An application of Vietnamese handwriting text recognition for information extraction from high school admission form

Pham The Bao, Le Tran Anh Dang, Nguyen Duy Tam, Nguyen Nhat Truong, Pham Cung Le Thien Vu, Trinh Tan Dat
Information Science Faculty, Sai Gon University, Ho Chi Minh City, Vietnam

Article Info

Article history:

Received Feb 22, 2022
Revised Aug 18, 2022
Accepted Sep 17, 2022

Keywords:

Attention
Convolutional neural networks
Handwriting
Long short-term memory
Recognition
Vietnamese

ABSTRACT

This paper presents an effective Vietnamese handwritten text recognition model by applying an improved convolutional recurrent neural networks (CRNNs) model to high school enrollment forms in Tay Ninh province, Vietnam. First, the proposed model extracts data areas containing text characters from forms. Then, we connect text boxes on the same row and divide the fields that containing text into three specific regions. Finally, we detect areas containing text characters for handwritten text recognition. We use word error rate (WER) to evaluate the recognition process and obtain a result of 0.3602. This result is one of the best solutions to the Vietnamese handwritten text recognition problem.

Corresponding Author:

Pham The Bao
Information Science Faculty, Sai Gon University
273 An Duong Vuong Street, Ward 3, District 5, Ho Chi Minh city, Vietnam
Email: trinhtandat@sgu.edu.vn, ptbao@sgu.edu.vn

1. INTRODUCTION

Every year in Tay Ninh province, the data entry of the entrance exam into the 10th grade of Tay Ninh high schools is done according to the typing process through the Department of Education and Training database interface of Tay Ninh. This province has ten schools that organize exams and admissions, so the annual data entry of student files across the entire province includes: 10835 files, the first phase is 4,000 files, the second phase is 6,835 files [1]. The data entry for students is all manually entered. From elementary schools to middle schools, from middle schools to high schools, and from high schools to the high school graduation exam, these data are re-entered every year. After being entered and stored, the data of each grade level is only used for the school years of that grade. When transferring files to another school level, these data are completely re-entered without inheritance. This problem costs the province's labor resources, time, and expense. Therefore, the creation of a system to support the digitization of candidates' registration forms is necessary to serve the entrance exam to high school in Tay Ninh province [2]–[4].

2. RELATED WORKS

Jaramillo *et al.* [5] presented the problem of processing offline handwritten text recognition handwriting text recognition (HTR) with reduced training data sets. Recent HTR solutions based on artificial neural networks show remarkable solutions in referenced databases. These deep neural networks include convolutional neural networks (CNNs) and long short-term memory (LSTM). In addition, connectionist temporal classification (CTC) is the key to avoiding character-level segmentation, greatly facilitating the labeling task. In 2018, Nguyen *et al.* [6] created an unconstrained Vietnamese online handwritten text database sampled from pen-based devices. The database stores handwritten text for paragraphs, lines, words, and characters, with the ground truth associated with every paragraph and line. We show detailed statistical analysis of handwritten text in this database and describe recognition experiments using several recent methods, including the bidirectional long short-term memory (Bi-LSTM) network. Overall, our database contains over 480,000 strokes from more than 380,000 characters, currently the largest database of handwritten documents online in Vietnam [7].

Nguyen *et al.* [8] mentioned convolutional recurrent neural networks (CRNNs) excel at scene text recognition. However, this model suffers from vanishing/exploding gradient problems when processing long text images, commonly found in scanned documents. This problem poses a significant challenge to overcome the goal of completely solving the optical character recognition (OCR) problem. Inspired by recently proposed memory-augmented neural networks (MANNs) for long-term sequential modeling, they introduced a new architecture called convolutional multi-way associative memory (CMAM) to address limitations of current CRNNs. Their architecture, which takes advantage of recent memory access mechanisms in MANNs, demonstrates superior performance over other CRNN counterparts in three real-world long-text OCR datasets. In addition, this paper reports new state-of-the-art IAM-OnDB results for both open and closed dataset settings. The system combines methods from sequence recognition with a new input encoding using Bézier curves. This combination leads to up to 10 times faster recognition than our previous system. Through a series of experiments, they determine the optimal configuration of their models and report the results of their setup on several additional public datasets**.** Additionally**,** in 2020, Carbune *et al.* described an online handwriting system that can support 102 languages using a deep neural network architecture. This new system has completely replaced our previous segment-and-decode-based system, reducing the error rate by 20-40% relative to most languages [9].

3. PROPOSED METHOD

3.1. Overview

Vietnamese handwriting recognition is much more complicated than print recognition because it varies widely depending on the writer, writing direction, speed and writing pressure. Although handwriting studies have made remarkable achievements, the recognition efficiency is not high compared to other recognition fields [10]–[13]. Therefore, this field of identification poses many potentials and is also a challenge for our research [14]. The article presents the method from normalizing the collected data, detecting the handwriting text container of the image and the model training process, the OCR Vietnamese handwriting recognition method using the CRNN model [15].

The general model for extracting and recognizing handwriting to extract information from the 10th-grade enrollment form in Tay Ninh province is shown in Figure 1. This model consists of 3 main parts: region extraction (Cropper), character extraction (Text detection), and string identifier (OCR). These are three problems that *need to be solved.*

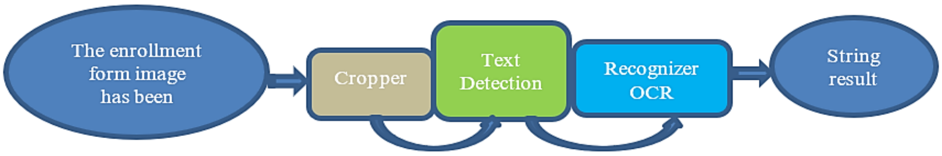
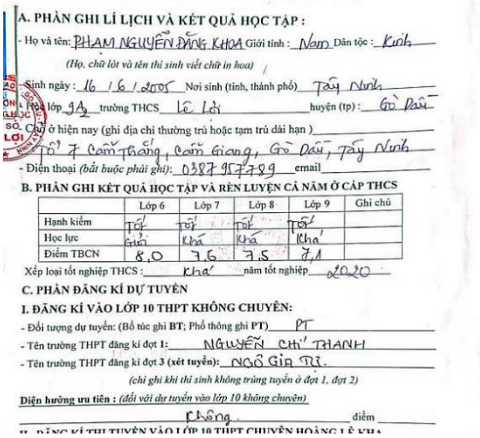


Figure 1. Overview model

3.2. Region extraction

We propose an algorithm to extract the data area from the scanned enrollment form image by only taking the critical information area, as shown in Figure 2, to remove unnecessary information. We then separate the information region into three regions called A, B, and C, shown in Figure 3(a) is region A, Figure 3(b) is region B, and Figure 3(c) is region C, to improve the performance of the extraction process information region when applying the efficient and accurate scene text (EAST) deep learning model [16] to the areas separated. Since the information in the form contains a scoreboard in region B, the table lines are noisy, causing difficulties in region extraction and character recognition. Therefore, the separation into three separate zones helps to improve the work efficiency. The enrollment form of Tay Ninh province is fixed, so we can separate these three regions based on heuristic thresholds.



A. PHẦN GHI LÍ LỊCH VÀ KẾT QUẢ HỌC TẬP :

- Họ và tên: PHAM NGUYỄN ĐĂNG KHOA Giới tính : Nam Dân tộc : Kinh

(Họ, chữ lót và tên thí sinh viết chữ in hoa)

- Sinh ngày : 16 / 6 / 2005 Nơi sinh (tỉnh, thành phố) Tây Ninh
- Học lớp 9A2 trường THCS Lê Lợi huyện (tp) : Gò Dầu
- Chỗ ở hiện nay (ghi địa chỉ thường trú hoặc tạm trú dài hạn) Tổ 7 Cẩm Thắng, Cẩm Giang, Gò Dầu, Tây Ninh
- Điện thoại *(bắt buộc phải ghi)*: 0387 957789 email

B. PHẦN GHI KẾT QUẢ HỌC TẬP VÀ RÈN LUYỆN CẢ NĂM Ở CẤP THCS

	Lớp 6	Lớp 7	Lớp 8	Lớp 9	Ghi chú
Hạnh kiểm	Tốt	Tốt	Tốt	Tốt	
Học lực	Giỏi	Khá	Khá	Khá	
Điểm TBCN	8,0	7,6	7,5	7,1	

Xếp loại tốt nghiệp THCS : Khá năm tốt nghiệp 2020

C. PHẦN ĐĂNG KÍ DỰ TUYỂN

I. ĐĂNG KÍ VÀO LỚP 10 THPT KHÔNG CHUYÊN:

- Đối tượng dự tuyển: (Bổ túc ghi BT; Phổ thông ghi PT) PT
- Tên trường THPT đăng kí đợt 1: NGUYỄN CHÍ THANH
- Tên trường THPT đăng kí đợt 3 (xét tuyển): NGÔ GIA TỰ.

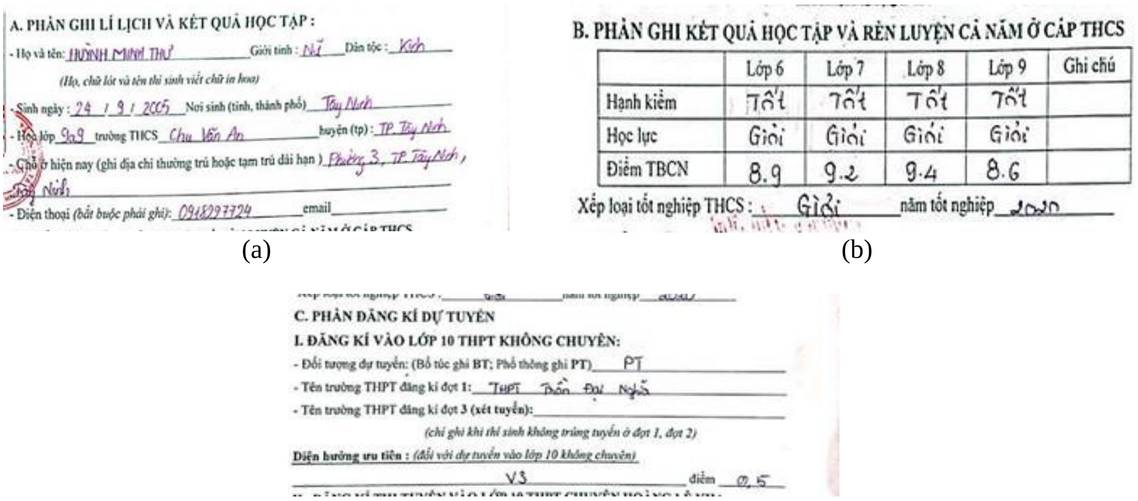
(chỉ ghi khi thí sinh không trúng tuyển ở đợt 1, đợt 2)

Diện hưởng ưu tiên : *(đối với dự tuyển vào lớp 10 không chuyên)*

Không. điểm

II. ĐĂNG KÍ THI TUYỂN VÀO LỚP 10 THPT CHUYÊN HOÀNG LÊ KHA

Figure 2. The main information container of the 10th grade enrollment form



A. PHẦN GHI LÍ LỊCH VÀ KẾT QUẢ HỌC TẬP :

- Họ và tên: HUỲNH MINH THƯ Giới tính : Nữ Dân tộc : Kinh

(Họ, chữ lót và tên thí sinh viết chữ in hoa)

- Sinh ngày : 24 / 9 / 2005 Nơi sinh (tỉnh, thành phố) Tây Ninh
- Học lớp 9a9 trường THCS Chu Văn An huyện (tp) : TP. Tây Ninh
- Chỗ ở hiện nay (ghi địa chỉ thường trú hoặc tạm trú dài hạn) Phường 3, TP. Tây Ninh, Tây Ninh
- Điện thoại *(bắt buộc phải ghi)*: 0918297724 email

(a)

B. PHẦN GHI KẾT QUẢ HỌC TẬP VÀ RÈN LUYỆN CẢ NĂM Ở CẤP THCS

	Lớp 6	Lớp 7	Lớp 8	Lớp 9	Ghi chú
Hạnh kiểm	Tốt	Tốt	Tốt	Tốt	
Học lực	Giỏi	Giỏi	Giỏi	Giỏi	
Điểm TBCN	8.9	9.2	9.4	8.6	

Xếp loại tốt nghiệp THCS : Giỏi năm tốt nghiệp 2020

(b)

C. PHẦN ĐĂNG KÍ DỰ TUYỂN

I. ĐĂNG KÍ VÀO LỚP 10 THPT KHÔNG CHUYÊN:

- Đối tượng dự tuyển: (Bổ túc ghi BT; Phổ thông ghi PT) PT
- Tên trường THPT đăng kí đợt 1: THPT Trần Đại Nghĩa
- Tên trường THPT đăng kí đợt 3 (xét tuyển):

(chỉ ghi khi thí sinh không trúng tuyển ở đợt 1, đợt 2)

Diện hưởng ưu tiên : *(đối với dự tuyển vào lớp 10 không chuyên)*

V3 điểm 0,5

(c)

Figure 3. Three separate regions; (a) area for recording student background information (Region A), (b) area for recording learning and training results (Region B), and (c) area of application registration (Region C)

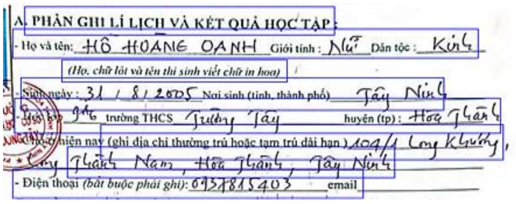
3.3. Character extraction

The deep learning method has the advantage of automatically learning features from the input information of the problem [17], [18]. We first apply the EAST model to detect text areas and create text boxes for image areas containing handwriting. This model is a powerful deep learning method used to detect texts presented on input images. It can find horizontal and rotated bounding boxes and can be used with any other text recognition method. The text detection system with EAST has eliminated redundant and intermediate steps and has only two stages. EAST uses a fully integrated network to generate text prediction words or lines directly. The generated predictions that can rotate the rectangle or the quadrilateral are further processed through the suppression step to yield the final output.

The EAST algorithm detects texts in the input image by creating a text box for each word or phrase, lead to many rectangular boxes for the detected words. Algorithm 1 is an algorithm to join the text box in each row to process image regions. As a result, the input image has many rows of information, and the output image also has many rows of text boxes. The output text lines are fed into OCR system in next step. Figure 4 illustrated results for the algorithm to join text boxes by row. In the final step of the EAST model to detect text boxes, based on advanced information such as fixed form, ratio of each field, position of each field, … We apply heurictic thresholds to separate large_boxes into separate fields per row to help the training step of data and other methods. The text box shows the correct semantics in the 10th-grade enrollment form in Tay Ninh province as shown in Figure 5.

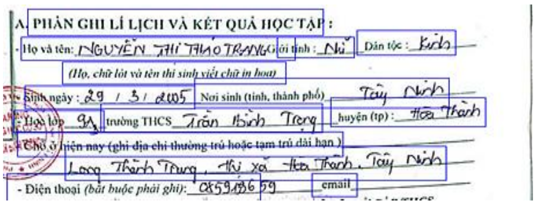
Algorithm 1: Algorithm to join text boxes row by row

```
Input: coordinates of the boxs put into stand_boxes.
Output: coordinates large_box = [(x_min, y_min, x_max, y_max)], As a result, the boxes are joined in rows.
1.   Put all the boxes put into stand_boxes [ ]
2.   Calculate the coordinates of the midpoint (y coordinates) of the Textbox to
     identify the Textboxes belonging to the same line, then sort the group_boxes in
     ascending order (y). Next, put them (the child group_box has been arranged as)
     into the same parent group_boxes
3.   Calculate (length) of group_boxs
4.   Loop group_box (from 0 to group_boxs) to calculate the coordinates of each
     text_box:
       x_min, x_max,
       y_min, y_max
5.   Calculate large_box = [(x_min, y_min, x_max, y_max)] {where are the top - left,
     bottom - right coordinates of each Textbox}
6.   Draw the large_boxes according to the calculated coordinates,
End the algorithm.
```



A. PHẦN GHI LÍ LỊCH VÀ KẾT QUẢ HỌC TẬP:
- Họ và tên: HÔ HOÀNG OANH Giới tính: Nữ Dân tộc: Kinh
(Họ, chữ lót và tên thí sinh viết chữ in hoa)
- Sinh ngày: 31 / 8 / 2005 Nơi sinh (tỉnh, thành phố) Tây Ninh
- Học lớp 9A6 trường THCS Trường Tây huyện (tp): Hòa Thành
- Chỗ ở hiện nay (ghi địa chỉ thường trú hoặc tạm trú dài hạn) 104/1 Long Khương, Long Thành Nam, Hòa Thành, Tây Ninh
- Điện thoại (bắt buộc phải ghi): 0937815403 email

Figure 4. Example of results of text box connection by row of area A



A. PHẦN GHI LÍ LỊCH VÀ KẾT QUẢ HỌC TẬP:
- Họ và tên: NGUYỄN THỊ THẢO TRANG Giới tính: Nữ Dân tộc: Kinh
(Họ, chữ lót và tên thí sinh viết chữ in hoa)
- Sinh ngày: 29 / 3 / 2005 Nơi sinh (tỉnh, thành phố) Tây Ninh
- Học lớp 9A trường THCS Trần Bình Trọng huyện (tp): Hòa Thành
- Chỗ ở hiện nay (ghi địa chỉ thường trú hoặc tạm trú dài hạn) Long Thành Trung, thị xã Hòa Thành, Tây Ninh
- Điện thoại (bắt buộc phải ghi): 085913659 email

Figure 5. Example of results of separated fields of area A

3.4. The string identifier

We propose a method to solve the OCR character recognition problem using CRNN and attention models to recognize Vietnamese handwriting in the 10th-grade enrollment form of Tay Ninh province [19]–[22]. The CRNN network model is a popular model that gives good results in print and handwriting recognition [23]. We have trained a CRNN model for Vietnamese handwriting recognition problems using the OCR technique with the dataset processed from the enrollment form. At the same time, we also provide a CRNN model for feature extraction and handwriting recognition, as shown in a Figure 6, trained on the enrollment form data set achieved relatively good results.

The CRNN model for the handwriting recognition problem presented in this paper consists of 2 parts: CNN and RNN + LSTM [24], [25]. Precisely, CNN extract features from the image. Therefore, the architecture of the CNN block must be suitable to receive input of size wxh. We place the output of the CNN block as the input of the RNN + LSTM block.

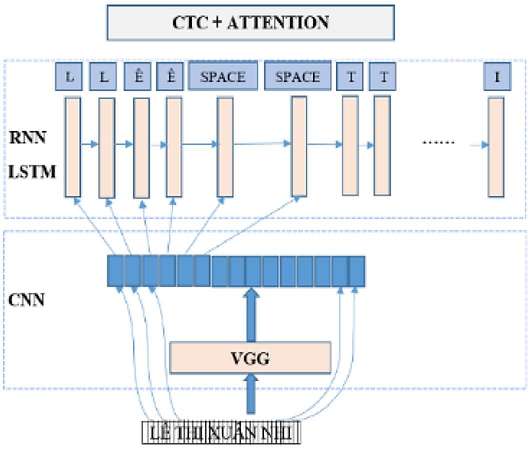


Figure 6. An identity pattern of CRNN

3.5. Model operation

The input image through the CNN block: the visual geometry group (VGG) component removes noise, reduces the dimensional space, and extracts features for output in the form of feature vectors (Feature map). Next, the RNN and LSTM block consist of two main components: RNN encoder and RNN decoder. RNN encoder helps to process the encoding features, RNN decoder as a decoder to process the output. Finally, CTC and ATTENTION improve the output by removing repeated characters and blanks (blank tokens) to produce a complete sentence. This problem is called output alignment or alignment problem [26]–[29].

3.6. Train the proposed network model

After locating the information areas to be extracted using the EAST model, we train the data against the proposed model. Due to the data collected, there are certain limitations described in section 4-experiments. So, in training data to produce an identification model to solve the OCR problem mentioned, we propose removing four fields (do not train data for these fields) such as conduct, graduation year, candidates for recruitment and school to register for the exam because the value is little changed. It does not guarantee the comprehensiveness of information in reality. Table 1 mentioned and specified the reason for the rejection.

Table 1. Ratio of data divided by fields

Numerical order	Name fields	Total number of experimental images	Number of images for training 80%	Number of photos to test 20%
		Total	1550	1240 310
1	Full name	307	246	61
2	Sex	39	31	8
3	Date of birth	93	74	19
4	Class	87	70	17
5	Secondary School	113	90	23
6	District (city)	90	72	18
7	Current accommodation	113	90	23
8	Phone	57	46	11
9	Academic ability	191	153	38
10	Grade Point Average for the whole year	168	134	34
11	Graduation High School Graduation	90	72	18
12	Priority Beneficiaries	113	90	23
13	Plus mark	92	74	18

We apply the word error rate (WER) measure to evaluate the word error rate in the data recognition of the trained model. After we get the results when using the built model to train and test the evaluation from the data set, we found that the recognition rate of handwritten Vietnamese characters is still low. Therefore, a spelling correction method was applied to improve the recognition rate. Spelling correction idea uses a result set that identifies incorrect results but approximates the correct results to compare with the complete data set, which is the fully collected contraints for comparison. We performed the spelling correction algorithm, experimented on the data fields and gave the following improved results.

- Fields that can correct spelling errors such as: Place of birth (name of province/city), district, secondary school name, gender, ethnicity, priority category, academic ability. Because these fields can collect all its occurrences.
- Fields that cannot be corrected include: Date of birth, phone number, full name, grade, grade point average. Since these fields have very large instances and possibly infinite numbers, complete statistics are not possible.

4. RESULTS AND DISCUSSION

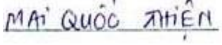
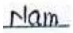
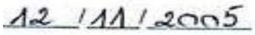
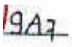
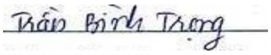
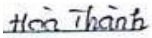
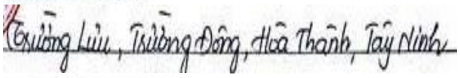
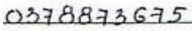
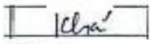
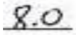
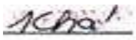
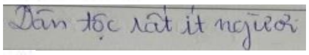
4.1. Dataset and implementation details

4.1.1. Data collection and pre-processing

We use the dataset collected from the 10^{th}-grade enrollment forms of three schools that organize the entrance exam (Nguyen Chi Thanh, Le Quy Don, and Tran Dai Nghia High School) in Tay Ninh province. The dataset was collected by scanning images of registration forms. Each form consists of 3 regions containing the following information: Region A-contains the personal information. Region B-includes a table of results of study and practice in four school years of candidates. Region C-contains registration information for grade 10 and corresponding priority points. Using the algorithm to separate the image into three regions has an accuracy of 100%. We used all 1550 images divided into 1240 images for training, 310 images for testing and evaluation. The data is divided according to the fields, as shown in Table 1.

For better training results, we have separated the data into separate fields shown in Table 2. Those are the input image files for the training of the network model. The output data is a text file containing the following information: full name, sex, date of birth, grade secondary school, district (city), current address, phone, academic ability, average mark of the whole year, valuation high school graduation, priority, inherited priority.

Table 2. Set of sample images divided by field for experiment

Numerical order	Name fields	Sample images have been separated
1	Full name	 MAi QUỐC THIỆN
2	Sex	 Nam
3	Date of birth	 12 /11/ 2005
4	Grade	 9A7
5	Secondary School	 Trần Bình Trọng
6	District (city)	 Hòa Thành
7	current address	 Trường Lưu, Trường Đông, Hòa Thành, Tây Ninh
8	Phone	 0378873675
9	Academic ability	 Khá
10	Average mark of the whole year	 8.0
11	Valuation High School Graduation	 1Khá
12	Priority	 Dân tộc rất ít người
13	inherited priority	1

4.1.2. Challenges of the collected data

Image quality much depends on the means of scanning equipment when collecting, the technique of taking pictures, the ambient light and the paper material of the information form, which greatly affects the image of the collected data. As well as the quality between the form images there are also differences. The big problem is Vietnamese hand-writing and the form has many fields, the information inside also depends on many factors such as: different writers, different types of pens, writing direction, and light density, character sharpness, writing speed and various scribbles. create the difficulty of the data set for the research problem.

4.2. Experimental environment

After collecting and normalizing the noise type of the dataset, we implemented and built the selected algorithms. We implemented the algorithms in Python 3.7 programming language with the configuration of training computer, testing CRNN model for handwriting recognition such as: Computer type-Lenovo, OS-Windows 10, Architecture-OS 64-bit, CPU-Intel Core i9 10900x 3.7g up 4.7g | 10 core | 20 thread, RAM-Gskill Trident Z RGB 128g/3600 (4x36g), HDD-SSD Samsung 970evo 1TB nvme m.2 pcie, Graphics Adapter-Graphics card that supports image processing: GPU 64GB, 128-bit-VGA: 2 x NVIDIA RTX 3090 24g Gddr6x.

4.3. Results of information detection based on EAST model

After applying the EAST model, the extraction results on the three regions reached the accuracy as shown in Table 3. Figure 7 shows the extraction ratio of the region containing text in the image of 3 regions A, B, and C. For region B, two algorithms are used (1. normal image processing and 2. deep learning algorithm with EAST model) [30].

Analysis: when creating a text box using the deep learning model-EAST, the detection rate of the region containing the text in the image is relatively high for regions A and C (87% and 83%). However, for region B (the region that has the learning results table of the student form), the accuracy is 54%, nearly 30% lower than in regions A and C. With conventional image processing algorithms for region B (with the same input data type), the result is 63%, higher than that of the EAST model (9%). The EAST model gives bad results for region B because this is a table. The table lines are noisy data that significantly affect the results of EAST.

Table 3. Result of detection rate for 3 regions A, B, C

	Number of test images	Total fields	Correct	Wrong	Ratio
Region A uses deep learning model-EAST	30	10 field x 30 image =300	260	40	87%
Region B uses conventional image processing algorithms	30	14 field x 30 image =420	265	155	*63%*
Region B uses deep learning model-EAST	30	14 field x 30 image =420	226	194	*54%*
Region C uses deep learning model-EAST	30	4 field x 30 image =120	100	20	83%

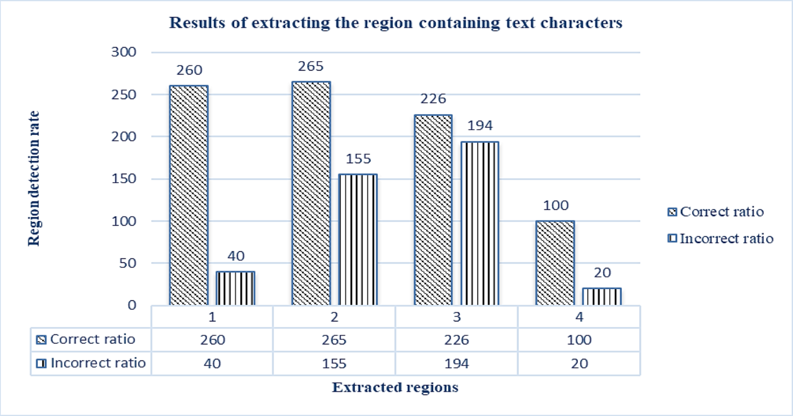


Figure 7. Extraction ratio of the region containing text in the image of 3 regions A, B, and C

4.4. Results of Vietnamese handwriting recognition using CRNN

Table 4 shows the results and using the OCR technique to check the recognition on a dataset with 1550 images, including 1240 train images and 310 test images. Table 5 shows the test results of correctly identified image regions, and Table 6 shows the results of wrongly recognized images after testing. Through the statistics of the results of training and evaluation, the OCR technique achieved an excellent rate with the WER measure of 36.02%. Each image has an average size of 32×525 with a recognition processing time of about 0.0471s. The total processing time of 310 images with an average size of 32×525 is 13,6214s.

Table 4. OCR training and identification results using the WER measure

Total number of experimental images	Number of training images	Number of test images	WER (%)
1550	1240	310	36,02

Table 5. Illustrated correctly recognized image region

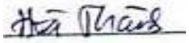
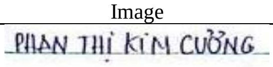
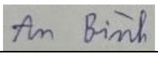
Name	Image	Label	Final	CTC	Attention
A_distric_010.jpg		Hòa Thành	Hòa Thành	HẢ	Hòa Thành
A_birthday_004.jpg		22/11/2005	22/11/2005	/11/k	22/11/2005

Table 6. Illustrated image region misidentified

Name	Image	Label	Final	CTC	Attention
a_name_test_084.jpg		PHAN THỊ KIM CƯỜNG	PHẠM THỊ TH	Px	PHẠM THỊ TH
a_shool_test_084.jpg		An Bình	Dân tộc rất	Ỹ	Dân tộc rất

5. CONCLUSION

This paper proposes a deep learning method to separate regions-EAST by applying the CRNN deep learning network model and OCR Vietnamese handwriting recognition technique on the 10th-grade enrollment form in Tay Ninh province. We implement a spelling correction algorithm to increase the efficiency of data recognition. Experiment results show that our model effectively utilizes data digitization in the education sector, potentially saving the provincial budget and human force, reducing data entry time, especially in many student records.

ACKNOWLEDGEMENTS

This research was supported by Sai Gon University under Fund Grant No. CSB2021-09.

REFERENCES

[1] N. Quang, "Hơn 4.000 học sinh toàn tỉnh thi tuyển sinh vào lớp 10 năm học 2020 – 2021," *Tayninh Magazine*, vol. 7, no. 2, 2020.

[2] M. T. Lệ, "Thông tin về tuyển sinh đầu cấp năm học 2020 – 2021," *Tayninh education Magazine*, vol. 3, no. 6, 2020.

[3] V. T. and T. Ba, "Hệ thống quản lý trường học vnEdu," *Khanhhoa Magazine*, vol. 8, no. 5, 2019.

[4] D. Nurseitov, K. Bostanbekov, D. Kurmankhojayev, A. Alimova, A. Abdallah, and R. Tolegenov, "Handwritten Kazakh and Russian (HKR) database for text recognition," *Multimedia Tools and Applications*, vol. 80, no. 21–23, pp. 33075–33097, 2021, doi: 10.1007/s11042-021-11399-6.

[5] J. C. Aradillas Jaramillo, J. J. Murillo-Fuentes, and P. M. Olmos, "Boosting handwriting text recognition in small databases with transfer learning," *Proceedings of International Conference on Frontiers in Handwriting Recognition, ICFHR*, vol. 2018-August, pp. 429–434, 2018, doi: 10.1109/ICFHR-2018.2018.00081.

[6] H. T. Nguyen, C. T. Nguyen, P. T. Bao, and M. Nakagawa, "A database of unconstrained Vietnamese online handwriting and recognition experiments by recurrent neural networks," *Pattern Recognition*, vol. 78, pp. 291–306, 2018, doi: 10.1016/j.patcog.2018.01.013.

[7] N. Toiganbayeva *et al.*, "KOHTD: Kazakh Offline Handwritten Text Dataset," 2021, [Online]. Available: http://arxiv.org/abs/2110.04075.

[8] D. Nguyen, N. Tran, and H. Le, "Improving Long Handwritten Text Line Recognition with Convolutional Multi-way Associative Memory," 2019, [Online]. Available: http://arxiv.org/abs/1911.01577.

[9] V. Carbune *et al.*, "Fast multi-language LSTM-based online handwriting recognition," *International Journal on Document Analysis and Recognition*, vol. 23, no. 2, pp. 89–102, 2020, doi: 10.1007/s10032-020-00350-4.

[10] M. Buta, L. Neumann, and J. Matas, "FASText: Efficient unconstrained scene text detector," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2015 International Conference on Computer Vision, ICCV 2015, pp. 1206–1214, 2015, doi: 10.1109/ICCV.2015.143.

[11] A. Gupta, A. Vedaldi, and A. Zisserman, "Synthetic Data for Text Localisation in Natural Images," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 2315–2324, 2016, doi: 10.1109/CVPR.2016.254.

[12] M. Jaderberg, K. Simonyan, A. Vedaldi, and A. Zisserman, "Reading Text in the Wild with Convolutional Neural Networks," *International Journal of Computer Vision*, vol. 116, no. 1, pp. 1–20, 2016, doi: 10.1007/s11263-015-0823-z.

[13] S. Tian, Y. Pan, C. Huang, S. Lu, K. Yu, and C. L. Tan, "Text flow: A unified text detection system in natural scene images," *Proceedings of the IEEE International Conference on Computer Vision*, vol. 2015 International Conference on Computer Vision, ICCV 2015, pp. 4651–4659, 2015, doi: 10.1109/ICCV.2015.528.

[14] O. R. D. and E. A. Varas, "Artificial Neural Networks for Stream Flow Prediction," *Journal of Hydraaulic Research*, vol. 40, no. 5, pp. 547–554, 2002.

[15] T. Q. Vinh, L. H. Duy, and N. T. Nhan, "Vietnamese handwritten character recognition using convolutional neural network," *IAES International Journal of Artificial Intelligence (IJ-AI)*, vol. 9, no. 2, p. 276, 2020, doi: 10.11591/ijai.v9.i2.pp276-281.

[16] X. Zhou *et al.*, "EAST: An efficient and accurate scene text detector," *Proceedings - 30th IEEE Conference on Computer Vision and Pattern Recognition, CVPR 2017*, vol. 2017-January, pp. 2642–2651, 2017, doi: 10.1109/CVPR.2017.283.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 770–778, 2016, doi: 10.1109/CVPR.2016.90.

[18] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, vol. 2016-December, pp. 2818–2826, 2016, doi: 10.1109/CVPR.2016.308.

[19] Q. Chen, "Evaluation of OCR Algorithms for Images with Different Spatial Resolutions and Noises," *University of Ottawa*, 2003.

[20] F. Gers, "Long short-term memory in recurrent neural networks," *Neural Computation*, vol. 9, no. 8, pp. 1735–1780, 2001, [Online]. Available: http://www7.informatik.tu-muenchen.de/~hochreit%0Ahttp://www.idsia.ch/~juergen%0Apapers://b6c7d293-c492-48a4-91d5-8fae456be1fa/Paper/p2558%5Cnfile:///C:/Users/Serguei/OneDrive/Documents/Papers/Long short-term memory in recurrent.pdf.

[21] A. D. Le, H. T. Nguyen, and M. Nakagawa, "Recognizing Unconstrained Vietnamese Handwriting by Attention Based Encoder Decoder Model," *Proceedings - 2018 International Conference on Advanced Computing and Applications, ACOMP 2018*, pp. 83–87, 2018, doi: 10.1109/ACOMP.2018.00021.

[22] P. J. Werbos, "Backpropagation Through Time: What It Does and How to Do It," *Proceedings of the IEEE*, vol. 78, no. 10, pp. 1550–1560, 1990, doi: 10.1109/5.58337.

[23] T. Mikolov, S. Kombrink, L. Burget, J. Černocký, and S. Khudanpur, "Extensions of recurrent neural network language model," *ICASSP, IEEE International Conference on Acoustics, Speech and Signal Processing - Proceedings*, pp. 5528–5531, 2011, doi: 10.1109/ICASSP.2011.5947611.

[24] Y. Kim, "Convolutional neural networks for sentence classification," *EMNLP 2014 - 2014 Conference on Empirical Methods in Natural Language Processing, Proceedings of the Conference*, pp. 1746–1751, 2014, doi: 10.3115/v1/d14-1181.

[25] Y. Goldberg, "Neural Network Methods for Natural Language Processing," *Synthesis Lectures on Human Language Technologies*, vol. 10, no. 1, pp. 1–311, 2017, doi: 10.2200/S00762ED1V01Y201703HLT037.

[26] D. Bahdanau, K. H. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

[27] M. Schall, M. P. Schambach, and M. O. Franz, "Multi-dimensional connectionist classification: Reading text in one step," *Proceedings-13th IAPR International Workshop on Document Analysis Systems, DAS 2018*, pp. 405–410, 2018, doi: 10.1109/DAS.2018.36.

[28] M. Yousef and T. E. Bishop, "OrigamiNet: Weakly-Supervised, Segmentation-Free, One-Step, Full Page Text Recognition by learning to unfold," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 14698–14707, 2020, doi: 10.1109/CVPR42600.2020.01472.

[29] A. Abdallah, M. Hamada, and D. Nurseitov, "Attention-based fully gated cnn-bgru for russian handwritten text," *Journal of Imaging*, vol. 6, no. 12, 2020, doi: 10.3390/jimaging6120141.

[30] D. Nurseitov, K. Bostanbekov, M. Kanatov, A. Alimova, A. Abdallah, and G. Abdimanap, "Classification of handwritten names of cities and handwritten text recognition using various deep learning models," *Advances in Science, Technology and Engineering Systems*, vol. 5, no. 5, pp. 934–943, 2020, doi: 10.25046/AJ0505114.

BIOGRAPHIES OF AUTHORS

Pham The Bao received his B.Sc. degree in Algebra from University of Natural Science-National University of HCM City, Vietnam in 1995. He also received MSc degree in Mathematical Foundation of Computer Science and Ph.D degree in Computer Science from University of Natural Science-National University of HCM City, Vietnam in 2000 and 2009, respectively. He was a lecturer and professor in Department of Computer Science, Faculty of Mathematics Computer Science, University of Natural Science, Vietnam from 1995 to 2018. He is currently dean and professor at Computer Science Department, Sai Gon University, Vietnam since 2019. He has published over 50 papers in international journals and conferences. His research includes image processing, pattern recognition and intelligent computing. He can be contacted at email: ptbao@sgu.edu.vn.

Le Tran Anh Dang received his B.E. (Control Engineering and Automation) in 2019 from Ho Chi Minh University of Technology, Vietnam. He is currently an AI engineer at Pharmacity Company, Ho Chi Minh city, Vietnam. He is also an AI researcher at Computer Science Laboratory, Sai Gon University, Viet Nam since 2020. His research includes machine learning, deep learning, speech recognition, computer vision, graph convolution neural network. He can be contacted at email: danglta1402@gmail.com.

Nguyen Duy Tam received his B.Sc. (Mathematics and Information technology) from Ho Chi Minh City University of Education in 2008 and received a master's degree in Computer Science from Saigon University in 2021, Vietnam. He is currently a computer science teacher at Tran Dai Nghia High School in Tay Ninh City. He is also an AI researcher at Computer Science Laboratory, Sai Gon University, Viet Nam since 2019. His research includes machine learning, deep learning, computer vision. He can be contacted at email: ndtam050185@gmail.com.

Nguyen Nhat Truong holds a BSc degree in Control and Automation from University of Technology-National University of Ho Chi Minh City in 2019. He is currently an AI and Machine Learning engineer at Vulcan Labs company, Ho Chi Minh city, Vietnam. He is also AI researcher at Computer Science Laboratory, Sai Gon University, Vietnam since 2020. His research includes speaker recognition speech signal processing, computer vision, time series and machine learning. He can be contacted at email: ntruonglhvl@gmail.com.

Pham Cung Le Thien Vu received his B.Sc. (Mathematics and Information technology) in 2015 and is studying for master's degree of Computer Science from Ho Chi Minh University of science, Vietnam. He is currently an AI and Machine Learning engineer at Heligate JSC company. He is also an AI researcher at Computer Science Laboratory, Sai Gon University, Vietnam since 2020. His research includes machine learning, deep learning, speaker recognition, computer vision, natural language processing. He can be contacted at email: phamcunglethienvu@gmail.com or vupclt@heligate.com.vn.

Trinh Tan Dat received a B.Sc. degree in Mathematics and Computer Sciences from University of Natural Science-National University of HCM City, Vietnam in 2010. He also received Master of Engineering (M.Eng.) and Ph.D degree in Electronics and Computer Engineering from Chonnam National University, Korea in 2013 and 2017, respectively. He is currently a lecturer at Computer Science Department, Sai Gon University, Vietnam since 2019. He is also an AI researcher at Computer Science Laboratory, Sai Gon University since 2019. His research areas of interest include speaker recognition, speech signal processing, computer vision and pattern recognition. He can be contacted at email: trinhtandat@sgu.edu.vn.