VNU Journal of Science: Mathematics – Physics, Vol. 36, No. 1 (2020) 64-70

VNU Journal of Science: Mathematics - Physics

Journal homepage: https://js.vnu.edu.vn/MaP

Original Article

An Efficient Method for Automatic Recognizing Text Fields on Identification Card

Nguyen Thi Thanh Tan[1*], Le Hong Lam[2], Nguyen Ha Nam[3]

[1]*Faculty of Information Technology, Electric Power University, Hanoi, Vietnam*

[2]*VNU Institute of Information Technology, 144 Xuan Thuy, Cau Giay, Hanoi, Vietnam*

Received 15 January 2020
Revised 21 February 2020; Accepted 26 February 2020

Abstract: The problem of optical character and handwriting recognition has been interested by researchers in long time ago. It has obtained great results in theory as well as practical applications. However, the accuracy of identification is still limited, especially in the case of low-quality input images. In this article, we propose an efficient method to recognize information fields for identification in ID card using Convolutional Neural Network (CNN) and Long Short-Term Memory networks (LSTM). The proposed method was trained in a large, various quality dataset including over three thousands ID card image samples. The implementation achieved better results compare to previous studies with the precision, recall and f-measure from over 95 up to over 99% out of all information fields to be recognized.

Keywords: HPC, academic, industrial applications, calculations.

1. Introduction

Identification (ID) Card is a personal card, providing basic information of citizen such as full name, date of birth, place of origin, place of permanent residence, nationality, religion, date and place of issue. In almost daily business, those information are required and usually extracted manually. It is not efficient process because we need a lot of time to input data one by one. Therefore, we need a method that processes automatically known as Optical Character Recognition (OCR) [1],[2].

*Corresponding author.
Email address: thanhtan.nt@gmail.com

https//doi.org/ 10.25073/2588-1124/vnumap.4456

A Vietnamese ID card usually contains text fields with different font styles and size. In many cases, the characters and also the other parts like rows, the seal, the signature was not well printed which cause the inaccurate information, like the overlap of characters [3]-[5]. In addition, by the time, the card is normally faded and blurred. In the literature, there are already existing works to improve the accuracy of ID card reading by different techniques before the recognition of optical characters. But for the Vietnamese ID Card, especially with the old form, it still lacks an efficient method to improve the quality of input data, reduce noise or time for the recognition task. In this paper, we propose an efficient method to recognize information fields for identification in ID card using Convolutional Neural Network (CNN) and Long Short-Term Memory networks (LSTM) [6],[7].

The paper is organized as follows: Section 2 presents our proposed method for automatic recognition of all personal information on the Identification Card. Section 3 provides the experimental evaluation; Section 4 is our conclusion and further work.

2. Computational methods

2.1. Details

We propose an adaptive method, as illustrated in the Fig.1 for automatic recognizing text fields from the Vietnamese ID, includes [8]:

- ✓ Image pre-proceeding.
- ✓ Analysis of table structure.
- ✓ Text zones detection
- ✓ Text lines segmentations.
- ✓ Text line recognition.

Image pre-proceeding: enhancing the quality of input data: As mentioned above, ID cards can be stained, moldy, crumpled and worn out over time [9],[10]. Therefore, improving and enhancing the quality of input image is necessary and important. Pre-processing was done in both front and back side of the card. It includes basic steps: Convert the color image to the gray-scale one; align tilt, smooth and create the binary image. Detecting and separating the ID card number: For the front side, the important information we need is the ID card Number, so that with this side we firstly detect and separate the ID Card Number field. However, due to the same color among the ID card Number, wavy lines, the national emblem and sometimes clothes of ID card holder; therefore, firstly we highlight the ID card.

Analysis of table structure: For the back side, the ROI is a table that contains different information. The table is formed by horizontal and vertical lines but those lines is usually blurred or dashed. Moreover, while stamping/printing and finger-print, the characters or the fingerprints may overlap with lines which makes it difficult to detect the table structure. Therefore, to determine table structure, the horizontal and vertical lines should be clearly defined. Since they have the same characteristic, we apply also a same algorithm to define them.

Text zones analysis and detection: The detection and segmentation of text zones is applied on the binary image block after separating national emblem, portrait, headings and ID card Number in the front side, or the text image defined from the table in the back side.

Text lines segmentation and normalization: The main purpose of this processing step is to segment blocks of text into separate text lines before recognizing them. For identification cards, text lines come in many different sizes, yet the relative position and scale of characters is an important feature for distinguishing characters in Vietnamese script and a variety of other scripts. Our text lines detection

method was proposed in [11]. The main idea of the method is to group together characters with same properties by walking through the document to form a text-line. Text line normalization plays an important role in applying CNN-LSTM networks to OCR in the next step. The normalization procedure for text line images is based on a dictionary composed of connected component shapes and associated baseline and x-height information. This dictionary is pre-computed based on a large sample of text lines with baselines and x-heights derived from alignments of the text line images with textual ground-truth, together with information about the relative position of Vietnamese characters to the baseline and x-height.

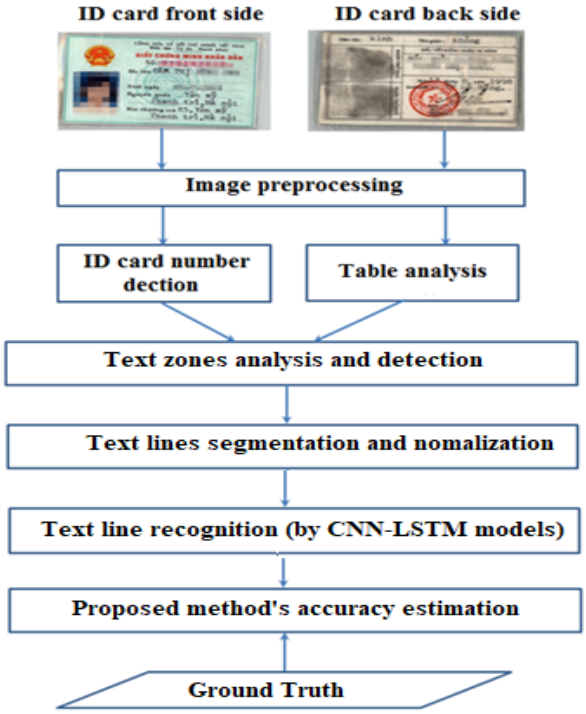

Figure 1. Automatic recognizing information fields on Identification Card.

When the baseline and x-height of a new text line need to be determined, the connected components are extracted from that text line and the associated probability densities for the baseline and x-height locations are retrieved. These densities are then mapped and locally averaged across the entire line, resulting in a probability map for the baseline and x-height lines across the entire text line. The resulting densities are then fitted with curves and used as the baseline and x-height for line size normalization. In line size normalization, the (possibly curved) baseline and x-height lines are mapped to two straight lines in a fixed size output text line image, with the pixels in between them rescaled using spline interpolation.

Text line recognition: The input of this step is the image of the text lines detected in the previous step. For recognition, we use the CNN-LSTM model. Our approach is purely data-driven and can be adapted with minimal manual effort not only to Vietnamese but also to different languages and scripts. Feature extraction from text images is realized using convolutional layers. Using the extracted features, we analyze the ability to model local context with both recurrent and fully convolutional sequence-to-sequence architectures. The alignment of the extracted features with ground-truth transcripts is realized via a CTC layer. This LSTM model will be presented more detail in the follow section.

2.2. Text line recognition model

The architecture of hybrid CNN-LSTM model for text line recognition is depicted in Fig. 2.

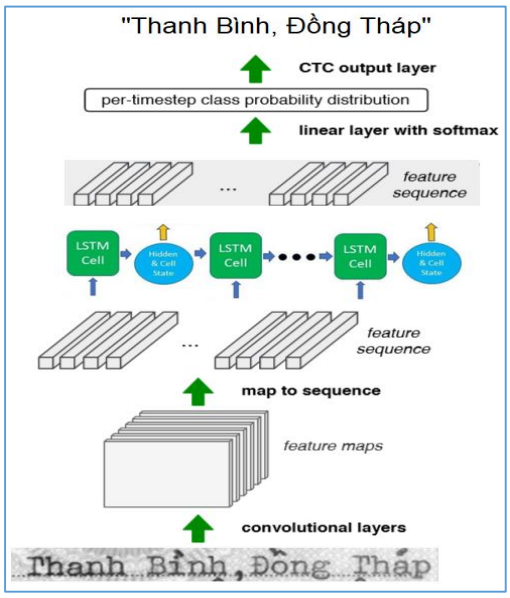

Figure 2: Recognizing information fields on ID card

The bottom part consists of convolutional layers that extract high-level features of an image. Activation maps obtained by the last convolutional layer are transformed into a feature sequence with the map to sequence operation. Specifically, 3D maps are sliced along their width dimension into 2D maps and then each map is flattened into a vector. The resulting feature sequence is fed to a bidirectional recurrent neural network with 256 hidden units in both directions. The output sequences from both layers are concatenated and fed to a linear layer with soft ax activation function to produce per-time step probability distribution over the set of available classes. The CTC output layer is employed to compute a loss between the network outputs and the ground truth transcriptions. During inference, CTC loss computation is replaced by greedy CTC decoding [12].

Table 1.Structure of CNN-LSTM model

Layers	**Output volume size**
Conv2d (3×3,64; stride: 1×1	32×W×64
Max pooling (2×2; stride: 2×2)	16×W/2×64
Conv2d (3×3,128; stride: 1×1)	16×W/2×128
Max pooling (2×2; stride: 2×2)	8×W/4×128
Map to sequence	W/4×1024
Dropout (50%)	–
Bidirectional LSTM (units: 2×256)	W/4×512
Dropout (50%)	–
Linear mapping (units: num classes)	W/4×*num classes*
CTC output layer	*Output sequence length*

Table 1 show more detail of the CNN-LSTM model. The model was trained via minibatch stochastic gradient descent using the Adaptive Moment Estimation (Adam) optimization method. The learning rate is decayed by a factor of 0.99 every 10000 iterations and has an initial value of 0.0001 for the model. Batch normalization [13] is applied after every convolutional block to speed up the training. The model was trained for approximately 300 epochs.

3. Results and discussion

To train and evaluate our ID Card OCR system we prepared several datasets, consisting of both real and synthetic documents. This section describes each in detail, as well as the preparation of training, validation, and test samples, data augmentation techniques, and the geometric normalization procedure.

The experiments are carried out using 3256 ID Card images, in which there are 1628 front-side-images and 1628 back-side images of ID cards. ID cards were collected from many provinces, in various qualities, font sizes, printing style and scanned at resolutions of 200dpi, 300dpi and 400dpi. Details of the experiments data are given in Table 1. The text lines were normalized to a height of 32 in preprocessing step.

Table 2. Experiment datasheet

Information fields	#Text lines	# Characters
ID card No.	1628	17908
Full name	1628	35216
Date of birth	1628	16280
Place of origin	2156	75460
Ethnic group	1628	6712
Religion	1628	1057
Date of issue	1628	37444
Place of issue	1628	19536
Total:	13552	209613

In order to evaluation of the result, we based on Precision, Recall and F-measure **Error! Reference source not found.**, which are calculated as following:

Precision **=** (Number of correct text line Recognized)/ /[Number of correct text line recognized + Number of incorrect text line recognized]

Recall = (Number of correct text line recognized)/ [Number of correct text line recognized + Number of unrecognizable text lines]

F-Measure =(2*Precision*Recall)/ (Precision+Recall)

The experimental results on the real datasheet are described more detail in Table 3.

We compare the text line recognition error rates of our system with two established commercial OCR products: ABBYY FineReader 11 [14] and with a popular open-source OCR library – Tesseract versions 4 [15] and Ocropus [16]. Recognition is performed at the text line level. The ground truth layout structure is used to crop samples from document images. The experiment results are showed on Figure 3.

Table 3. Accuracy of text line recognition

Information field	#Text lines	Precision (%)	Recall (%)	F-Measure (%)
ID card No.	1628	98.8	98.62	97.7
Full name	1628	97.9	97.40	97.53
Date of birth	1628	98.57	98.13	98.21
Place of origin	2156	96.09	95.60	95.86
Ethnic group	1628	99.24	99.02	99.11
Religion	1628	99.1	98.93	99.01
Date of issue	1628	96.53	96.08	96.21
Place of issue	1628	95.71	95.44	95.59

The results presented in this paper show that the CNN-LSTM model yields good OCR results for Vietnamese ID card recognition. Our benchmarks suggest that error rates for CNN-LSTM based OCR without a language model are considerably lower than those achieved by segmentation based approaches.

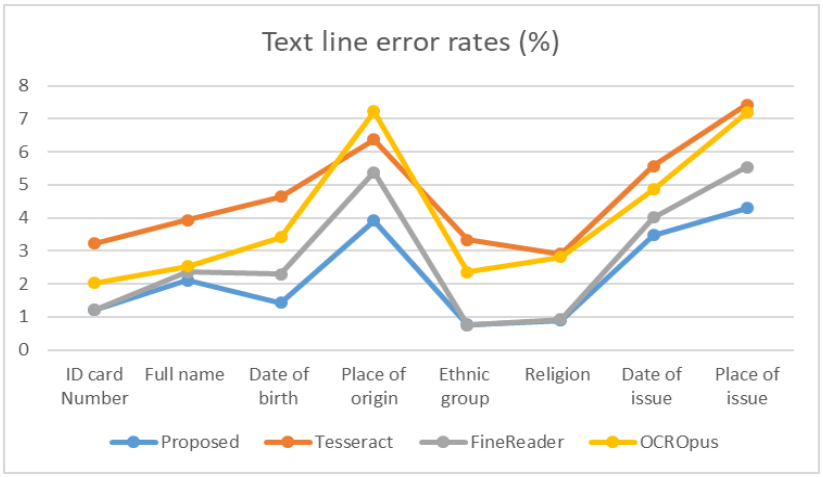

Figure 3. Compare text line error rates of systems.

A common and valid concern with OCR systems based on machine learning or neural network techniques is whether they will generalize successfully to new data.

We would ordinarily determine the error rate of a system by taking a data set, dividing it into training and test sets, train the system on the training set and evaluate on the test set.

There are several indications that LSTM-based approaches generalize much better to unseen samples than previous machine learning methods applied to OCR.

During LSTM training, we often observe very low error rates long before even one epoch of training has been completed, meaning that there has likely not been an opportunity to "overtrain". LSTM-based systems have been found to generalize well to novel data by other practitioners.

4. Conclusions

The article proposed a solution for recognition the text fields, which is suitable for identification automatic data input) of personal information on Vietnamese ID card. Based on its specific feature, the detection and segmentation are divided into two separated step for the back side and the front side. Ours recognition engine has built based on the CNN and multidimensional LSTM networks.

The implementation achieved better results compare to previous studies with the precision, recall and f-measure from over 95 up to over 99% out of all information fields to be recognized.

Acknowledgments

This work has been sponsored and funded by Ho Chi Minh City University of Food Industry under Contract No. 149/ HD-DCT.

References

[1] T.M. Breuel, A.U. Hasan, M.A. Azawi, F. Shafait, High-performance ocr for printed english and fraktur using lstm networks, Proc. 12th Int. Conf. on Document Analysis and Recognition (2013) 683 - 687.

[2] N.T.T. Tan, N.T. Khanh, A Method for Segmentation of Vietnamese Identification Card Text Fields, Advanced Computer Science and Applications, 10 (2019) 415-421.

[3] E. Sabir, S. Rawls, P. Natarajan, Implicit language model in lstmf or ocr, Proc. 14th IAPR Int. Conf. Document Analysis and Recognition, (2017) 27–31.

[4] M.R. Yousefi, M.R. Soheili, T.M. Breuel, D. Stricker, A comparison of 1d and 2d lstm architectures for the recognition of handwritten Arabic, Proc. of SPIE-IS&T Electronic Imaging, (2015), doi 10.1117/12.2075930.

[5] P. Lyu, M. Liao, C. Yao, W. Wu, X. Bai, Mask textspotter: An end to-end trainable neural network for spotting text with arbitrary shapes, Proc. European Conf. on Computer Vision, (2018) 1 - 16.

[6] K. He, X. Zhang, S. Ren, J. Sun, Deep residual learning for image recognition, Proc. IEEE Conf. on Computer Vision and Pattern Recognition, (2016) 770 - 778.

[7] M.R. Yousefi, M.R. Soheili, T.M. Breuel, D. Stricker, A comparison of 1d and 2d lstm architectures for the recognition of handwritten Arabic, Proc. of SPIE-IS&T Electronic Imaging, (2015), doi 10.1117/12.2075930.

[8] W. Satyawan, M.O. Pratama, R. Jannati, G. Muhammad, B. Fajar, H. Hamzah, R. Fikri, K. Kristian, Citizen Id Card Detection using Image Processing and Optical Character Recognition, IOP Conf. Series: Journal of Physics, (2019) 1 – 6, doi: 10.1088/1742-6596/1235/1/012049.

[9] T.M. Breuel, A.U. Hasan, M.A. Azawi, F. Shafait, High-performance ocr for printed english and fraktur using lstm networks, Proc. 12th Int. Conf. on Document Analysis and Recognition (2013) 683 - 687.

[10] R. Smith, Limits on the application of frequency-based language models to ocr, Proc. Int. Conf. Document Analysis and Recognition, (2011) 538–542.

[11] B. Shi, X. Bai, C. Yao, An end-to-end trainable neural network for image-based sequence recognition and its application to scene text recognition, IEEE Transactions on Pattern Analysis and Machine Intelligence 39 (2017) 2298–2304.

[12] S. Ioffe, C. Szegedy, Batch normalization: Accelerating deep network training by reducing internal covariate shift, Proc. of the 32nd Int.Conf. on Machine Learning, (2015) 448–456.

[13] D. Kingma, J.B. Adam, A method for stochastic optimization, Proc. ICLR Int.Conf. on Learning Representations, (2015) 1 - 15.

[14] ABBYY FineReader Engine for OCR. https://www.abbyy.com/en-eu/finereader, 2019 (accessed 05 October 2019).

[15] Tesseract Open Source OCR Engine (main repository). https://tesseract-ocr.github.io, 2019 (accessed 03 September 2019).

[16] Python-based tools for document analysis and OCR. https://github.com/tmbdev/ocropy, 2019 (accessed 25 September 2019).