An End-to-End Trainable Thai OCR System Using Deep Recurrent Neural Network

รัฐศาสตร์ เฮงประเสริฐ* และ สุรเดช อิณทกรณ์
สาขาเทคโนโลยีสารสนเทศ คณะศิลปศาสตร์และวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์

Ratthasart Hengprasert* and Suradej Intagorn
major Information Technology, faculty of Liberal Arts and Science, Kasetsart University

Abstract

In this paper, we present an end-to-end trainable model to recognize a Thai word from an image. Compared with other previous Thai OCR system, our system has distinctive features that can handle arbitrary length of Thai word without character segmentation and high level visual features are learned from data. The neural network model is composed of 2 main modules: Convolutional Layer and Recurrent Neural Network (LSTM).

Index Terms- *Neural Network; Optical Character Recognition; Convolution Neural Network; Recurrent Neural Network; Image Processing*

Introduction

The recognition process of Thai characters are difficult because two main reasons. First, visual representation of Thai character is sophisticated. Some of them are very similar in visual representation, for example (ค,ศ), (ข,ฃ),(ช,ฆ) and (ฎ,ฏ). Second, there are four levels of characters in each line depending on the type of characters, for example, consonants or vowels. Figure 1 shows level in a Thai sentence

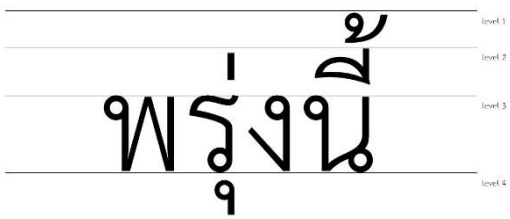

Figure 1. level in a Thai sentence [1].

The level in a Thai sentence makes the character segmentation module in the preprocessing step in Thai OCR more complicated than some other languages.

According to Tanprasert, preprocessing step is the one of most crucial step of the Thai

* ผู้ประสานงาน (Corresponding Author)
e-mail: ratthasat241998@gmail.com

OCR system.Noise cleaning and pixel binarization are applied in order to remove unwanted pixel from the images which results in more accurate character segmentation.

In contrast, our OCR system do no need the character segmentation module. Feature extraction are also learned from data via convolution layers.

ReLated Work

Tanprasert et al used an artificial neural network to recognize Thai sentences from images. They present challenges and solution in Thai OCR system. Several preprocessing steps are proposed including character segmentation step. They further improved their works for recognizing images that contain both Thai and English character by another preprocessing step called Kohonen self-organization. (Tanprasert, 1997)

Shi et al presented a novel end-to-end method to recognize a text in images. Their system do not need the character segmentation module. Instead, they extract high level visual feature by using convolution layer and then treats the problems as a sequence of high level features from convolution layer to the deep recurrent layer to predict the text. (Shi Baoguang, 2016)

Methodology

Data Set

There are 8,142,880 images in our dataset from 10 fonts which are angsab, Charm-Regular, THSarabunNew Bold, SOV_wayo, Kanit-Regular, TH Mali Grade6 Bold, TH Charm of AU, TH Srisakdi, TH Krub, TH Charmonman. There are 7735736 in the training set and 407144 in the test set. Most of the words in our corpus are from Thai Wikipedia. The images are assumed to contain only a single word. We can easily extend this work from word level to line level by increasing the size of the network. However, training the larger network requires more powerful computational resources (such as high-end GPU). Due to limitation of our hardware, we limited our works to the word level which can still demonstrate the ability to handle the character segmentation problem.

Preprocessing

In previous Thai OCR systems, several preprocessing steps are required. However, in this work, there are only thre simple preprocessing steps. First, we converted the images to gray scale images. Second, we resize the images from an arbitrary size to width = 128 pixels and height = 32 pixels. Third, we normalize the value from 0-255 to standard Gaussian of pixel values with a mean of 0.0 and a standard deviation of 1.0. Figure 2 shows the preprocessing step in this work.

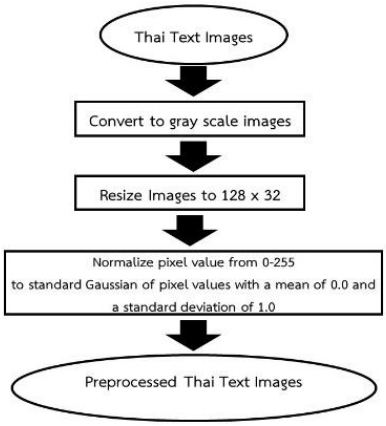

Figure 2. The preprocessing diagram

The Convolutional Recurrent Network

Our model and codes are based on the work from Shi Baoguang and K. Simonyan. The idea of this model is to extract high level features by using Convolutional layers which is called feature maps. Then these feature maps are slice from into 32 time steps to feed into the bidirectional LSTM (Recurrent Layer). Figure 5 shows the output from Convolutional layers. The output can be interpreted as high level visual feature corresponding to position in the images. The size of these features are smaller than the original size of the input image because of max pooling layer in the convolutional layers. The size of each feature is 32x1 and there are 256 features (the input image size is 128x32). Then these features are sliced into 32 time steps. Please note that each time step is just a position from left to right of the image, for example, time step = 1 is the features of the 4 leftmost column of the image. Time step is just a naming convention to the input to the recurrent neural network. The authors used the recurrent network to deal with the varying output size problem, for example, outputs size of จาน and ล้มลุก are different. จาน has 3 characters and ล้มลุก has characters. The maximum output length in this work is 32. It can be increased by designing the larger network. Figure 3 shows the training phase of the model. The prediction of the model is compared with the ground truth and propagate the errors back to each layer in each epoch until errors cannot be reduced on the training set. (Shi Baoguang, 2016) (K. Simonyan, 2014)

Figure 4 shows the prediction phase of the model. Please note that the recurrent neural network can deal with varying size of output by output the blank output at some time steps. We can see that output of the second and fourth time step are blank output.

Figure 5 shows the output of convolutional layer which the size of each feature is 32x1 and there are 256 features. Figure 6 shows the input of recurrent layer. The 32x1 feature map are sliced into 32 time steps to be compatible with the recurrent layer.

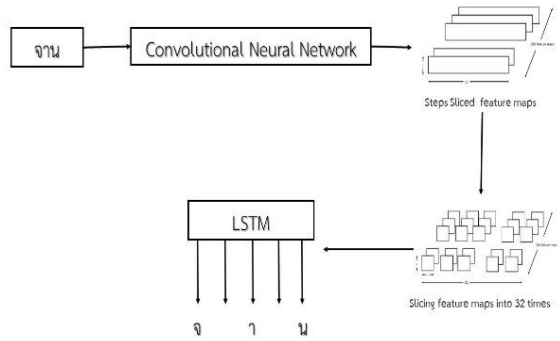

Figure 3. The network architecture

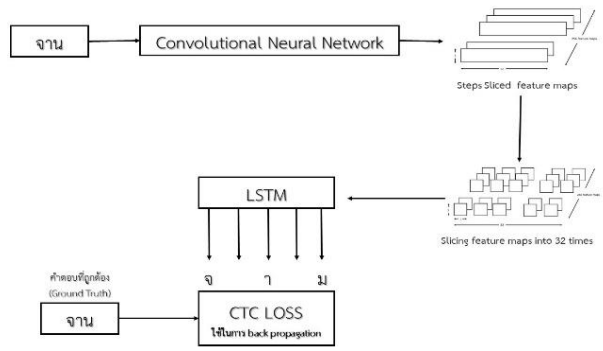

Figure 4. the prediction phase of the model.

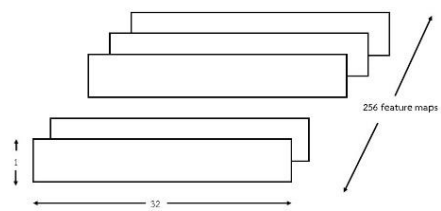

Figure 5. The output of convolutional layer which the size of each feature is 32x1 and there are 256 features.

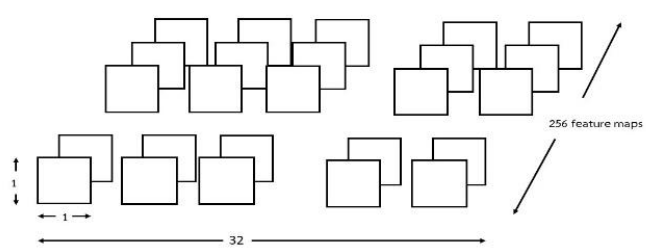

Figure 6. The input of recurrent layer. The 32x1 feature map are sliced into 32 time steps

Implementation

Python is used as the main programming language. The standard data science libraries are used such as numpy, scikit-learn and etc. The network is implemented by tensor flow. Flask is used to create the micro services. HTML, Javascript, CSS, JQuery and Jinja2 are used to create web application front end.

Results

Evaluation Metric

In this work, we measure the two level of accuracies, the word level and character level. The word level is quite simple if the prediction is same as the ground truth, we count it as correct. The character accuracy is measured by Levenshtein distance

Word and character accuracy results

The table below shows the accuracy of our system.

Table 1. Word and character accuracies

Metric	Value
Word Accuracy	92.23
Character Accuracy	98.01

Demo

In this work, we implemented our demo by using flask, HTML, Javascript, CSS, JQuery and Jinja2. The figures below shows the GUI demo and inputs from 10 fonts.

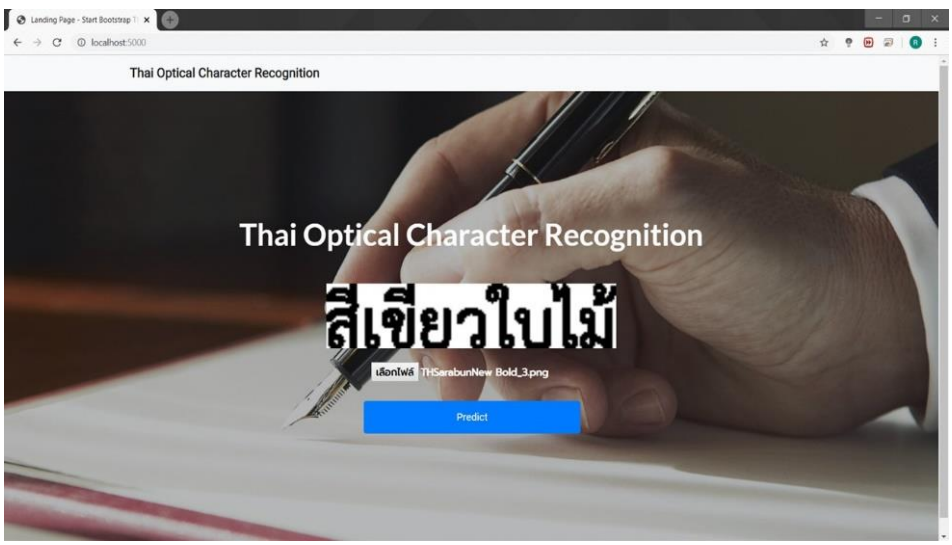

Figure 7. The Demo

Conclusion

In this work, we present an end-to-end system that can recognize the text in images. There are no character segmentation in the preprocessing step in this system. The model consists of convolutional and recurrent layers. We achieved the word accuracy at 92% and character level at 98%. We also implement the web application as a demo of our system.

คณะผู้เขียน/ ผู้เขียน

นายรัฐศาสตร์ เฮงประเสริฐ

สาขาเทคโนโลยีสารสนเทศ คณะศิลปศาสตร์และวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ เลขที่ 1 หมู่ 6 ต.กำแพงแสน อ.กำแพงแสน จ.นครปฐม 73140

e-mail : ratthasat241998@gmail.com

อ.ดร.สุรเดช อินทกรณ์

สาขาเทคโนโลยีสารสนเทศ คณะศิลปศาสตร์และวิทยาศาสตร์ มหาวิทยาลัยเกษตรศาสตร์ เลขที่ 1 หมู่ 6 ต.กำแพงแสน อ.กำแพงแสน จ.นครปฐม 73140

e-mail : suradej.i@ku.th