2018 18th International Conference on Control, Automation and Systems (ICCAS 2018)
Oct. 17~20, 2018; YongPyong Resort, PyeongChang, GangWon, Korea

Number Recognition of Parts Book Schematics using Convolutional Recurrent Neural Network

Erdal Genc, Hee Ran Shin, Jang Sik Park

Department of Electrical and Electronic Engineering, Kyunsung University,
48434. 309, Suyeong-ro, Nam-gu, Busan, Republic of Korea(erdal.genc09@gmail.com, heeran3646@gmail.com, jsipark@ks.ac.kr)

Abstract: OCR (Optical Character Recognition) has been becoming a vital method to recognize digits, letters, symbols and so on. The main idea is basically the conversion of data files which consists of handwritten or machine-written digits or characters into a type to let the machine make edits and read. This way, it lets computers read articles or books. They can also read images and make the conversion to a text file by using OCR. There are two important benefits of OCR. First, is the enhancement of the device to operate more productively even if the number of employees is decreased. Secondly, is the increase in the efficiency of the storage. This paper compares two state-of-the-art OCR algorithms in a simulated environment by using modified dataset. Simulation results are shown in part 4.

Keywords: OCR, MLP, Deep Learning, Computer Vision, RNN, LSTM, Image Processing

1. INTRODUCTION

OCR is being used for years to detect or recognize the digits or characters from a schematic, image etc. If it is desired to use those characters or numbers in the image, it can be obtained by using OCR easily. These numbers and characters may be vital for various cases. They might be used as ranking, layout, organization or even more effective uses.

OCR can be used in many different areas such as banks, education, publishing articles, the entrance of data for business documents (e.g. check, passport, invoice, bank statement, and receipt), traffic (plate number recognition).

In [1], authors use Tesseract(an OCR algorithm developed by Hewlett-Packard Laboratories between 1985-1994, currently being developed by Google) to recognize the text from an image that may be an electronic conversion of images, handwritten or printed article, a scanned paper or a photo taken by a camera. After recognition of the text by OCR, the obtained data might be stored in any kind of file. To search for the string stored in the file, Boyer-Moore string search algorithm [2] is used.

In [3] an OCR system is developed by combining CNN(Convolutional Neural Network) and 1-D LSTMs(Long Short-Term Memory). Feature extraction is made by CNN, whereas, sequence modeling is done by LSTM. The results are obtained by using English and Arabic handwritten data, machine print data(English). They believe that their approach is more straightforward than the current 2-D LSTM model.

In [4], in order to analyze the complex gene-network, identification of articles discussing cis-regulatory elements and modules are important. One of their approach for identification is grayscale imaging is not accurate enough. Therefore, OCR has a vital role in this situation. They acquire the data from the images by using OCR.

In this paper, we present a benchmark of the accuracy of two different algorithms by using our own dataset. As simulation results, CRNN(Convolutional Recurrent Neural Network) shows better performance than conventional one.

2. OPTICAL CHARACTER RECOGNITION

2.1 Multilayer Perceptron

Multilayer perceptron (MLP) is the type of feedforward only artificial neural network that is composed of at least 3 layers of nodes: an input layer, hidden layer, and an output layer. Backpropagation [5] is a technique of supervised learning, is used in order to train an MLP model. What distinguishes an MLP from a linear perceptron is that it can recognize non-linearity in the data. They are considered as original neural networks since they include a hidden layer.

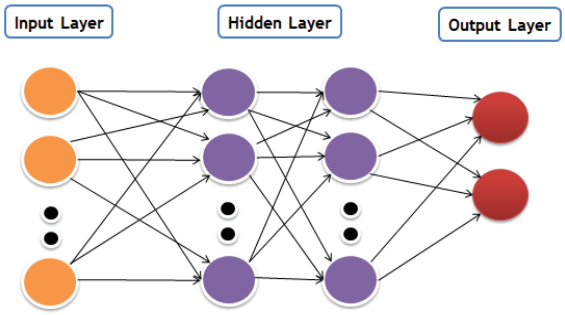

Figure 1. Multilayer Perceptron Architecture.

2.2 OCR

OCR is a process of recognition of text characters which are printed out or written by a computer. These characters might be involved in handwritten or printed text, scanned document, etc. It is commercially used in many different areas such as banking, data recording, plagiarism. During early processes of OCR, only one font was recognizable. Due to advancements in

machine learning and image processing fields, recognizable image formats and image fonts have skyrocketed.

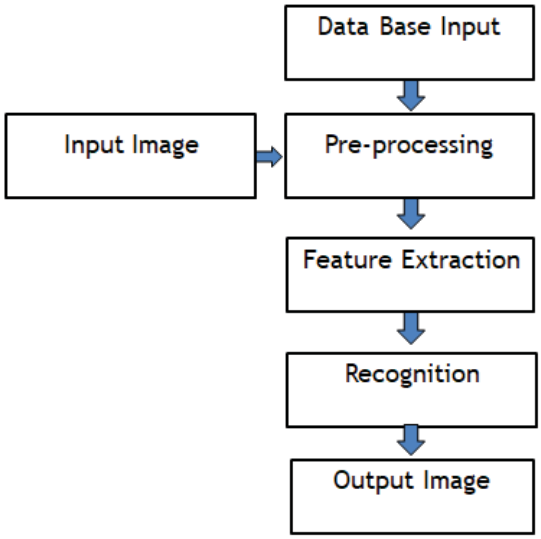

Figure 2. OCR process diagram.

A modern example of OCR MODI which is developed by Microsoft is currently considered as state-of-the-art. Even though the recognition rate is among the highest with 70-80%, this result is still not decent enough. Nowadays with the increasing performance rate of deep learning applications, authors propose a deep learning approach to the OCR low recognition problem.

3. OCR FOR PARTS BOOK USING CRNN

3.1 CRNN

A CRNN is an amalgamation of convolutional and recurrent neural networks that are mainly used on image recognition tasks. CNNs made a breakthrough for extraction independent of spatial variations, whereas, RNNs, which also includes LSTMs, make it possible to work within time domain due to their architecture.

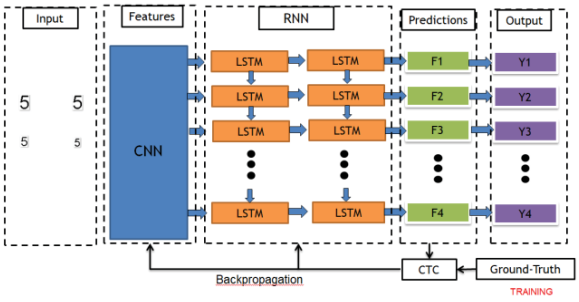

Figure 3.Architecture of CRNN.

3.2 OCR for Parts Book

A parts book is a guidebook which gives the information about the certain machinery parts. In Figure 5. Numbers represented in black color are original colors. Red colors are the OCR predicted numbers. Problems may arise from the parts of the image. For instance, certain drawings may be recognized as numbers, as well as guiding lines can be recognized as digit 1.

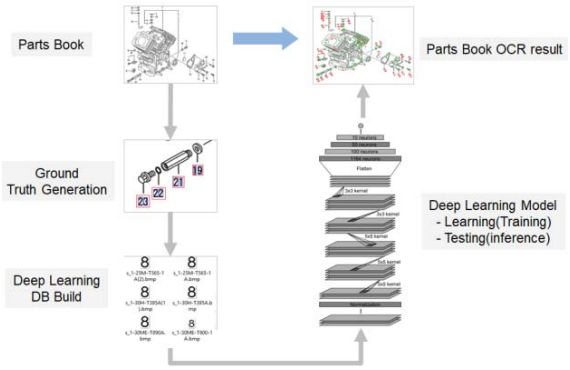

Figure 4. Block Diagram of OCR process for parts book.

4. SIMULATION RESULTS AND CONSIDERATIONS

4.1 Environments of OCR

The developed system was realized using Ubuntu 16.04 OS with Caffe framework[6] written in Python 2.7. Computer specifications are: 64GB RAM, 500GB Samsung SSD, Intel Xeon E5-2650 @ 2.30 GHz x 40, Nvidia Quadro M5000-SSE2.

Initially, a digital version of a parts book is obtained. Through the means of number extraction, ground truth images are generated. Using extracted images, a database of digits(0,1,…,9) is created that consists of 55,000 images.

First, an MLP using OpenCV artificial neural network architecture as a reference is trained using the dataset. The created dataset is used to train the network. For testing purposes, 10,000 images have been used. Training process took 1 hour and 33 minutes.

Erstwhile, using the dataset, the proposed architecture is trained with the following specifications; learning rate is 0.0001, maximum iteration is 50,000, weight decay is 0.0005. Training process took 5 hours 15 minutes.

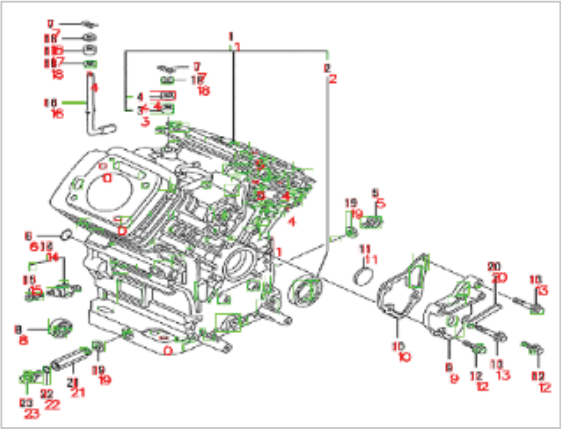

Figure 5. An Example of schematic recognition by OCR.

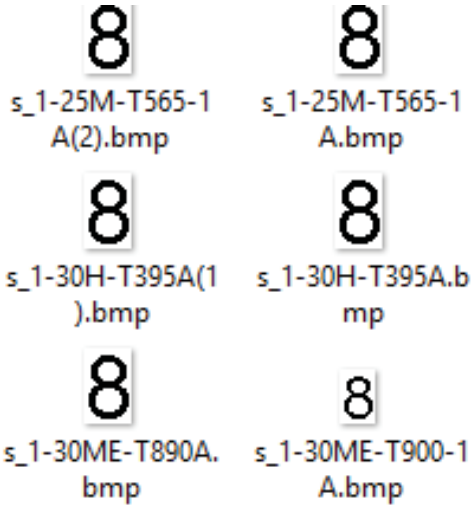

Figure 6. Example of the image dataset.

4.2 Simulation results

Even though MLP has been trained in a shorter period of time, CRNN deep learning method has performed more accurately during the testing process. Results are shown below in Table 1.

Table 1. Testing results comparison

	MLP	CRNN
Accuracy	91.91%	99.46%

From the table, it can be seen that CRNN heavily outperformed MLP based approach.

5. CONCLUSIONS

It can be clearly understood that combination of two deep learning architectures has better performance than the conventional method.

For future work, instead of digit extraction from the given images, the whole image might be presented as training dataset. That way, ground truth database creation time might be shortened while the application areas might be broadened.

ACKNOWLEDGMENTS

REFERENCES

[1] P.A. Wankhede, and S.W. Mohod, "A different image content-based retrievals using OCR techniques", International Conference of Electronics, Communication and Aerospace Technology, April 2017.

[2] R.S. Boyer, and J.S. Moore, "A Fast String Searching Algorithm", Association for Computing Machinery, vol. 20, no. 10, pp. 762-772, October 1977.

[3] S. Rawls, H. Cao, S. Kumar, and P. Natarajan, "Combining Convolutional Neural Networks and LSTMs for Segmentation-Free OCR", IAPR International Conference on Document Analysis and Recognition, vol. 10, no. 1109, pp. 155-160, November 2017.

[4] H. Shatkay, R. Narayanaswamy, S. S. Nagaral, N. Harrington, R. Mv, G. Somanath, R. Tarpine, K. Schutter, T. Johnstone, D. Blostein, S. Istrail, and C. Kambhamettu, "OCR-Based Image Features for Biomedical Image and Article Classification: Identifying Documents Relevant to Genomic Cis-Regulatory Elements", in Proceedings of the ACM Conference on Bioinformatics, Computational Biology and Biomedicine, pp. 98-104, October 2012.

[5] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "

[6] ImageNet classification with deep convolutional neural networks" in Proceedings of the 25th International Conference on Neural Information Processing Systems, vol. 1, pp. 1097-1105, December 2012.

[7] Y. Jia, E. Shelhamer, J. Donahue, S. Karayev, J Long, R. Girshick, S. Guadarrama, and T. Darrell, "Caffe, Convolutional Architecture for Fast Feature Embedding", In Proceedings of the 22nd ACM international conference on Multimedia, pp. 675-678, November 03-07 2014.