An Approach to Extract Information from Academic Transcripts of HUST

Nguyen Quang Hieu, Nguyen Le Quy Duong, Le Quang Hoa and Nguyen Quang Dat

May 9, 2023

An approach to extract information from academic transcripts of HUST

Nguyen Quang Hieu[1a], Nguyen Le Quy Duong[2], Le Quang Hoa[1b] (*), Nguyen Quang Dat[3]

[1] School of Applied Math. and Informatics, Hanoi University of Science and technology, Hanoi, Vietnam
(a) *hieu.nq185351@sis.hust.edu.vn*;
(b) *hoa.lequang1@hust.edu.vn*

[2] School of Information and Communications Technology, Hanoi University of Science and technology, Hanoi, Vietnam
duong.nlq210242@sis.hust.edu.vn

[3] HUS High School for Gifted Students, Hanoi University of Science, Vietnam National University, Hanoi, Vietnam
nguyenquangdat@hus.edu.vn

(*) Corresponding author.

Abstract. In many Vietnamese schools, grades are still being inputted into the database manually, which is not only inefficient but also prone to human error. Thus, the automation of this process is highly necessary, which can only be achieved if we can extract information from academic transcripts. In this paper, we test our improved CRNN model in extracting information from 126 transcripts, with 1008 vertical lines, 3859 horizontal lines, and 2139 handwritten test scores. Then, this model is compared to the Baseline model. The results show that our model significantly outperforms the Baseline model with an accuracy of 99.6% in recognizing vertical lines, 100% in recognizing horizontal lines, and 96.11% in recognizing handwritten test scores.

Keywords: Academic transcript; Image Processing; CRNN; CTC; Digit string recognition

1 Introduction

At Hanoi University of Science & Technology, after every exam, scores are recorded in academic transcripts and then transferred to the school's database by teachers.

Until now, this process has been done manually, which is time-consuming for the teachers and may lead to accidental mistakes such as the scores inputted being incorrect or the scores being assigned to the wrong students.

Currently, machine learning methods have been applied to automate these processes [1] [2] [3]. It also helps to free up manpower. By utilizing Image-Processing Techniques and Deep Learning, we can automate this procedure with a system that can extract necessary data.

This paper consists of 5 parts. The first part is the introduction. The second part is the previous studies of some authors in the world, on related image processing research methods. The third part is the method studied in this paper, including our proposed method. The fourth part is our results based on real data. And the last part is conclusion and acknowledgment.

2 Related works

Digit recognition is extremely useful. This is especially the case for schools where the process of inputting grades into database is still being done manually. In such schools, the assistance of digit recognition can significantly increase accuracy and reduce the time allotted to this process.

In 2018, Liu et al. [4] proposed a hybrid CNN-LSTM algorithm to recognize when a defect occurs in CO_2 welding. The algorithm is tested against 500 molten pool photos from the key laboratory of Robotics and Welding Technology of Guilin University of Aerospace Technology. The results of the CNN-LSTM algorithm were considered to be better than those of other basic algorithms (CNN, LSTM, CNN-3), with an accuracy of 85%, 88%, and 80% for 32x32, 64x64, and 128x128 images, respectively.

In 2019, Rehman et al. [5] utilized a hybrid CNN-LSTM model to analyze opinions in people's online comments. Against the IMDB movie reviews dataset and the Amazon movie reviews dataset, this model performs better than traditional machine learning techniques, with a precision of 91%, recall of 86%, an F-measure of 88%, and an accuracy of 91%.

In 2020, Yang et al. [6] compared the performance of CNN-LSTM model with analytical analysis and FEA methods in detecting the natural frequency of six different beams. The objective was to evaluate the first, second, and third-order frequencies of each beam. The CNN-LSTM model was concluded to be superior in both the test on robustness, with 96.6%, 93.7%, and 95.1% accuracy, respectively, and the test on extrapolability, with 95.4%, 92%, and 92.5% accuracy, respectively.

In 2019, Sivaram et al. [7] proposed a hybrid deep convolutional recurrent neural network to detect facial landmarks. The proposed model outperforms existing methods, such as CFAN, SDM, TCDN, and RCPR, on the FaceWarehouse database, with 99% precision, 99% recall, 99% F1-Score, 98.65% Accuracy, and 98.65% AUC or ROC.

In 2018, Xiangyang et al. [8] utilized a hybrid CNN-LSTM model to tackle the problem of text classification. On the Chinese news dataset (published by the Sogou Lab), the model proved superior to traditional KNN, SVM, CNN, and LSTM, with a precision of 90.13% under the CBOW model and 90.68% under the Skip-Gram model.

In 2017, Yin et al. [9] used CNN-RNN and C3D hybrid networks to detect emotions in videos from the AFEW 6.0 database. The objective was to assign one of seven emotions, namely anger, disgust, fear, happiness, sadness, surprise, and neutral, to each video in the test dataset. It was found that with 1 CNN-RNN model and 3 C3D models, an accuracy of 59.02% was achieved, surpassing the baseline accuracy of 40.47% and last year's highest accuracy of 53.8%.

In 2017, Zhan et el. [10] introduced a new RNN-CTC model to recognize digit sequences in three datasets, namely CVL HDS, ORAND-CAR (which is divided into two subsets, namely CAR-A and CAR-B), and G-Captcha. Even though the proposed model only achieved a recognition rate of 27.07% for the CVL dataset due to the fact that a segmentation-free method was used, the model outperformed other state-of-the-art methods on the CAR-A, CAR-B, and G-Captcha datasets, with recognition rates of 89.75%, 91.14%, and 95.15%, respectively.

3 Methodology

3.1 Convolutional Neural Networks

A convolutional neural network can be successfully used to solve most computer vision problems. It has been the space structure of image and fully connected network, and its characteristics make it more efficient than other traditional techniques. Ever since CNN was introduced, it has undergone countless optimization.

However, when it comes to deeper networks, a degradation problem arises. To tackle this, He et al. proposed a deep residual learning framework, ResNet. The basic structure of ResNet is shortcut connections. Shortcut connections are ones that pass over one or more layers. With this type of connection, we can deal with the vanishing gradient problem and build deeper networks; in other words, better feature representations can be acquired. In practice, shortcut connections can be adjusted on a case-by-case basis, depending on each specific problem.

In our proposed model, we design a 10-layer residual network that doesn't have any global pooling layers. To prevent divergence, we avoid using excessively deep CNN. Moreover, we make full use of residual learning to improve gradient propagation.

3.2 Recurrent Neural Networks

A Recurrent neural network is a neural network model where multiple connections between neurons create a directed cycle. With self-connection, it has the advantage of utilizing contextual data when mapping between input and output sequences. However, for traditional RNN, the range of accessible context in practice is limited because of the vanishing gradient problem. A viable solution is to apply a memory structure to RNN, which results in what's known as a long short-term memory (LSTM) cell. It is shown that this LSTM version of RNN is capable of addressing some inherent problems of traditional RNN as well as learning to solve long-term dependency problems. Presently, LSTM has become one of the most commonly used RNNs.

Regarding the sequence labeling problem, it is helpful to have access to context in both the future and the past. However, standard LSTM only takes into account past information and disregards future context. An alternative option is to create another LSTM to process information in reverse, which is called bidirectional LSTM, commonly abbreviated to Bi-LSTM. Bi-LSTM presents every training sequence forward and backward to two different LSTM layers, which are linked to the same output layer. This structure supplies the output layer with complete context, both in the past and the future, for all points in the input sequence.

3.3 Handwritten digit string recognition with CTC

Sequence characters recognition is a common problem of OCR. In this paper, we proposed an approach to recognize handwritten digit string. The main idea is using a recurrent neural network to recognize sequence information in images, after being extract feature by convolutional neural network, then get the final predicted results by the output connectionist temporary classification layer.

The input image is one-dimensional tensor (after resizing 40x100x1). For feature extraction, a backbone network is constructed with convolutional, max-pooling layers, and residual network. After every convolution layer, we performed batch normalization to prevent internal covariance shift. Output of feature extraction block are fed as a sequence into labeling block. To advoid vanishing gradient problem, we use two Bi-LSTM layers instead of traditional RNN layer. Finally, a fully connected layer is used to reduce the dimension to 13, before passing CTC layer. CTC layer has two main purposes, one is to calculate the loss, the other is to decode the output. The full architecture is shown in figure 1

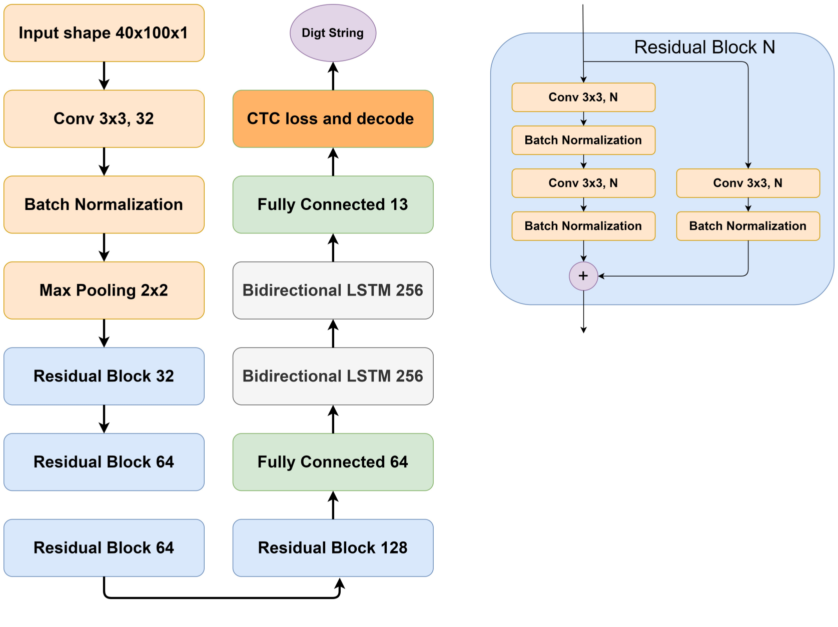


Fig. 1: Our proposed architecture model

3.4 Proposed method

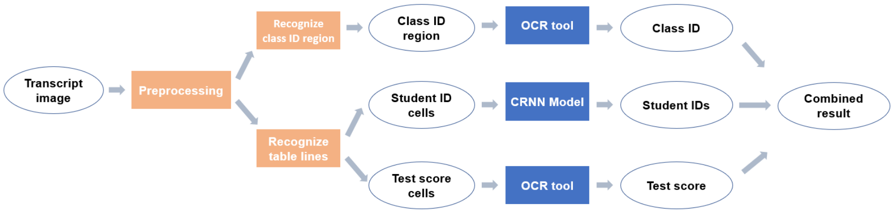
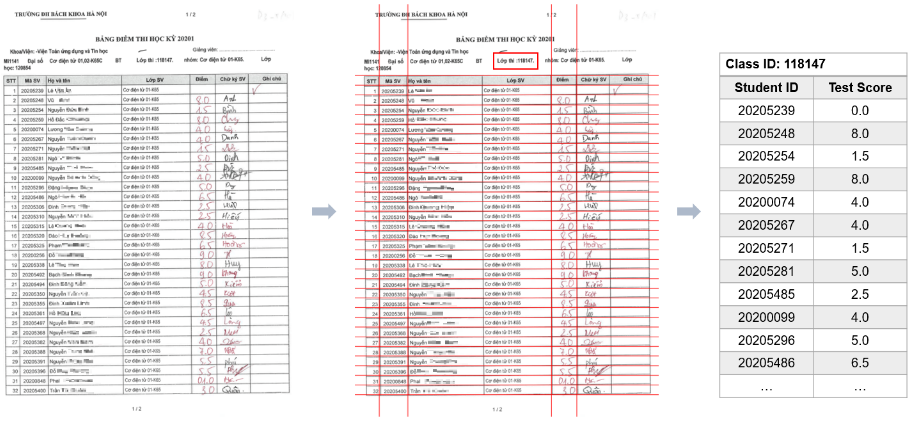


Fig. 2: Our proposed method flowchart

The first step of our method is image preprocessing. Transcript images are binarize, removing noises by Gaussian filter. We deskew the images by Projection profile algorithm. For class ID recognition, we use Template matching followed by OCR tools. To recognize and calculate coordinates of lines in transcript, horizontal and vertical masks generated by morphological operations are fed into Hough transformation. After having full coordinates of lines, cells of student IDs and test scores are cropped. For student IDs, are sequence of printed digts, can easily recognized by available OCR tools. In our method, we use Tesseract-OCR, which is inherently very accurate in recognizing printed characters. For test scores, we use our Handwritten digits recognition model with CTC decode. Finally, student IDs and test scores are combined.



Class ID: 118147

Student ID	Test Score
20205239	0.0
20205248	8.0
20205254	1.5
20205259	8.0
20200074	4.0
20205267	4.0
20205271	1.5
20205281	5.0
20205485	2.5
20200099	4.0
20205296	5.0
20205486	6.5
...	...

Fig. 3: Result of automatic score-inputting system

4 Experiment and results

4.1 Evaluation of Image-preprocessing

By using a dataset consisting of images of 126 academic transcripts with 1008 vertical lines and 3859 horizontal lines, the results of the Baseline model and my improved model in detecting lines are shown below:

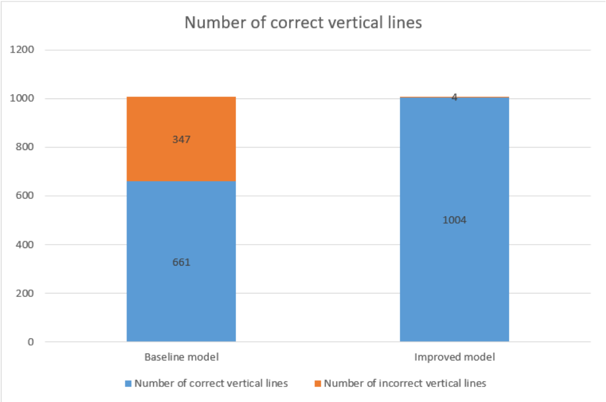


Fig. 4: Results of the two models in detecting vertical lines

The Baseline model achieved an accuracy of 65.57% for vertical lines, whereas the improved model had an accuracy of 99.6%.

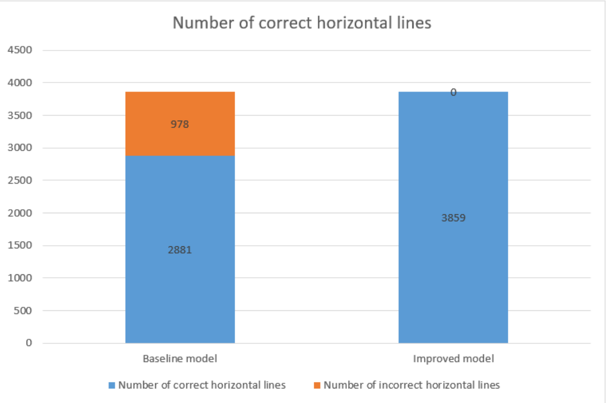


Fig. 5: Results of the two models in detecting horizontal lines

The Baseline model achieved an accuracy of 74.65% for horizontal lines, whereas the improved model had an absolute accuracy of 100%.

4.2 Evaluation of models in recognizing handwritten test scores

By using an extracted dataset with 2139 handwritten test scores, the results of the CNN Baseline model and the my CRNN model are shown below:

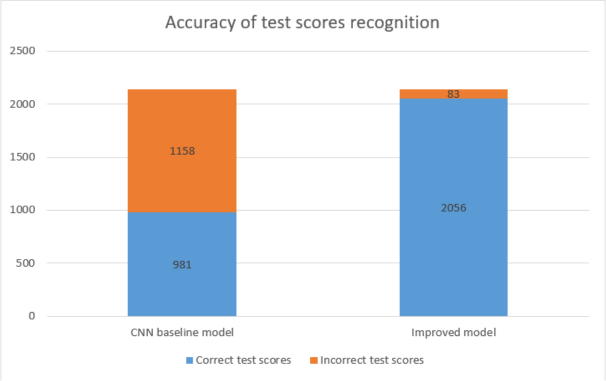


Fig. 6: Results of the two models in recognizing handwritten test scores

The Baseline model achieved an accuracy of approximately 45.86%, whereas my improved CRNN model had an accuracy of 96.11%.

4.3 Evaluation of automatic score-inputting system

To evaluate the accuracy of the entire score-inputting system, we tested it on a dataset of 75 scanned academic transcripts with 162 images of size 1653 x 2337.

a) Evaluation of Baseline model

Results of the Baseline model:

- The model was able to accurately detect lines in 92 images and recognize class IDs of 51 images (20 academic transcripts), achieving an accuracy of 31.4%.
- Among 3596 student IDs, the model correctly recognized 2201 IDs, achieving an accuracy of 61.2% (The majority of images in which the lines were accurately detected all had their student IDs recognized by the model).
- Among 3596 test scores, the model was able to accurately recognize 1532 test scores, achieving an accuracy of 42.6%.

b) Evaluation of improved model

Results of the improved model:

- In 22 images, the model misidentified 1 vertical line. However, these misidentified lines didn't affect the columns of data that needed to be recognized. Horizontal lines, on the other hand, were all accurately detected. In all 162 images, the model correctly recognized the class IDs with an accuracy of 100%.

- Among 3596 student IDs, the model correctly recognized 3481 IDs, achieving an accuracy of 96.8%.
- Among 3596 test scores, the model was able to accurately recognize 3451 test scores, achieving an accuracy of 95.9%.

5 Conclusion

In this research paper, we have introduced a new approach to the case of handwritten test scores into the computer. When using additional auxiliary features on the printout such as horizontal and vertical lines on the A4 paper, we have achieved very good results in clearly separating handwritten letters and numbers, thereby increasing adds precision to reading handwritten letters and numbers into numeric data.

In the future, we will put more research on some related issues, in order to further increase the accuracy:

- Identify several records of the same person.

- Identify both letters and numbers at the same time (points are written in both numbers and words in the one handwritten paper)

6 Acknowledgment

We, the authors, would like to thank the researchers of Hanoi University of Science and Technology (Hanoi, Vietnam), who helped us a lot in collecting data and testing the model application in practice.

References

1. Vu Minh Duc, Tran Ngoc Thang, "Text Spotting in Vietnamese Documents", In: Anh, N.L., Koh, SJ., Nguyen, T.D.L., Lloret, J., Nguyen, T.T. (eds) ***Intelligent Systems and Networks***. Lecture Notes in Networks and Systems, vol 471. Springer, Singapore. https://doi.org/10.1007/978-981-19-3394-3_17
2. Ho Trong Anh, Tran Anh Tuan, Hoang Phi Long, Le Hai Ha, Tran Ngoc Thang, "Multi Deep Learning Model for Building Footprint Extraction from High Resolution Remote Sensing Image". In: Anh, N.L., Koh, SJ., Nguyen, T.D.L., Lloret, J., Nguyen, T.T. (eds) ***Intelligent Systems and Networks***. Lecture Notes in Networks and Systems, vol 471. Springer, Singapore. https://doi.org/10.1007/978-981-19-3394-3_29
3. Long Phi Hoang, Dung Duy Le, Tuan Anh Tran, Thang Tran Ngoc, "Improving Pareto Front Learning via Multi-Sample Hypernetworks", https://doi.org/10.48550/arXiv.2212.01130
4. T. Liu, J. Bao, J. Wang, Y. Zhang, "A Hybrid CNN–LSTM Algorithm for Online Defect Recognition of CO2 Welding". *Sensors*, vol 18, No. 12, 4369, 2018. https://doi.org/10.3390/s18124369
5. Rehman, A.U., Malik, A.K., Raza, B. et al. A "Hybrid CNN-LSTM Model for Improving Accuracy of Movie Reviews Sentiment Analysis". *Multimed Tools Application*, vol 78, pp 26597–26613, 2019. https://doi.org/10.1007/s11042-019-07788-7

6. Ruoyu Yang, Shubhendu Kumar Singh, Mostafa Tavakkoli, Nikta Amiri, Yongchao Yang, M. Amin Karami, Rahul Rai, ”CNN-LSTM deep learning architecture for computer vision-based modal frequency detection”, *Mechanical Systems and Signal Processing*, vol 144, page 106885, 2020. https://doi.org/10.1016/j.ymssp.2020.106885
7. M Sivaram, V Porkodi, Amin Salih Mohammed, V Manikandan, Lebanese French University, Iraqi Kurdistan, ”DETECTION OF ACCURATE FACIAL DETECTION USING HYBRID DEEP CONVOLUTIONAL RECURRENT NEURAL NETWORK”, *ICTACT Journal on Soft Computing*, vol 9. No. 2, pp 1844-1850, 2019. https://doi.org/10.21917/ijsc.2019.0256
8. Xiangyang She and Di Zhang, ”Text Classification Based on Hybrid CNN-LSTM Hybrid Model,” *2018 11th International Symposium on Computational Intelligence and Design* (ISCID), pp. 185-189, 2018. https://doi.org/10.1109/ISCID.2018.10144.
9. Yin Fan, Xiangju Lu, Dian Li, Yuanliu Liu, ”Video-based emotion recognition using CNN-RNN and C3D hybrid networks”, *Conference: International Conference on Multimodal*, 2017 https://doi.org/10.1145/2993148.2997632.
10. Hongjian Zhan, Qingqing Wang, Yue Lu, ”Handwritten Digit String Recognition by Combination of Residual Network and RNN-CTC”. In: Liu, D., Xie, S., Li, Y., Zhao, D., El-Alfy, ES. (eds) *Neural Information Processing.* ICONIP 2017. Lecture Notes in Computer Science(), vol 10639. Springer, Cham. https://doi.org/10.1007/978-3-319-70136-3_62
11. Williams, Ronald J.; Hinton, Geoffrey E.; Rumelhart, David E. ”Learning representations by back-propagating errors”. *Nature* vol 323 (6088), pp 533–536, 1986. https://doi.org/10.1038/323533a0