution is able to capture all the low-level features e.g., borders, colour, gradient and orientation. The design also adjusts to a high-level usability with additional layers, offering a network with such a sound understanding of an image, similar to how humans do. The purpose of a convolution procedure is to remove the high-level features as edges of the source images. The method has an impact, i.e., the reshaped characteristic is reduced in dimensionality compared to the input [4].

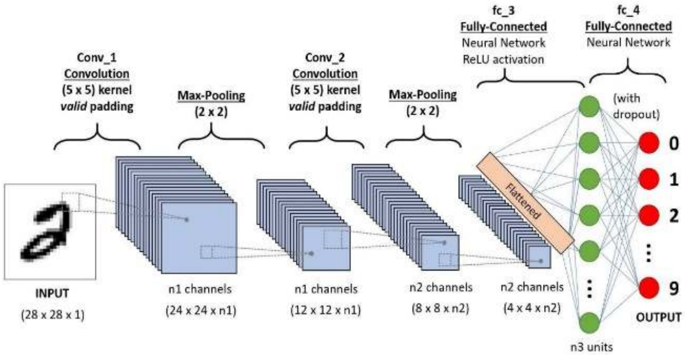

Figure 1. Process of Convolutional Neural Network

2.2. *Pooling*

A second key component of CNN is pooling. The purpose is to slowly raise the spatial scale of the image description to reduce the computational complexity within the network. In each application, pooling can be of different types - maximum pool, minimum pool, average pool and adaptive pool. Maximum pool (Max-Pooling) is a popular approach for CNN method. It reduces the computing power by reducing the dimensionality by keeping the dominant features that are rotational and translational variants that preserve the model's efficient training cycle. Max-Pooling gives maximum value from its image part covered within the kernel. But on the other side, average pooling provides better noise suppression. It discards the noisy activations and also conducts de-noise along with reduction of the dimensionality [4].

2.3. *Fully-connected neural network*

Fully connected neural network are used at the last part of CNN. They are simply artificial neural network which are fully connected. Weights which are involved with the network are computed during the training session. Fully connected neural network receives the end result of convolution/pooling operation and computes the most-matched label that represents the image. This part makes the connection of the image feature vector with the class of the image. The outputs from the convolution/pooling operation are multiplied by the weights associated with the network connection path. The result is then passed through an activation function [5].

3. Methods

HCR have been some of the long studied and significant topics of machine learning and computer vision [1]. We now discuss our methodology how the characters are recognized from texts in image format in a document. The proposed model illustrated by Figure 2 has four stages for the classification and detection purpose: Pre-processing, Segmentation & Clipping, Feature extraction and Classification & Recognition.

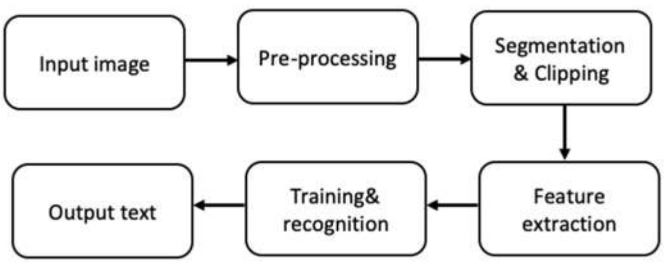

Figure 2. Proposed system architecture

3.1. Pre-Processing
Pre-processing takes input image is to perform cleaning tasks. It effectively enhances the image by noise removal. Furthermore, images may be required to be in greyscale or binary formats which are done in this stage [10].

3.2. Segmentation
After the input images are pre-processed, individual characters are separated using a segmentation technique. These characters are then stored into a sequence of images. Then boarders in each character image are eliminated if the boarder is available. Next, individual characters are scaled to specific size [10].

3.3. Feature Extraction
Feature extraction is made on the segmented characters. In our case, the features are extracted using CNN with ReLU activation function as shown in Figure 1. CNN works on each character image to form a matrix of reduced size using convolution and pooling. Finally, the reduced matrix is compacted to a vector form using the ReLU function. This vector is regarded as feature vector [5].

3.4. Classification and Recognition
The derived feature vector is used as individual input to formulate corresponding class. During the training phase, the parameters, biases, and weights are calculated. The calculated parameters, biases, and weights are used in the testing phase for classification and recognition purposes [11].

4. Results and Discussion
The CNN algorithm was implemented in MATLAB R2015a under Windows 7 operating system. The implemented program was run on intel core i7-2640 CPU with 4GB RAM. The test results are given in Table 1. The table provides three columns – in which the first column provides number of training images; the second column presents the number of testing sets and the last column lists the accuracy obtained by the CNN method for correctly classified images.

It can be noticed from the experiment that the average accuracy increases with higher number of training images, since the higher number of images in the training produces more accurate information on the training parameters, which subsequently improves the accuracy in classification during testing phase. One checks that the accuracy obtained from 200 training images as 65.32% is improved gradually with increasing training images. The accuracy reaches to 92.91% with the 1000 training images. Thus, further increment of training images will continue to enhance the accuracy towards to certain limit – which cannot be exceeded due to numerical errors, and the constraints on the CNN capability of image differentiability for labels.

Table 1. Test Results for Handwriting Character Recognition with NIST Dataset.

No. of Training Images	No. of Testing Images	Average Accuracy (%)
200	200	65.32%
300	200	74.43%
500	200	80.84%
600	200	85.21%
800	200	87.65%
1000	200	92.91%

5. Conclusion
In this research, a deep learning technique CNN is implemented for handwritten character recognition. The main focus of this work is to investigate CNN capability to recognize the characters from NIST dataset with a high degree of accuracy. It was found that the accuracy obtained from 200 training images

as 65.32% is improved gradually with increasing training images. The accuracy reaches to 92.91% with the 1000 training images. Thus, further increment of training images will continue to enhance the accuracy towards to certain limit.

Acknowledgments
This work was supported by Fundamental Research Grant Scheme (FRGS), Ministry of Education, Malaysia under Grant No. FRGS/1/2018/ICT04/UMP/02/2 and RDU190117

References

[1] Plamondon R and Srihari S N 2000 *IEEE Trans. on Patt. Anal. and Mach. Intelligence* **22** 63
[2] Kato N, Suzuki M, Omachi S I, Aso H and Nemoto Y 1999 *IEEE Trans. on Patt. Anal. and Mach. Intelligence* **21** 258
[3] Islam M M, Tayan O, Islam M R, Islam M S, Nooruddin S, Kabir M N and Islam M R 2020 *IEEE Access* **8** 166117
[4] Gu J, Wang Z, Kuen J, Ma L, Shahroudy A, Shuai B, Liu T, Wang X, Wang G, Cai J and Chen T 2018 *Patt. Recognition* **77** 354
[5] Ptucha R, Such F P, Pillai S, Brockler F, Singh V and Hutkowski P 2019 *Patt. Recognition* **88** 604
[6] Chaudhuri B B and Adak C 2017 *Patt. Recognition* **61** 282
[7] Gan J, Wang W and Lu K 2020 *Patt. Recog. Letters* **129** 190
[8] Fardous A and Afroge S 2019 *Proc. Int. Conf. on Elect. Comp. and Comm. Engineering* **4** 1
[9] Das D, Nayak D R, Dash R and Majhi B 2020 *Multi. Tools and Applications* **79** 33023
[10] Khoon L L and Chuin L S 2016 *Int. J. of Soft. Eng. and Comp. Systems* **2** 10
[11] Sallam A A, Kabir M N, Alginahi Y M, Jamal A and Esmeel T K *2020 Proc. Int. Colloq. on Sign. Proc. & Its Applications* **3** 255