Robust and Real-Time Deep Learning System for Checking Student Attendance

Vinh Dinh Nguyen, Khanh Xuan Hong Nguyen Tran, Vu Cong Nguyen, and Narayan C. Debnath
Eastern International University, Nam Ky Khoi Nghia Street, New City, Binh Duong, Vietnam
Email: {vinh.nguyen, khanh.tran.set15, vu.nguyen, narayan.debnath}@eiu.edu.vn

***Abstract*—A face detection and identification algorithm is an interesting research topic. The performance of existing face detection and identification systems works well under normal lighting conditions, while their performance is not stable under difficult conditions due to noise and illumination changes. Therefore, this research aims to develop a robust and real-time deep learning system for student face detection and identification to overcome these current limitations. The proposed method investigates both benefits of state-of-the-art deep learning models and local patterns to create a robust frame-work for detecting and checking student attendance. Comprehensive experimental results show that the proposed method obtained stable results under various normal and difficult indoor conditions. The proposed method obtains the detection rate of 93.55% and 89.25% under normal and difficult indoor conditions, respectively. The proposed method obtains the identification rate of 87.79% and 85.19% under normal and difficult indoor conditions, respectively.**

***Index Terms*—face detection, face recognition, attendance system, deep learning, local binary pattern, multiple features**

I. Introduction

Nowadays, face detection and recognition play an important role in many real-world applications, such as security, a smart home, and a student attendance system. The student attendance systems are often designed to check whether a student comes to a class or not. In general, the student attendance system often includes two main steps: face detection and face recognition. A face detection step is often designed to localize a student's face. A face recognition step is then applied to identify the student's name. However, there are still challenges to accurately detect and identify student information due to vavious illumination changes and occlusion. Recently, Bhrath *et al.* have proposed a method to detect and verify student information by using Haar cascade to detect a student face's location and using a histogram of local binary pattern to identify student information [1]. However, the performance of this system does not work well under various environments because of illumination changes and pose variances.

Therefore, with the motivation from the research [2], this paper proposed a new approach for detecting and identifying student information by using the deep learning approach and local pattern. First, we used the Single-Shot Detector (SSD) along with ResNet [3] to localize a student face. Second, from the results of the face detection step, we then introduce a pre-processing step based-local patterns to enhances input features for training. Third, the deep learning-based Zeiler&Fergus method [4] is then used to extract embedding features (128-d feature vectors) from each student face. Finally, the embedding feature is then input to the proposed deep neural network to identify a student information. Fig. 1 shows the workflow of the proposed system.

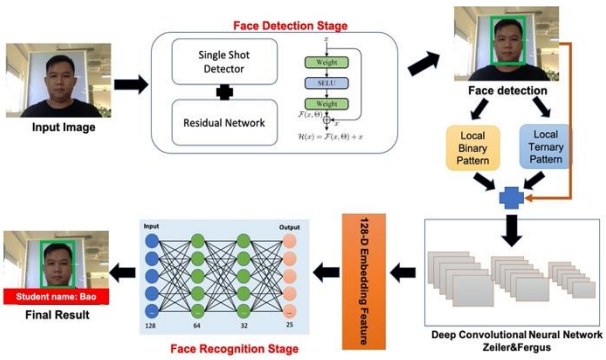

Figure 1. The workflow of the proposed system.

Experimental results show that the proposed method obtained better performance than [1] in terms of accuracy under various indoor conditions. The main contributions of this paper can be listed below:

1) Investigate the benefits of local patterns to enhance the features for training the deep neural networks for face detection and recognition system.
2) Improve the accuracy of existing student attendance based on Haar-cascade and local binary pattern systems by using the deep learning approach.

The paper is organized as follows. We review the existing face detection and recognition systems in Section II. Section III describes the proposed system. Section IV presents experimental results under various indoor environments. Finally, we conclude the paper in Section V.

Manuscript received February 11, 2021; revised August 26, 2021.

doi: 10.12720/jait.12.4.296-301

II. Related Works

There are a large number of face detection and face identification research. However, this paper only reviews the most relevant methods of our proposed method. In general, we can classify the face detection algorithm into two main approaches: feature-based methods and image-based method as discussed in [5].

Three basic techniques, including shape-based models, low level-based features, and advanced-based features, are often used in feature-based approaches. The shape-based models were designed to extract features of human eyes, nose, and mouth by building a statistical facial model. The shape-based model includes three main techniques: snakes [6], point distribution model [7], and deformable templates [8]. For the low level-based approach, most algorithms are designed to detect face by using basic features, such as skin color, motion, gray scale, or edge information. Various color models, including RGB [9], HSV [10], YCbCr [11], and CIELAB [12], have been used to design a face detection system in skin color-based method. Another approach is to localize the human face by considering the motion information between two continuous frames [6]. Analyzing that the gray level inside eyebrows and lip appears darker than their neighbor regions [7], many face detection methods have been developed by considering the local gray minima in segmented regions. In addition, edges features were used to detect a human face by applying a median filter, Sobel operator, and tracking based methods [13]. In recently, many algorithms have been proposed to detect a human face under illumination changes by finding the structural features using Local Binary Pattern, AdaBoost, or Gabor feature-based method [14]-[16].

For the image-based approaches, there are three common techniques, including neural network, linear-based method, and statistical approaches (such as PCA and SVM) as discussed in [6]. The first neural network was designed to detect the front-view of the human face using five layers [17]. Recently, several deep learning-based approaches have been proposed to detect the human face under various poses and occlusions [18], [19].

III. The Proposed Method

Existing student detection and identification systems were not stable under difficult indoor conditions due to illumination changes and occlusions. To overcome these limitations, we proposed a new approach by integrating local patterns information to enhance the quality of input features for the proposed deep learning model. The proposed system is described in Fig. 1. The proposed system consists of four main steps: (1) a face detection step is used to find the location of the face. (2) the detected face is then inputted to the local pattern module to extract stable features. (3) The multiple features-based local patterns is then inputted to the deep convolution neural network learning-based Zeiler&Fergus method to learn an embedded face features. (4) The embedded face features is then used to input to the proposed deep neural network to identify the face information, such as student's name.

A. Face Detection Module

There are many deep learning-based methods for face detection such as Faster RCNN [20], YOLO [21], and SSD [22]. Among the three of them, the SSD is the best because it can satisfy both real-time processing and high accuracy. Here, we aim to use the SSD along with base-net Residual Network [23] to detect student faces. We replace VGG with ResNet-101 to avoid the problem of saturation and decline as mentioned in [24]. Fig. 2 and Fig. 3 show the result of the face detection stage of the proposed method.

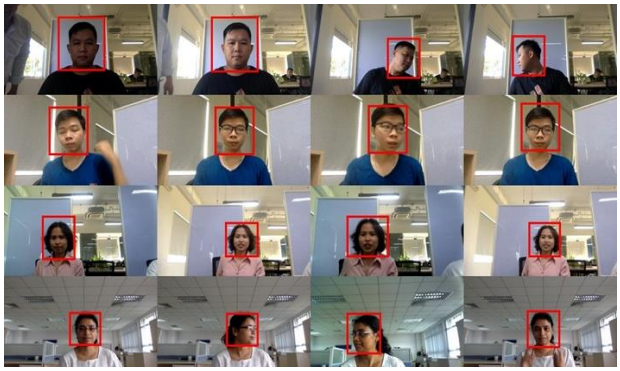

Figure 2. Experimental results of the proposed system under various lighting conditions using EIU dataset.

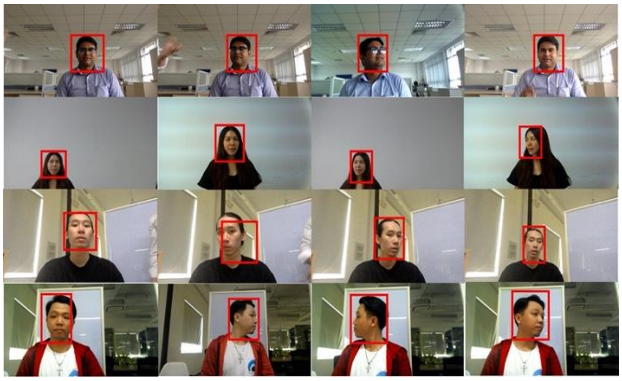

Figure 3. Experimental results of the proposed system under various lighting conditions using EIU dataset.

B. Pre-processing Stage for Face Identification

After localizing the face's location, the next stage is to identify the face information. However, the existing face identification system is easy to affect by noises such as illumination changes.

Recently, the benefits of the local pattern [25] have been demonstrated in various research applications, such as pedestrian recognition, traffic light detection, and recognition. LBP is a feature-based method that was developed to extract the features of the local region by considering the information of neighbor pixels. Given an input image *I*, and a central pixel $p_c = (x, y)$ in the image, the local binary pattern is defined as follows:

$$p_c = (x, y)$$
$$\mathrm{B}_{N,R}^{LBP}(p_c) = \sum_{n=1}^{N} t(p_n, p_c) \times 2^n \tag{1}$$
$$t(p_n, p_c) = \begin{cases} 1, & I(p_n) \geq I(p_c) \\ 0, & else \end{cases}$$

where $I(p_c)$ and $I(p_n)$ are the gray values of the central pixel and the n-th neighbor pixel, respectively. R is the radius of the neighborhood. Moreover, N neighborhood pixels can be located outside of the image grids. Their gray values can be calculated using bilinear interpolation, and the coordinate of p_n is determined by

$$(x_n, y_n) = \left[x_c + R\cos\left(\frac{2\pi n}{N}\right), y_c - R\sin\left(\frac{2\pi n}{N}\right) \right] \tag{2}$$

Fig. 4 shows the results of the local binary pattern on three image channels.

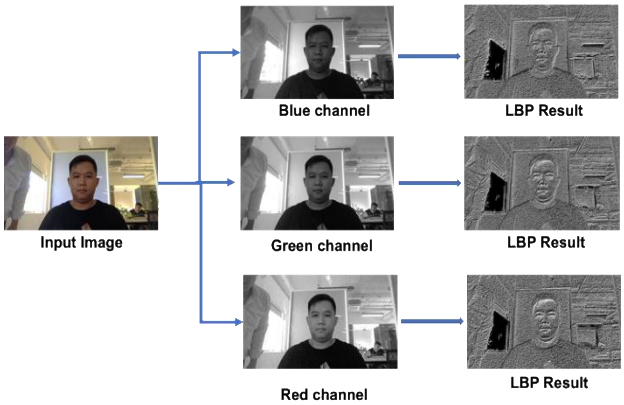

Figure 4. Results of a local binary pattern on three channels of the input image.

The Local Ternary Pattern (LTP) [26] is proposed in order to decrease noise dependence, especially near the uniform region. LTP produces output values with a positive part ($\mathrm{B}_{N,R}^{LTP_{positive}}$) and a negative part ($\mathrm{B}_{N,R}^{LTP_{negative}}$) by using a pre-defined threshold, as follows:

$$\mathrm{B}_{N,R}^{LTP_{positive}}(p_c, \lambda) = \sum_{n=1}^{N} \eta_{positive}(p_n, p_c, \lambda) \times 2^n$$
$$\eta_{positive}(p_n, p_c, \lambda) = \begin{cases} 1, & I(p_n) > I(p_c) + \lambda \\ 0, & I(p_n) > I(p_c) - \lambda \ \ \& \ \ I(p_n) < I(p_c) + \lambda \end{cases} \tag{3}$$
$$\mathrm{B}_{N,R}^{LTP_{negative}}(p_c, \lambda) = \sum_{n=1}^{N} \eta_{negative}(p_n, p_c, \lambda) \times 2^n$$
$$\eta_{negative}(p_n, p_c, \lambda) = \begin{cases} 1, & I(p_n) < I(p_c) - \lambda \\ 0, & I(p_n) > I(p_c) - \lambda \ \ \& \ \ I(p_n) < I(p_c) + \lambda \end{cases}$$

where $\mathrm{B}_{N,R}^{LTP_{positive}}$ and $\mathrm{B}_{N,R}^{LTP_{negative}}$ are the encoding results using positive and negative operators, respectively. Fig. 5 shows the results of the local ternary pattern with negative and positive operators on three image channels.

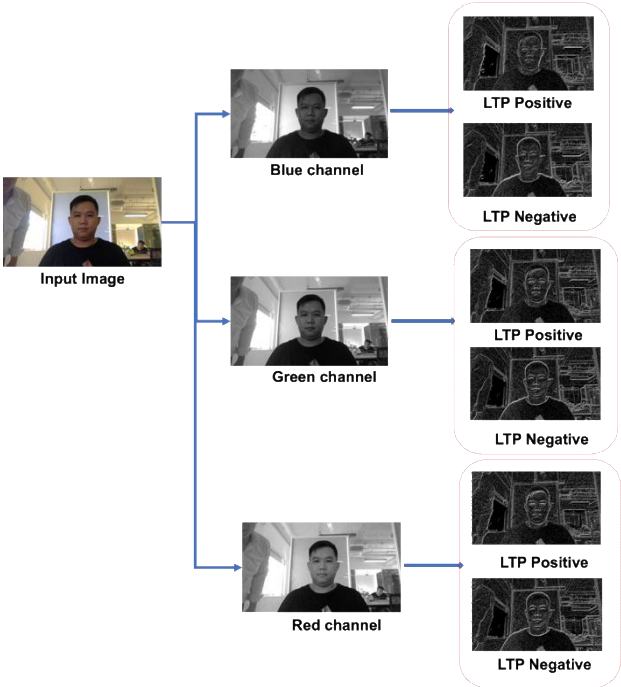

Figure 5. Results of a local ternary pattern on three channels of the input image.

Motivated by the results [27], where the multiple local patterns have been investigated to improve the performance of the vehicle detection system. Therefore, this research investigates the benefit of multiple local patterns to produce a robust feature for the face identification step by using local patterns, local binary pattern, and local ternary pattern, on three image channels as follows:

$$\mathrm{M}_{N,R}^{c}(p) = \alpha_B \times \frac{\mathrm{B}_{N,R}^{LBP,c}(p) + \mathrm{B}_{N,R}^{LTP_{positive}}(p, \lambda) + \mathrm{B}_{N,R}^{LTP_{negative}}(p, \lambda)}{3} + \alpha_R \times I_{N,R}^{c}(p)$$
$$\alpha_B + \alpha_R = 1 \tag{4}$$

where $\mathrm{M}_{N,R}^{c}(p)$ is the encoding result of the proposed system by fusing the benefits of local binary pattern and local ternary pattern. α_B and α_R are thresholds to determine how much contribution of the local patterns and the raw input to the final model.

After obtaining the encoder results by using the proposed system, the embedding feature-based deep learning FaceNet [2] was used to produce 128-d embedding features as shown in Fig. 1.

C. Face Identification

From the results of embedding features, this research proposed a deep neural network to classify the student face as described in Fig. 6.

The proposed network consists of three layers including, 128 units for an input layer, 64 units for a hidden layer, 32 units for a hidden layer, and 25 units for an output layer.

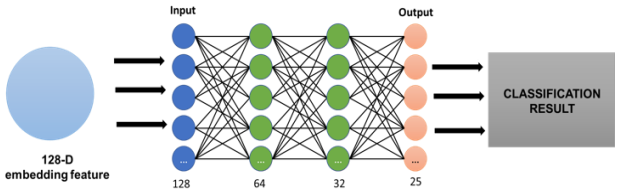

Figure 6. The proposed model for face identification using deep neural network architecture.

IV. Experimetnal Results

A. Dataset and System Configuration

To prepare a dataset for training and testing, we set up a system to capture students' images at Eastern International University by using two USB cameras HD720P Logitech C525 as shown in Fig. 7. The proposed system was used to capture 1000 images for 25 students (40 images for each student). Fig. 2 and Fig. 3 show our dataset that was captured under various indoor conditions. To evaluate the processing time of the proposed system, we implemented it on a computer equipped with Intel Core i5 CPUs, 6GB of RAM, and a GTX 1060 GPU.

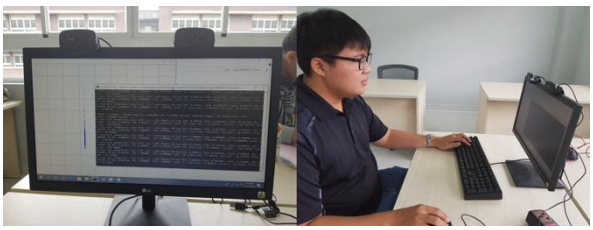

Figure 7. A configuration of the proposed system

B. Evaluation

To evaluate the performance of the proposed system, we use the detection rate (D_{ratio}) to measure the accuracy of the face detection step as follows:

$$D_{ratio} = \frac{T_d}{O} \tag{5}$$

where T_d is the number of correct detections by satisfying the Intersection over Union (*IOU*). This research sets *IOU* to 0.7. *O* is the total number of testing images.

We use the following metric to evaluate the identification rate (I_{ratio}):

$$I_{ratio} = \frac{N_c}{Z} \tag{6}$$

where N_c is the number of correct identifications, and *Z* is the total number of evaluating images.

C. Results and Discussion

We compared our proposed method with the most related method [1]. The dataset was divided into two categories that include normal indoor conditions and illumination changes indoor conditions. We use 500 images for training, 250 images for cross-validation, and 250 images for testing from 1000 captured images.

To verify how the local pattern feature affects the performance of the proposed system, we conducted an experiment to evaluate the detection rate under normal and difficult indoor conditions as shown in Fig. 8. Experimental result show that the proposed method obtained better accuracy than feature-based method [1]. The detection rate of the proposed method with local pattern, the proposed method without a local pattern, and the feature-based method [1] are 93.55 %, 90.15%, and 88.7% under normal lighting conditions, respectively. The detection rate of the proposed method with a local pattern, the proposed method without a local pattern, and the feature-based method [1] are 89.25 %, 87.45 %, and 85.27% under difficult lighting conditions, respectively. Thus, the local pattern can improve the performance of the proposed method under both normal and difficult indoor conditions.

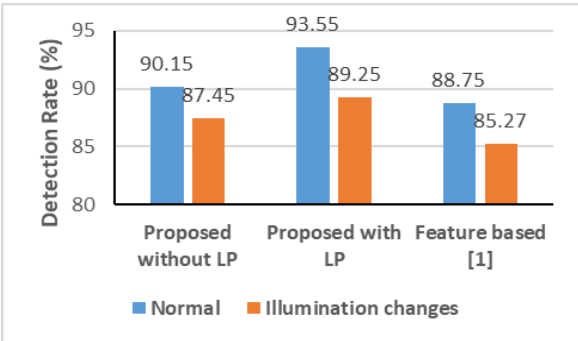

Figure 8. The detection rate of the proposed method without using local patterns, the proposed method with using local patterns, and the feature-based method [1] under various indoor conditions.

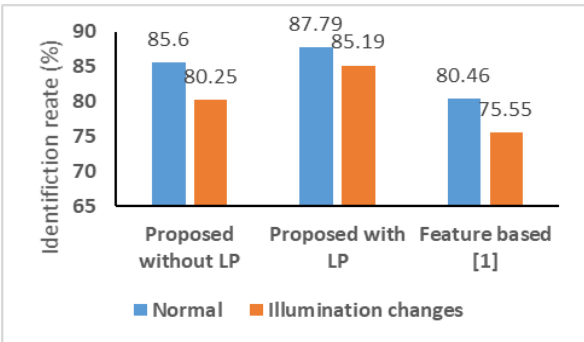

Figure 9. Identification rate of the proposed method without using local patterns, the proposed method with using local patterns, and the feature based-method [1] under various indoor conditions.

To evaluate the performance of the proposed system in terms of identification rate, we conducted more experiments on the testing dataset as shown in Fig. 9. Experimental results show that the proposed method obtained a better identification rate than the feature-based method [1]. The identification rate of the proposed method with a local pattern, the proposed method without a local pattern, and the feature-based method [1] are 87.79 %, 85.60%, and 80.46% under normal lighting conditions, respectively. The detection rate of the proposed method with a local pattern, the proposed method without a local pattern, and the feature-based method [1] are 85.19%, 80.50%, and 75.55% under

difficult lighting conditions, respectively. Thus, the local pattern can improve the performance of the proposed method in terms of identification rate under both normal and difficult indoor conditions. The proposed method obtained real-time processing by using a 1280x720 image resolution.

V. Conclusion

Face dectection and identification systems play a very importance to many real-world applications. However, the existing face detection system does not work well under difficult conditions. Therefore, this research investigates a simple and efficient framework to detect and identify the face (student) information by using the benefit of deep learning neural network and local-based features. Experimental results yield that the proposed system does stable results under both normal and difficult indoor conditions. However, the performance of the proposed system is decreased under outdoor conditions where many unknow-noise factors might appear. Therefore, we plan to investigate a robust end-to-end deep learning-based system to overcome this limitation in the future.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Khanh Xuan Nguyen Tran collected and analyzed the data. Vu Cong Nguyen conducted experiments on the face detection. Vinh Dinh Nguyen designed and developed a deep learning model for face identification. Narayan C. Debnath wrote and corrected the grammar of the manuscript. All authors had approved the final version.

Acknowledgment

This research was supported by Artificial Intelligence (AI) Lab, Eastern International University, Binh Duong, Vietnam.

References

[1] B. T. Chinimilli, T. Anjali, A. Kotturi, V. R. Kaipu, and J. V. Mandapati, "Face recognition based attendance system using haar cascade and local binary pattern histogram algorithm," in *Proc. International Conference on Trends in Electronics and Informatics*, 2020, pp. 701-704.

[2] F. Schroff, D. Kalenicheko, and J. Phibin, "FaceNet: A unified embedding for face recognition and clustering," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2015.

[3] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, 2016.

[4] M. D. Zeiler and R. Fergus, "Visualizing and understanding convolutional networks," in *Proc. European Conference on Computer Vision*, 2014.

[5] A. Kumar, A. Kau, and M. Kumar, "Face detection techniques: A review," *Artificial Intelligence Review*, vol. 52, pp. 927-948, 2019

[6] H. Erik and B. K. Low, "Face detection: A survey," *Computer Vision Image Understanding*, vol. 83, pp. 236-274, 2001.

[7] T. F. Cootes, D. H. Cooper, C. J. Taylor, and J. Graham, "Trainable method of parametric shape description," *Image and Vision Computing*, vol. 10, no. 5, pp. 289-294, 1992.

[8] A. L. Yuille, P. W. Hallinan, and D. S. Cohen, "Feature extraction from faces using deformable templates," *Int. J. Comput. Vis.*, vol. 8, pp. 99-111, 1992.

[9] B, Dhivakar, C. Sridevi, S. Selvakumar, and P. Guhan, "Face detection and recognition using skin color," in *Proc. IEEE 3rd International Conference on Signal Processing, Communication and Networking*, 2015, pp. 1-7.

[10] H. F. Hashem, "Adaptive technique for human face detection using HSV color space and neural networks," in *Proc. 26th National Radio Science Conference*, 2009, pp. 1-7.

[11] X. Zhu, D. Ren, Z. Jing, L. Yan, and S. Lei, "Comparative research of the common face detection methods," in *Proc. 2nd International Conference on Computer Science and Network Technology*, 2012, pp. 1528-1533.

[12] L. Zou and S. Kamata, "Face detection in color images based on skin color models," in *Proc. IEEE Region 10 Conferences*, 2010, pp. 681-686.

[13] S. Anila and N. Devarajan, "Simple and fast face detection system based on edges," *International Journal Universal Computer Science*, vol. 2, no. 1, pp. 54-58, 2010.

[14] H. Zhang, Y. Xie, and C. Xu, "A classifier training method for face detection based on AdaBoost," in *Proc. International Conference on Transportation, Mechanical, and Electrical Engineering*, 2011, pp. 731-734.

[15] H. Filali, J. Riffi, A. M. Mahraz, and H. Tairi, "Multiple face detection based on machine learning," in *Proc. International Conference on Intelligent Systems and Computer Vision*, 2018, pp. 1-8.

[16] M. Sharif, A. Khalid, M. Raza, and S. Mohsin, "Face recognition using Gabor filters," *Journal of Applied Computer Science and Mathematics*, vol. 11, no. 5, pp. 53-57, 2011.

[17] M. A. Kramer, "Nonlinear principal component analysis using auto associative neural networks," *Am. Inst. Chem. Eng. J.*, vol. 37, no. 2, pp. 233-243, 1991.

[18] S. S. Farfade, M. Saberian, and L. J. Li. "Multi-view face detection using deep convolutional neural networks," in *Proc. the International Conference on Multimedia Retrieval*, 2015, pp. 1-8.

[19] S. Liao, A. K. Jain, and S. Z. Li, "A fast and accurate unconstrained face detector," *IEEE Trans. Pattern. Anal. Mach. Intell.*, vol. 38, no. 2, pp. 211-223, 2016.

[20] S. Ren, K. He, and R. Girshick, "Faster R-CNN: Towards real-time object detection with region proposal networks," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 39, no. 6, pp. 1137-1149, 2017.

[21] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, 2016.

[22] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C. Y. Fu, and A. C. Berg, "SSD: Single shot MultiBox detector," in *Proc. European Conference on Computer Vision*, 2016, pp. 21-37.

[23] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. Conference on Computer Vision and Pattern Recognition*, 2016.

[24] X. Lu, X. Kang, S. Nishide, and F. Ren, "Object detection based on SSD-ResNet," in *Proc. International Conference on Cloud Computing and Intelligence Systems*, 2019.

[25] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 24, no. 7, pp. 971-987, July 2002.

[26] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," in *Proc. International workshop on Analysis and Modeling of Faces and Gestures*, 2007, vol. 4778, pp. 168-182.

[27] V. D. Nguyen, D. T. Tran, J. Y. Byun, and J. W. Jeon, "Real-Time vehicle detection using an effective region proposal-based depth and 3-channel pattern," *IEEE Transactions on Intelligent Transportation Systems*, vol. 20, no. 10, pp. 3634-3646, 2019.

Vinh Dinh Nguyen received the B.Sc. (*magna cum laude*) degree in computer science from Nong Lam University, Ho Chi Minh City, Vietnam, in 2007, and the M.S. and Ph.D. degrees in electrical and computer engineering from Sungkyunkwan University, Suwon, South Korea, in 2012 and 2015, respectively. From 2015 to 2019 he was a research professor at College of Information and Computer Engineering, Sungkyunkwan University, South Korea. Since 2019, he has been with the Department of Software Engineering, School of Computing and Information Technology, Eastern International University, Vietnam as a Lecturer and Researcher. His research interests include deep learning, computer vision, image processing, and graphics processing unit computing.

Vu Cong Nguyen received the B.Sc. degree in mathematics and computer science from the University of Science, Ho Chi Minh City, Vietnam, in 1991, and a Ph.D. degree in computational chemistry from Innsbruck University, Austria in 1996. Since 2014 he has been with the Department of Software Engineering, School of Computing and Information Technology, Eastern International University, Vietnam as a lecturer and researcher. His research interests include software engineering, data science, machine learning, and computer vision.

Khanh Xuan Nguyen Tran is currently finishing his B.Sc. degree in software engineering at the Department of Software Engineering, School of Computing and Information Technology, Eastern International University, Vietnam. His interests include software engineering, machine learning, and deep learning.

Narayan C. Debnath is currently the Founding Dean of the School of Computing and Information Technology at Eastern International University, Vietnam. He is also serving as the Head of the Department of Software Engineering at Eastern International University, Vietnam. Dr. Debnath has been the Director of the International Society for Computers and their Applications (ISCA), USA since 2014. Formerly, Dr. Debnath served as a Full Professor of Computer Science at Winona State University, Minnesota, USA for 28 years (1989-2017). He was elected as the Chairperson of the Computer Science Department at Winona State University for 3 consecutive terms and assumed the role of the Chairperson of the Computer Science Department at Winona State University for 7 years (2010-2017).