

Vietnamese handwritten character recognition using convolutional neural network

Truong Quang Vinh, Le Hoai Duy, Nguyen Thanh Nhan

Faculty of Electrical and Electronics Engineering, Ho Chi Minh City University of Technology (HCMUT), Vietnam
National University – Ho Chi Minh, Vietnam

Article Info

Article history:

Received Dec 12, 2019

Revised Feb 10, 2020

Accepted Apr 27, 2020

Keywords:

Handwriting recognition
Convolutional Neural Network
Overfitting problem
Vietnamese handwriting

ABSTRACT

Handwriting recognition is one of the core applications of computer vision for real-word problems and it has been gaining more interest because of the progression in this field. This paper presents an efficient model for Vietnamese handwritten character recognition by Convolutional Neural Networks (CNNs) – a kind of deep neural network model which can achieve high performance on hard recognition tasks. The proposed architecture of the CNN network for Vietnamese handwritten character recognition consists of five hidden layers in which the first 3 layers are convolutional layers and the last 2 layers are fully-connected layers. Overfitting problem is also minimized by using dropout techniques with the reasonable drop rate. The experimental results show that our model achieves approximately 97% accuracy.

This is an open access article under the [CC BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



Corresponding Author:

Truong Quang Vinh,
Faculty of Electrical and Electronics Engineering,
Ho Chi Minh City University of Technology (HCMUT), Vietnam National University – Ho Chi Minh,
268 Ly Thuong Kiet, district 10, Ho Chi Minh city, Vietnam.
Email: tqvinh@hcmut.edu.vn

1. INTRODUCTION

Handwritten character recognition has recently become one of the most challenging research areas when it comes to pattern recognition and image classification [1-3]. The purpose of handwritten character recognition is to create the ability of computer to identify and interpret intelligible handwritten. Generally, there are two approaches - Online Handwritten recognition [4-5] which uses movement of the pen as input data and Offline Handwritten Recognition [6-7] which uses only information of input scanned images. There are many difficulties of handwriting recognition due to the fact that each person has a unique writing style and people often write so quickly that makes their character's shape unclear and rarely standardized [8-9]. Many researchers and scientists have concentrated on new techniques and methods that would address these problems with a higher rate of accuracy. Some competitions were opened to find the best algorithm for handwriting recognition with the dedicated dataset [10-11].

In the past few decades, many breakthroughs have been made in machine learning, especially Deep Learning. This resulted in the creation of various algorithms that have shown their improved efficiency in many important areas such as computer vision, natural language processing, and big data. Many machine learning algorithms including Support Vector Machines (SVM) [12], Naive Bayes Classification [13], and deep learning algorithms such as Convolutional Neural Networks (CNNs) [14], Recurrent Neural Networks (RNNs) [15], combined with the development beyond the computing speed of hardware devices have created models that can outperform human performance in many computer vision tasks, especially classification tasks such as handwriting recognition.

Many researches on handwritten character recognition for English, Japanese, and Chinese language have been published [16-18]. However, there are few of studies on Vietnamese handwriting recognition. Vietnamese alphabet is based on the Latin script with the addition of nine accent marks to create additional sounds, and five strokes to express the tone of words. Therefore, it is a challenge for researchers to develop algorithms to recognize Vietnamese handwritten characters. Some authors presented Vietnamese handwriting recognition methods using support vector machine (SVM) with feature extraction methods [19-20]. However, the performance depends on the robustness of feature extraction methods and thus the accuracy is decreased when the handwritten characters are natural and unconstrained. Some models of deep learning such as CNN, RNN, and Bidirectional Long Short-Term Memory have been applied to improve the recognition accuracy rate [21-22]. Nevertheless, accent marks and strokes in Vietnamese language make it difficult to recognize handwriting correctly.

In this paper, Vietnamese handwritten character dataset for training a supervised model and a CNN architecture for Vietnamese handwritten character recognition are presented. A technique called dropout is applied to prevent our model from overfitting, to guarantee the generalization property and to improve its performance. This technique includes a unit's temporary removal from the network. This removed unit is selected in a random way only during the training stage. The goal of our work is to create a Vietnamese handwritten character dataset containing 29 Vietnamese characters and to build a deep learning model that can achieve a high degree of accuracy for Vietnamese handwritten character recognition.

The rest of the paper is composed as follows. In section II, the overview of the CNN model and the proposed CNN architecture for Vietnamese handwritten character recognition are presented. In section III, the experimental results are shown in detail. Finally, in the last section, the conclusion is given.

2. RESEARCH METHOD

In this section, the overview architecture of our offline Vietnamese handwritten character recognition system based on Convolutional neural network is presented. Besides, a briefly discussion of the fundamental features of CNN and a dropout technique which combines fully-connected layers to prevent overfitting phenomenon for the propose model are described.

2.1. Convolutional neural networks (CNNs)

CNN is the current state-of-the-art model architecture for image classification tasks [23]. It has been employed to solve a variety of tasks including image recognition problems, natural language processing challenges by taking advantage of the way the input images sensibly constrain the architecture. CNNs learns features of images on sub-regions instead of scaling full images. In particular, CNNs arrange layers and their neurons in 3 dimensionalities: width, height and depth. CNNs passes the images which contains the raw pixel data through a series of filters to extract and learn higher-level features, which can be used for classification phase of the model. CNN is built from three components. Convolutional layers (Conv layers) are the major building block of a Convolutional Network which make use of a specific amount of convolution filters for the image. For each sub-region, Conv layers perform convolution operation of the sub-region considered as a matrix of pixel values and the filter considered as a matrix of real numbers to produce a single value in the output feature map. Convolution layers keep the spatial relationship between pixels by learning image features using sub-region of input data. An extra operation called ReLU has been used after every Convolution layer. ReLU acts as a non-linear activation function introducing non-linearity into the model, which is really common in real-world data.

Pooling layers involve down-sampling the image data extracted by the convolutional layers to reduce the dimension of each feature map but still retains the most important information so as to minimize processing time. A commonly used pooling algorithm is known as max pooling, which defines sub-regions of feature map by picking out the largest element and getting rid of other elements.

Dense (fully connected) layers are traditional Multi-layer Perceptron, which classify the features extracted by the convolutional layers and are down-sampled by the pooling layers. In a dense layer, every node in the layer is connected to every node in the preceding layer. For classification task, an activation function called softmax regression is used most often.

Typically, a CNN consists of convolutional layers, each of which is followed by a pooling layer. They are called convolutional modules and responsible for performing feature extraction. The last convolutional module proceeds after one or more dense layers that performs classification task.

2.2. Dropout technique

A large neural network contains multiple non-linear hidden layers can have a very expressive performance on training set but achieve a bad result on test set [24-25]. This is overfitting phenomenon. This

is the case when sampling noise is brought about by the lack of training data for a complex model. Dropout is a technical solution to this problem. This drop units (hidden and visible) from the neural network during training. The number of dropped units will be determined by probability p (dropout rate). Such dropped units are chosen randomly.

2.3. ReLU activation function

ReLU (Rectified Linear Unit) function is a non-linear function. Today, many researchers and scientists use ReLU function to introduce non-linear property into deep neural network model instead of using other activation functions. ReLU function makes the training of the model faster than regular activation functions such as sigmoid and tanh functions, because it does not compute exponentials and divisions. The formula of ReLU is $f(x) = \max(0, x)$.

2.4. The CNN architecture for Vietnamese handwritten character recognition:

The proposed CNN architecture is shown in Figure 1. It contains five hidden layers with weights. The first three are convolutional layers and the remaining two are fully-connected layers. Each convolutional layer is followed by one ReLU activation function and one max pooling layer with a filter of size 2×2 and a stride of 2 pixels. Also, the first fully-connected layer is followed by one dropout layer.

The first convolutional layer applies 32 kernels of size 11×11 (extracting 11×11 sub-region) with a stride of 2 pixels. It takes a 100×100 image as an input. The second convolutional layer uses outputs of the first convolutional layer as an input, and filters it with 64 kernels of size 5×5 with a stride of 2 pixels. The third convolutional layer has 128 kernels of size 3×3 which are connected to the outputs of the second convolutional layer. The first fully-connected layers (dense layer) have 1024 neurons. They are combined with dropout regularization rate of 0.4 (40% of neurons will be dropped during training). The last fully-connected layers (dense layer) have 29 neurons equal to the number of class. The output layer is a softmax layer which takes the output of last fully-connected layers as an input and will return the raw values for our predictions. These values are the probability values that represent the probability of each class's sample.

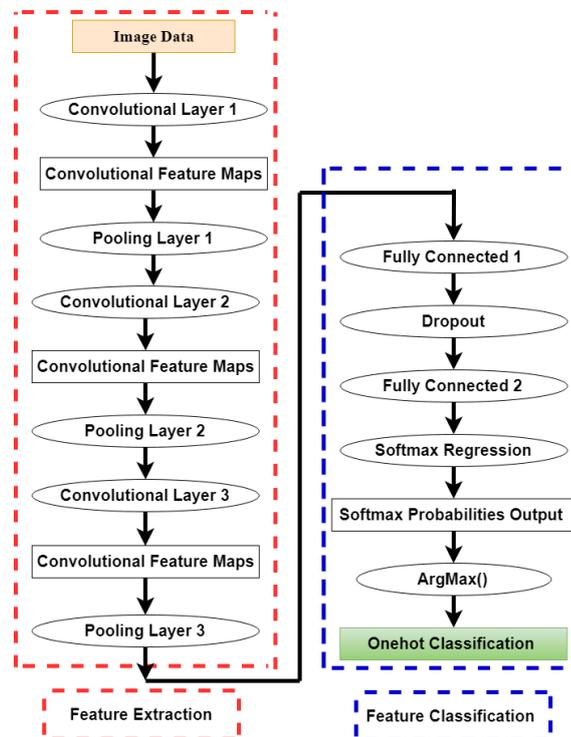


Figure 1. An illustration of the architecture of our CNN

The architecture is the combination of two blocks: Feature Extraction block and Feature Classification block. The Feature Extraction takes an original image as an input and extract abstractive feature by using Convolutional layer. The Feature Classification takes the output of the Feature Extraction as

an input and feeds these features into a deep neural network to perform the classification task. In the Feature Classification block, dropout unit is used to prevent overfitting phenomenon, and thus improve the performance of the model.

3. EXPERIMENTAL RESULTS

In this section, the process to collect the images of Vietnamese handwritten characters to build the dataset is described. Next, the training process and the testing results are presented. Finally, the demo application which can recognize Vietnamese handwritten characters is shown.

3.1. Dataset

Until recently, datasets of labeled images of Vietnamese handwritten characters have not been published. The 29 Vietnamese characters was constructed from 23 Latin alphabet and 6 unique characters. Therefore, we have collected six Vietnamese special character examples written by 500 different persons combine with modified examples of 23 Latin alphabet in NIST Special Database 19 which contains binary images of handwritten characters to create Vietnamese handwritten character dataset for the purpose of training and testing a supervised learning model for classification task. The six Vietnamese special character set is composed of 15,000 patterns that was collected by two methods. The first method is that an android application is developed to collect handwriting images on mobile device. The application allows users write a letter on screen and then send this image of letter to our server. Although this method is very fast for collecting and processing data because it is done automatically, the fact that it is quite difficult to reach a sufficient number of users ensuring the diversity of data sources. Therefore, the second method which uses handwriting on collected papers of HCMUT students is proceeded. The students wrote characters on grid paper. Then, we scanned those papers and separated them into single character images. The original black and white images were size-normalized and centered in a 100x100 box. This method takes a lot of effort to process data manually, but it ensures the diversity of writers. By combining these two methods, we have achieved two criteria: to collect data quickly and make sure the variety of data sources.

Our final Vietnamese handwritten character dataset contains 72,500 labelled images of 29 Vietnamese handwritten characters that was constructed from 57,500 patterns of NIST Dataset and 15,000 patterns of our six Vietnamese unique character. It was divided into three sets: 58,000 examples in training set, 7,250 examples in validation set, and 7,250 examples in testing set.

In order to collect data conveniently, we develop an Android mobile application called Hody Paint for collecting data from internet users as shown in Figure 2. Users can draw a character on the painting window, send an image of this character to our database by upload button, press clear button, and repeat these steps above for new letters.

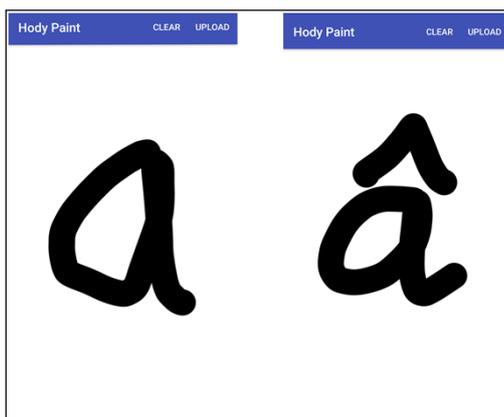


Figure 2. Mobile application for collecting data

Figure 3a shows some image samples of unique Vietnamese characters that were collected by our two methods and Figure 3b shows some samples of 23 Latin alphabets that were picked from NIST Special Database 19. Generally, data in training set, validation set, and test set should follow the same probability distribution [26]. Therefore, when building our dataset for machine learning or deep learning system, we

have performed a preprocessing including zooming, resizing, and binarizing on 23 Latin alphabets that come from NIST Special Database 19 to make all images in our dataset have the same distribution.

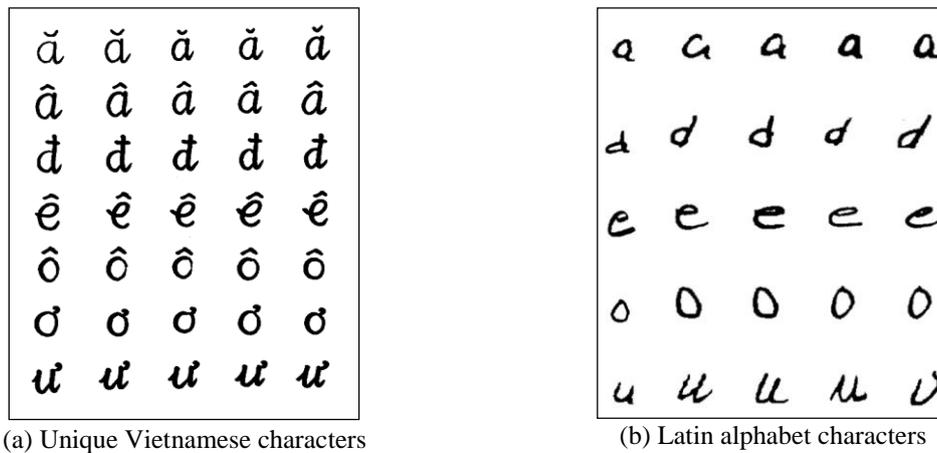


Figure 3. Examples of Vietnamese characters in our dataset

3.2. Training process

The training process was done by the Google Collaboratory server. For both training phase and testing phase, we chose categorical cross-entropy function to be a loss function for our CNN architecture. It is a softmax activation plus a cross-entropy loss. This function takes the measurements of how closely the model's predictions match the target classes in performing multi-class classification operation. The proposed model was trained by using mini-batch gradient descent as optimization algorithm. It is a variant of gradient descent algorithm. The mini-batch gradient descent is chosen because it provides a computationally more efficient process than stochastic gradient descent. In this work, mini-batch gradient descent was configured with a batch size of 128 and learning rate of 0.001 as the optimization algorithm. An equal learning rate was employed for all layers. The network was trained for 70,000 steps equivalent to 155 epochs through 58,000 examples of 29 characters in the training set. In addition to loss metric, the accuracy metric was used to evaluate training process.

3.3. Testing results

The accuracy metric is utilized to evaluate the model performance on our Vietnamese handwritten character dataset. The accuracy metric is defined as the ratio between the number of correct predicted samples and the total samples. The proposed model achieved an accuracy 97.2% on test set (top-1 test set error rate of 2.8%) after training 69,010 steps. Figure 4 shows the accuracy graph of our model on the test set and Figure 5 presents the training and validation loss.

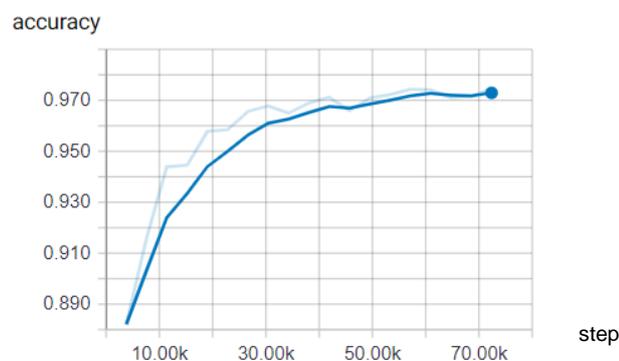


Figure 4. The accuracy of our model on test set as a function of iteration

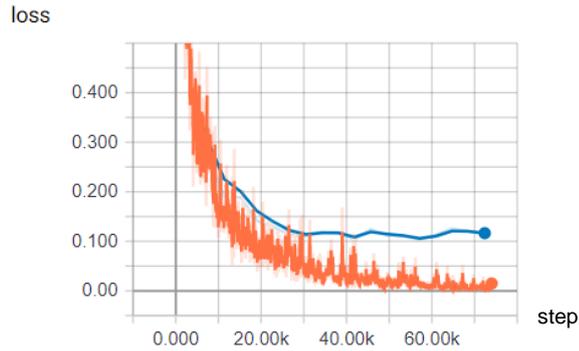


Figure 5. Training (—) and validation (—) loss of our model as a function of iteration

In order to compare our algorithm with other models, two methods of [19-20] were used for references. The proposed method in [19] tries to segment the accented characters into two parts: the root character and the accent. Then, the SVM model is applied to classify the Vietnamese characters. In some cases, this method meets problem in which the character is not written in one connected component. Therefore, the segmentation process can be failed. The method of the authors in [20] uses contour and project histogram features with SVM classification to recognize handwritten characters. However, those features are not enough to characterize Vietnamese handwritten characters with unconstrained accents and strokes. Our CNN model can solve the problem of natural and unconstrained Vietnamese characters. The feature extraction block by 3 convolutional layers can learn higher-level features, and then the feature classification block which are built by 2 fully connected layers with dropout techniques can prevent overfitting issue and thus provide the best performance of handwriting recognition. As a result, the comparison which is described in Table 1 shows that the proposed method outperforms two other methods.

Table 1. The comparison between the proposed method and others

	De Cao Tran [19]	Thach Tran Van et al. [20]	The proposed
Model	Accented character segmentation + SVM classification	Contour and project histogram-based feature extraction + SVM	CNN (3 convolutional layers + 2 fully connected layers)
Dataset	UNIPEN + IRONOFF + self-collected Vietnamese characters (39,800 samples)	Self-collected Vietnamese characters (27,000 samples)	NIST dataset + self-collected Vietnamese characters (72,500 samples)
Accuracy	91.3%	93.23%	97.2%

3.4. Demo application for Vietnamese handwritten character recognition

In order to demonstrate a real application for Vietnamese handwritten character recognition, we designed a desktop software by Qt toolkit based on our pretrained CNN model. The application can take an input as an image of a Vietnamese character that was drawn directly into painting window of the app or selected from an existing image and then produce a digital character of this character in input image. The application has two modes: predict directly image from painting window (online mode) and predict from external image (offline mode). For online mode, the application captures painting window and makes prediction based on this captured image. Users also can change brush weight and painting window size in this mode. For offline mode, users can select an image that contains a character need to recognize by Open button. The result window shows two information of prediction: top 5 score of prediction in detail mode and the highest score as the final result in Result option. This application can run on both Windows OS and Linux OS. Figure 6 shows the UI of Vietnamese handwritten character recognition application. The demo video of the proposed method has been uploaded to YouTube at <https://youtu.be/NY8GIqIh660>.

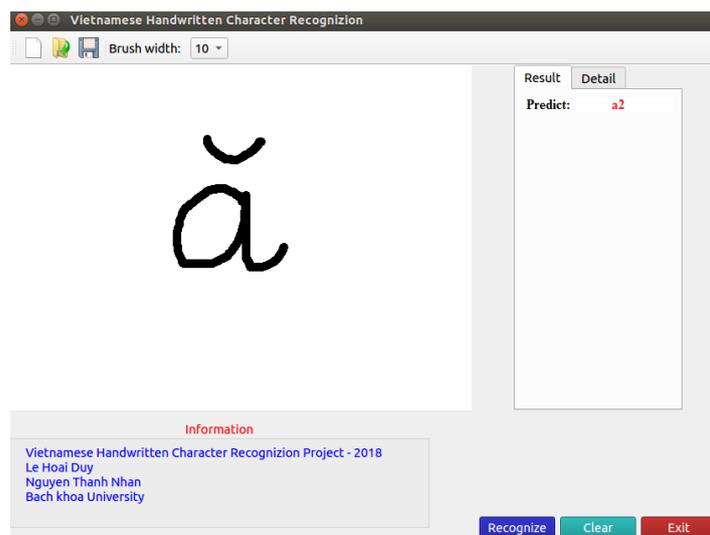


Figure 6. Vietnamese handwritten character recognition application

4. CONCLUSION

In this paper, we presented an efficient CNN architecture for Vietnamese handwritten character recognition. Our CNN model is built with 3 convolutional layers and 2 fully connected layers. A dropout technique is combined with the fully connected layers to prevent overfitting phenomenon. The experiments on a handwriting database show that our model can achieve approximately 97% accuracy. Moreover, a Qt application for Vietnamese handwritten character recognition was designed to demonstrate the proposed CNN architecture.

REFERENCES

- [1] A. & M. S. & R. S. & M. S. & D. S. Priya, "Online and offline character recognition: A survey", *International Conference on Communication and Signal Processing (ICCSP)*, 2016.
- [2] M. S. a. N. Sahu, "A Survey on Handwritten Character Recognition (HCR) Techniques for English Alphabets", *Advances in Vision Computing: An International Journal (AVC)*, vol. 3, March 2016.
- [3] Zouhaira Noubigh, Monji Kherallah, "A survey on handwriting recognition based on the trajectory recovery technique", *2017 1st International Workshop on Arabic Script Analysis and Recognition (ASAR)*, 2017.
- [4] S. España-Boquera, M.J. Castro-Bleda, J. Gorbe-Moya, F. Zamora-Martinez, "Improving Offline Handwritten Text Recognition with Hybrid HMM/ANN Models", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, no. 4, pp. 767 - 779, 2011.
- [5] Yanwei Wang ; Xiaoqing Ding ; Changsong Liu, "Topic Language Model Adaption for Recognition of Homologous Offline Handwritten Chinese Text Image", *IEEE Signal Processing Letters*, vol. 21, no.5, pp. 550 - 553, 2014.
- [6] Daniel Keysers, Thomas Deselaers, Henry A. Rowley, Li-Lun Wang, Victor Carbune, "Multi-Language Online Handwriting Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1180 - 1194, 2017.
- [7] Zecheng Xie, Zenghui Sun, Lianwen Jin, Hao Ni, Terry Lyons, "Learning Spatial-Semantic Context with Fully Convolutional Recurrent Network for Online Handwritten Chinese Text Recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 8, pp. 1903 - 1917, 2018.
- [8] S. Attigeri, "Neural Network based Handwritten Character Recognition system", *International Journal of Engineering And Computer Science*, vol. 7, no. 3, pp. 23761-23768, March 2018.
- [9] N. T. M. K. Rania MAALEJ, "Online Arabic Handwriting Recognition with Dropout applied in Deep Recurrent Neural Networks", *12th IAPR Workshop on Document Analysis Systems*, 2016.
- [10] Cheng-Lin Liu, Fei Yin, Qiu-Feng Wang, Da-Han Wang, "ICDAR 2011 Chinese Handwriting Recognition Competition", *2011 International Conference on Document Analysis and Recognition*, 2011.
- [11] Michael Murdock, Jack Reese, Shawn Reid, "ICFHR2016 Competition on Local Attribute Detection for Handwriting Recognition", *2016 15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2016.
- [12] Bouchra El qacimy, Mounir Ait kerroum, Ahmed Hammouch, "Handwritten digit recognition based on DCT features and SVM classifier", *2014 Second World Conference on Complex Systems (WCCS)*, 2014.

- [13] Selda Güney, Ceyhun Çakar, "Comparison of expectation maximization and Naïve Bayes algorithms in character recognition", *2016 24th Signal Processing and Communication Application Conference (SIU)*, 2016.
- [14] M. Rajalakshmi, P. Saranya, P. Shanmugavadivu, "Pattern Recognition-Recognition of Handwritten Document Using Convolutional Neural Networks", *2019 IEEE International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, 2019.
- [15] K. Dutta, P. Krishnan, M. Mathew, C.V. Jawahar, "Improving CNN-RNN Hybrid Networks for Handwriting Recognition", *2018 16th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2019.
- [16] Pritam Dhande, Reena Kharat, "Recognition of cursive English handwritten characters", *2017 International Conference on Trends in Electronics and Informatics (ICEI)*, 2017.
- [17] Qiu-Feng Wang, Fei Yin, Cheng-Lin Liu, "Handwritten Chinese Text Recognition by Integrating Multiple Contexts", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 8, pp. 1469 - 1481, 2012.
- [18] Xiang-Dong Zhou, Da-Han Wang, Feng Tian, Cheng-Lin Liu, Masaki Nakagawa, "Handwritten Chinese/Japanese Text Recognition Using Semi-Markov Conditional Random Fields", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, no. 10, pp. 2413 - 2426, 2013.
- [19] De Cao Tran, "An efficient method for online Vietnamese handwritten character recognition", *The Third Symposium on Information and Communication Technology – SoICT*, 2012.
- [20] Thach Tran Van, Phi Nguyen Huu, Trang Hoang, "Isolated vietnamese handwriting recognition embedded system applied combined feature extraction method", *International Conference on Advanced Technologies for Communications (ATC)*, 2015.
- [21] H. T. Nguyen, C. T. Nguyen, P. T. Bao, and M. Nakagawa, "Preparation of an Unconstrained Vietnamese Online Handwriting Database and Recognition Experiments by Recurrent Neural Networks", *15th International Conference on Frontiers in Handwriting Recognition (ICFHR)*, 2016.
- [22] Anh Duc Le, Hung Tuan Nguyen, Masaki Nakagawa, "Recognizing Unconstrained Vietnamese Handwriting By Attention Based Encoder Decoder Model", *International Conference on Advanced Computing and Applications (ACOMP)*, 2018.
- [23] W. & W. Z. Rawat, "Deep Convolutional Neural Networks for Image Classification: A Comprehensive Review", *Journal of Neural Computation*, vol. 29, iss.9, pp.2352-2449, Sep. 2017.
- [24] C. G. S. Lawrence, "Overfitting and neural networks: conjugate gradient and backpropagation", *IEEE-INNS-ENNS International Joint Conference on Neural Networks IJCNN*, 2000.
- [25] G. H. A. K. I. S. R. S. Nitish Srivastava, "Dropout: A Simple Way to Prevent Neural Networks from overfitting", *Journal of Machine Learning Research*, vol.15, 2014.
- [26] J. Quiñero-Candela, M. Sugiyama, A. Schwaighofer, N. D. Lawrence, "Dataset Shift in Machine Learning", *Neural Information Processing series*, The MIT Press, ISBN 978-0-262-17005-5, 2008.

BIOGRAPHIES OF AUTHORS



Dr. Truong Quang Vinh received the B.E. degree from Ho Chi Minh City University of Technology, Vietnam, in 1999, the M.E. degree in Computer Science from the Asian Institute of Technologies, Bangkok, Thailand, in 2003; and Ph.D. degree in Computer Engineering at Chonnam National University, Korea, in 2010. Currently, he is a lecturer in Department of Electronics, Faculty of Electrical and Electronics Engineering, HCM City University of Technology, Vietnam National University – Ho Chi Minh, Vietnam. His homepage: <http://www4.hcmut.edu.vn/~tqvinh/>



Mr. Le Hoai Duy received the B.E. degree in Electronics from Ho Chi Minh City University of Technology, Vietnam in 2018. Now he works as an Artificial Intelligence Engineer at Emage Development Co.Ltd, Vietnam.



Mr. Nguyen Thanh Nhan received the B.E. degree in Electronics from Ho Chi Minh City University of Technology, Vietnam in 2018. Currently, he works as an Vision Engineer at Emage Development Co.Ltd, Vietnam.