Lightweight Models for Face Mask Detection And Face Recognition

Tien Ho-Phuoc, Hanh T. M. Tran*, Duc Hoang Xuan, Tuan Dao-Duy, Hoang Le Uyen Thuc

Abstract—The Covid-19 pandemic has abruptly changed our daily life: face mask and wearing a mask have become popular, and are occasionally obligatory for some public activities. As a result, several methods were proposed to address the problem of face mask detection and/or face recognition. While these methods showed promising results, they often require high computational resource due to sophisticated deep learning models. In this paper, we will propose lightweight methods to detect face masks and recognize human faces simultaneously, in the perspective of implementing them in real-time applications. Our proposed methods are based mainly on Random Forest and MobileNetV2, and are trained and tested with our own dataset collected from Vietnamese faces. The experiment shows that Random Forest can address face mask detection and face recognition with high accuracy. Moreover, a combination of Random Forest and MobileNetV2 can still improve the performance of detection and recognition while keeping the method at relatively low computational complexity.

Index Terms—Face mask detection; Face recognition; MobileNetV2 model; Random Forest Classifier.

1. Introduction

FACE recognition, which is one branch of biometric identification, is mostly used recently to recognize an individual's identity [1]. Unlike other biometric identification techniques such as password, ID cards or fingerprint, face recognition does not require direct touch on the sensors. Therefore, face recognition can be safely used during the global COVID-19 pandemic crisis. However, mandatory mask-wearing policy in public places as a method to slow the virus transmission [2], has yielded certain difficulties in recognizing faces, e.g., the masked face data is insufficient, the grand part of the face including the nose is occluded, the color and the texture of masks are diversity, etc.

In this paper, we perform two tasks related to face recognition. The first one is to check whether a person is wearing a mask. This application aims to assist the organizations in automatically monitoring the implementation of mask-wearing policy. The second one is to enlarge the existing face recognition systems to recognize a human face regardless of whether they are wearing a mask or not.

Since the Covid-19 pandemic, the two tasks above have become widely popular topics and have been well studied in the literature. While the problem of detection and recognition was generally addressed by the traditional machine learning approach, recently the deep learning based approach has helped to solve this problem more effectively [3]–[8].

Tien Ho-Phuoc, Hanh T. M. Tran, Duc Hoang Xuan, Tuan Dao-Duy and Hoang Le Uyen Thuc are with the University of Danang - University of Science and Technology, Danang, Vietnam.

**Corresponding author: Hanh T. M. Tran (e-mail: hanhtran@dut.udn.vn)*
Manuscript received May 04, 2022; revised May 27, 2022; accepted June 22, 2022.
Digital Object Identifier 10.31130/ud-jst.2022.243ICT

In [3], the authors propose a system to detect whether a person uses a face mask. It can receive real-time video as input and, hence, may be used in a smart city to identify persons not wearing a mask and inform the relevant authority in order to limit the propagation of the virus in the context of the Covid-19 pandemic. Chavda et al. [4] exploit a two-staged CNN architecture for face mask detection. In the first stage, the model extracts a human face; and in the second one, it determines whether this face is masked or not. Interestingly, the unmasked class also contains masks that are not properly used and the situation in which a person puts a hand on his or her face.

Based on the high object detection capability of the YOLOv3 model, the authors in [5] propose a system to detect and localize a human face with or without a mask. This system can also process video in real-time and provide promising detection accuracy for the purpose of human safety during the pandemic.

Another line of human face related research that has emerged actively since the Covid-19 crisis is masked face recognition. This kind of face recognition is challenging as an important part of a face is missing, and we need to effectively use the remaining information to correctly recognize a person. Hariri [6] addresses this problem based on occlusion removal and deep learning-based features. In particular, the method tries in the first step to remove the masked face region and, then, use CNNs, for example VGG or ResNet, to extract features from the obtained regions that are mainly eyes and forehead. Finally, a classic MLP (Multilayer Perceptron) is used to classify persons.

Different from the above classification approach [6], Deng et al. [7] addresses the masked face recognition by comparing a pair of face images to determine whether the two faces are from the same person. To obtain this

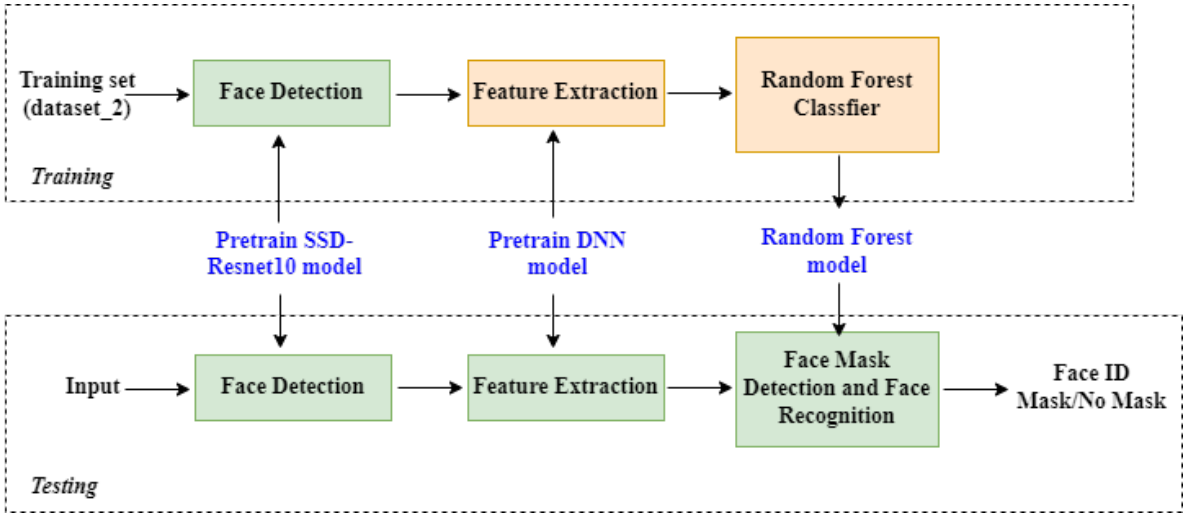

Fig. 1: Face Mask Detection and Face Recognition using Random Forest Classifier.

purpose, the authors propose to use the Large Margin Cosine Loss, which is expected to be able to extract key features of masked faces. Moreover, for better feature extraction the method also uses an attention module in order to focus more on the area that is not covered by the mask.

Lightweight masked face recognition models are highly expected as they may be very useful for recognition applications in public areas. In [8], the authors compare different lightweight CNNs for this problem. Specifically, by evaluating three CNNs – MobilenetV2, DenseNet, and NASNetMobile –, they showed that MobilenetV2 is slightly better than the two others.

In this paper, we propose methods to address face mask detection and face recognition simultaneously by using MobileNetV2 and Random Forest since a combination of these two tasks may promise interesting practical applications. Moreover, in order to satisfy such applications, which are often real-time, our methods have low computational complexity. Random Forest belongs to the traditional machine learning approach and is very effective for classification. Meanwhile, MobilenetV2 exploits high representation and recognition capability of deep learning, but has relatively low resource requirement. Our methods are evaluated on datasets that we capture using laptop and smart-phone camera with the slightly change (about 30 degree left, right, up and down) of face pose from the neutral pose. The evaluation results show that our combination of MobileNetV2 and Random Forest improves the detection and recognition accuracy.

In summary, the contributions of the paper are as follows:

- We build models that are capable of detecting human face masks and recognizing faces simultaneously.
- The paper aims at lightweight models, which can be used for real-time applications.
- The combination of MobileNetV2 and Random Forest can improve the performance of detection and recognition.

2. Methods

2.1. "Random Forest" - Face mask detection and face recognition based on Random Forest

Random Forest constructs a number of decision trees at training time and can be used for regression or classification [9]. For regression, the output of Random Forest is the average prediction of all individual decision trees. For classification, Random Forest provides the output that is the class selected by a majority of trees. Hence, Random Forest can be considered as an ensemble learning method and helps to reduce overfitting.

In Fig. 1, we propose a method for face mask detection and face recognition based on Random Forest. In this method, a pre-trained SSD model [10] combined with Resnet10 is used to detect faces in a frame. The feature of each face is then extracted by the DNN model [11], which is pre-trained with the FaceScrub [12] and CASIA-WebFace [13] datasets. Consequently, each face is represented by a 128 D embedding vector. Such vectors and corresponding labels are used to train the Random Forest (the upper part of Fig. 1). In the test phase (the lower part of Fig. 1), an image input also goes through the SSD-Resnet10 and DNN models to generate a 128D embedding vector, which is then classified by the trained Random Forest.

It is important to note that in this method, the Random Forest recognizes a person's identifier and, at the same time, determines whether this person wears a mask or not. This is possible thanks to the fact that the dataset used for Random Forest's training considers a person wearing mask and this same person but not wearing mask as two distinct classes. Such dataset will be described in details in the experiment (dataset_2).

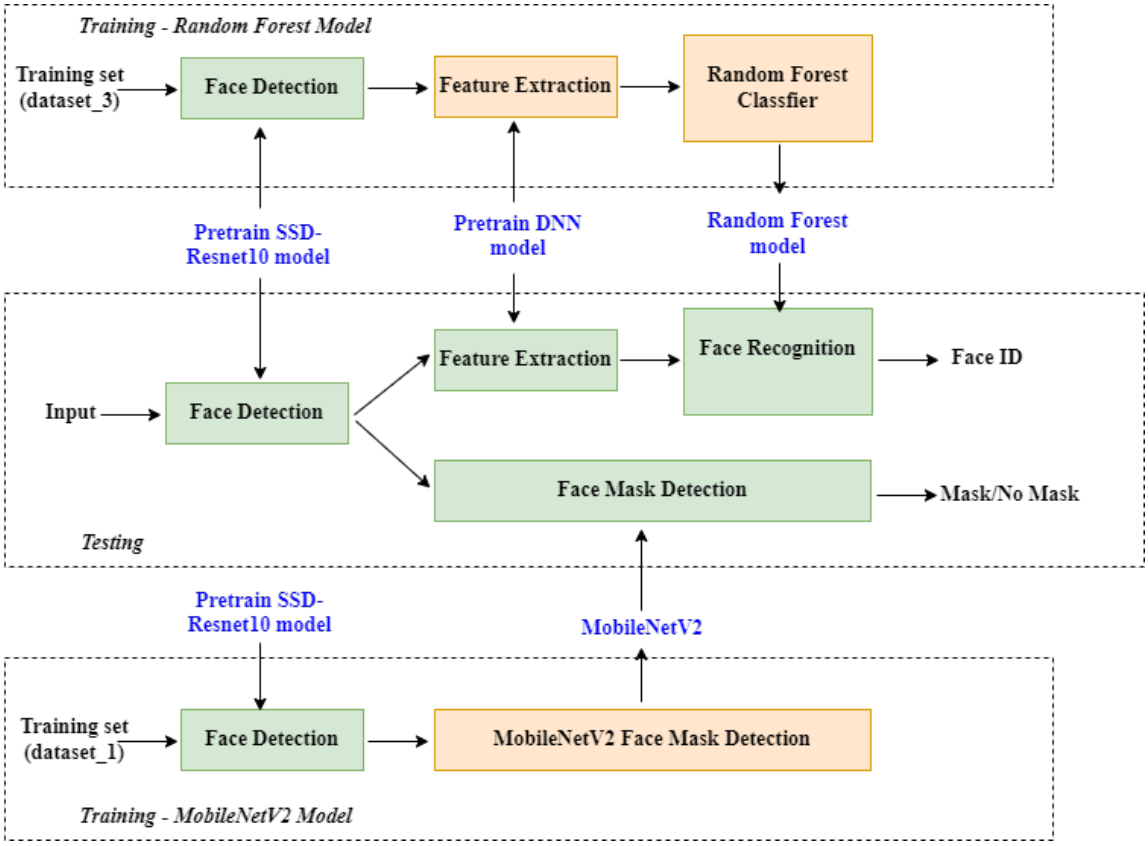

Fig. 2: Face Mask Detection and Face Recognition using MobileNetV2 and Random Forest Classifier.

2.2. MobileNetV2

MobileNetV2 is an efficient architecture aiming at low-resource devices: it requires less operations and memory while retaining high accuracy [14]. MobileNetV2 has been widely used in a large number of deep learning based models for various applications [14]–[16]. The particularity of MobileNetV2 is the inverted residual with linear bottleneck. The input and output of this block have low dimension and do not contain non-linearities. The intermediate layer of the block expands features to high dimension and this non-linear representation is carried out by a lightweight depthwise convolution. This special convolution applies a single convolutional filter per input channel, followed by a 1x1 convolution (or pointwise convolution) to generate new features by linear combinations of channels [14]. Table 1 summarizes the MobileNetV2 architecture: "t" represents the expansion factor, "c" the number of output channels, "n" the number of times a block to be repeated, and "s" the stride.

In this paper, we use MobileNetV2 for face mask detection. The input of MobileNetV2 is the face detected by the pre-trained SSD-ResNet10 as in Fig. 1. The training and test phases of MobileNetV2 are described in the bottom and middle of Fig. 2.

In the next subsection, we present a combination of MobileNetV2 and Random Forest in order to improve accuracy of face mask detection and face recognition.

TABLE 1: MobileNetV2 architecture [14]

Input	Layer/Block	t	c	n	s
$224^2 \times 3$	Conv2d	–	32	1	2
$112^2 \times 32$	Bottleneck	1	16	1	1
$112^2 \times 16$	Bottleneck	6	24	2	2
$56^2 \times 24$	Bottleneck	6	32	3	2
$28^2 \times 32$	Bottleneck	6	64	4	2
$14^2 \times 64$	Bottleneck	6	96	4	1
$14^2 \times 96$	Bottleneck	6	160	3	2
$7^2 \times 160$	Bottleneck	6	320	1	1
$7^2 \times 320$	Conv2d 1x1	–	1280	1	1
$7^2 \times 1280$	Avgpool 7x7	–	1	–	–
$1 \times 1 \times 1280$	Conv2d 1x1	–	k	–	–

2.3. "MobileNetV2 + Random Forest" - Face mask detection and face recognition based on MobileNetV2 and Random Forest

In this method (Fig. 2), face mask detection and face recognition are carried out by two distinct modules. Concretely, MobileNetV2 is responsible for face mask detection while Random Forest is used to recognize a person's identifier. The functioning of the Random Forest in Fig. 2 is quite similar to that of the Random Forest in Fig. 1. However, there is an important

difference: the Random Forest in Fig. 2 only recognizes a person's identifier and does not determine whether this person wears a mask or not. In other words, a person with mask and this same person without mask are considered as one class or represented by the same identifier. Once again, this is due to the dataset used for training the Random Forest in Fig. 2. We will describe this dataset in details in the experiment.

Looking at the middle part of Fig. 2, we can also see the difference between the traditional machine learning approach and the deep learning based approach. For the traditional approach (the upper branch), a module of feature extraction (carried out by pre-trained DNN) is separated from a classifier or recognizer (carried out by Random Forest). Meanwhile, the deep learning based MobileNetV2 module (the lower branch) combines feature extraction and classification/recognition. The method presented in this subsection tries to obtain a good compromise between accuracy and low complexity by combining the two approaches.

3. Experiments and Results

3.1. Datasets

In this paper, models are trained and evaluated using three datasets that the authors captured ourselves using smart-phone and laptop camera. The details of these datasets are as follows:

- Dataset 1: is used to train and test MobileNetV2 for face mask detection (Fig. 2). The dataset includes two classes: Mask and No Mask, in which there are in total 975 images for Mask class and 1,000 images for No Mask class. We collected these images using smart-phone and laptop camera. In order to avoid over-fitting, we increase the size of training set by data augmentation. We implemented random rotation, random shift, zoom with nearest neighbour interpolation method, shear and horizontal flip. In order to build the face mask dataset, we used six types of face masks with four colors: white, black, light blue and light grey.
- Dataset 2: is used to train and test Random Forest Classifier for face mask detection and face recognition (Fig. 1). All 1975 images in Dataset 1 are used to form Dataset 2 which includes 20 classes of 10 people with two classes (Mask and No Mask) each person.
- Dataset 3: is used to train Random Forest Classifier and test the face mask detection and face recognition system combining MobileNetV2 and Random Forest (Fig. 2). We form this dataset by combining two classes in Dataset 2 into one class. Therefore this dataset includes 10 classes of 10 people with the same total number of images as the Dataset 2. Fig. 3 shows images in two classes in which each class contains face mask images and no face mask images.

When using these three datasets, we split each dataset into training set and test set with the ratio of 8:2.

3.2. Results

3.2.1. MobileNetV2 for face mask detection

We removed the last layer of pretrained MobileNetV2 model on ImageNet dataset [17], then added four layers on the top: Average Pooling layer, Fully connected layer, Dropout layer and Fully connected layer. The parameters of pretrained MobileNet model were frozen and the last four layers was trained on Dataset 1 by minimizing the cross-entropy cost function [18] as in Eq. 1 with Adam optimizer [19]:

$$L(y,\hat{y}) = -\frac{1}{n}\sum_{i=1}^{n}[y_i.log(\hat{y_i}) + (1-y_i).log(1-\hat{y_i})] \quad (1)$$

where n is the total number of images in training set. $\hat{y_i}$ is an output of the model and y_i is a corresponding ground truth.

In order to evaluate the performance of the model, we also used the accuracy represented as in Eq. 2.

$$acc = \frac{TP+TN}{TP+FP+TN+FN} \quad (2)$$

where TP is the total number of faces with masks that are correctly detected, TN is the total number of faces without masks that are correctly detected, FP is the total number of faces without mask in the images that are incorrectly predicted to faces with masks and FN is the total number of faces with masks that are incorrectly predicted to faces without masks.

The training, testing accuracy and loss of the MobileNetv2 model on the Dataset 1 are shown in Fig. 4. The accuracy increases significantly after four epochs.

3.2.2. MobileNetV2 and Random Forest for face mask detection and face recognition

Table 2 shows the comparison between the two methods: Random Forest (RF), the combination of MobileNetV2 and RF used for face mask detection (FMD) and face recognition (FR). The table shows that the combination of MobileNetV2 and Random Forest (MobileNetV2 + RF) improves the face mask detection (FMD) accuracy and face recognition (FR) accuracy. The FMD + FR column in this table considers both mask detection and face ID recognition accuracy simultaneously. In these methods, we use Random Forest with 50 trees.

TABLE 2: Comparison of the models accuracy.

Method	Accuracy (%)		
	FMD	FR	FMD + FR
Random Forest (RF)	98.7	94.5	93.7
MobileNetV2 + RF	100	95	95

Fig. 5 shows the detection and recognition results using MobileNetV2 and Random Forest. The qualitative results show that the system can correctly detect and recognize multiple faces with or without mask in the images.

Fig. 3: An example of two classes in Dataset 3 in which each class contains images of both mask and no mask faces.

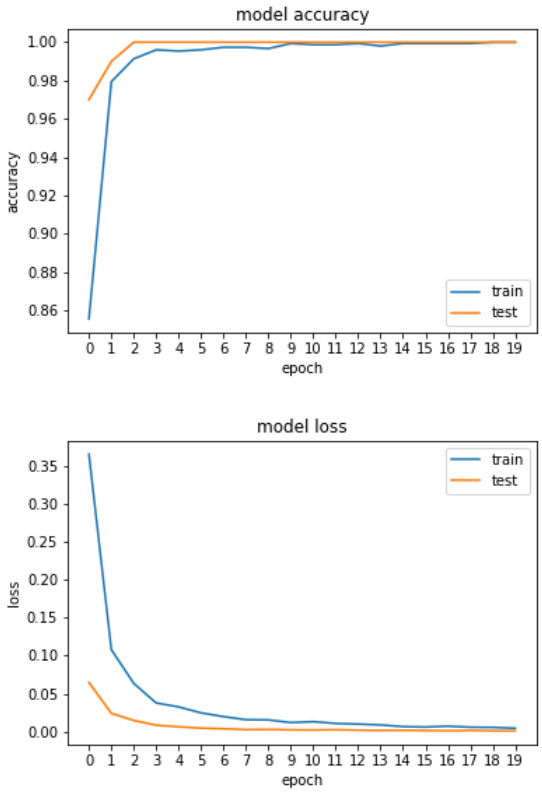

Fig. 4: Training, testing accuracy and loss of MobileNetV2 on Dataset 1.

Fig. 5: Face mask detection and face recognition qualitative results of the combination method.

4. Conclusion

In this work, we propose to combine MobileNetV2 model for face mask detection and Random Forest for face recognition. The combination of two models improves the accuracy of both detection and recognition. Moreover, the result show that the method can detect and recognize multiple faces with or without mask. However, our method has not been evaluated on CCTV videos with many human faces in far distances. In the future, we will improve the method to detect and recognize human faces with mask in the CCTV videos.

Acknowledgment

We would like to thank Mr. Hai Ho Van, the University of Danang - University of Science and Technology, for his help on the dataset used in this paper.

References

[1] Juneja K. and Rana C., "An Extensive Study on Traditional-to-Recent Transformation on Face Recognition System", *Wireless Pers Commun* 118, 3075–3128, 2021.

[2] Cheng, Yafang, et al. "Face Masks Effectively Limit the Probability of SARS-CoV-2 Transmission", *Science*, vol. 372, no. 6549, 20 May 2021.

[3] M. Rahman, M. Manik, M. Islam, S. Mahmud, and J. Kim, "An Automated System to Limit COVID-19 Using Facial Mask Detection in Smart City Network", *IEEE International IOT, Electronics and Mechatronics Conference (IEMTRONICS)*, 2020.

[4] A. Chavda, J. Dsouza, S. Badgujar, and A. Damani, "Multi-Stage CNN Architecture for Face Mask Detection", *International Conference for Convergence in Technology (I2CT)*, 2021.

[5] M. Bhuiyan, S. Khushbu, and M. Islam, "A Deep Learning Based Assistive System to Classify COVID-19 Face Mask for Human Safety with YOLOv3", *International Conference on Computing, Communication and Networking Technologies (ICCCNT)*, 2020.

[6] W. Hariri, "Efficient masked face recognition method during the COVID-19 pandemic", *Signal, Image and Video Processing* 16, 605–612, 2022.

[7] H. Deng, Z. Feng, G. Qian, X. Lv, H. Li, and G. Li, "MFCosface: A Masked-Face Recognition Algorithm Based on Large Margin Cosine Loss", *Applied Sciences*. 2021; 11(16):7310.

[8] A. Alawi and A. Qasem, "Lightweight CNN-based Models for Masked Face Recognition", *International Congress of Advanced Technology and Engineering (ICOTEN)*, 2021

[9] L. Breiman, "Random Forest", *Machine Learning* 45, 5-32,2001

[10] L. Wei, D. Anguelov, D. Erhan, C., S. Reed, C. Fu, and A. Berg, "Ssd: Single shot multibox detector," *European conference on computer vision*, pp. 21-37, Springer, 2016.

[11] OpenFace, [online] Available: https://cmusatyalab.github.io/openface/training-new-models/
[12] Facescrub database, [online] Available: Vintage.winklerbros.net/facescrub.html
[13] D. Yi, Z. Lei, S. Liao, and S. Li, "Learning face representation from scratch," *arXiv:1411.7923*, 2014.
[14] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. Chen, "Mobilenetv2: Inverted residuals and linear bottlenecks," *IEEE conference on computer vision and pattern recognition*, pp. 4510-4520, 2018.
[15] C.Chieh, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," *arXiv:1706.05587*, 2017.
[16] M. Francesco and A. Santone, "Transfer learning for mobile real-time face mask detection and localization," *Journal of the American Medical Informatics Association* 28, no. 7, 2021, 1548-1554.
[17] [online] Available: https://www.image-net.org/
[18] Murphy, Kevin P., *Machine learning: a probabilistic perspective*, MIT press, 2012.
[19] Diederik Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv*:1412.6980, 2014.

Tien Ho-Phuoc received his Engineer degree in Telecommunications from Ho Chi Minh City University of Technology in 2004; his Master and PhD degrees in Signal, Image, Speech and Telecommunications from University of Grenoble in 2006 and 2010 respectively. Tien Ho-Phuoc is now a lecturer/researcher at the University of Da Nang - University of Science and Technology. His research interests include super resolution, visual attention, detection, recognition, and deep learning

Hanh T. M. Tran is currently a Lecturer with the Department of Electronics and Telecommunications, the University of Danang - University of Science and Technology, Vietnam, where she joined since 2009. She received the B.Eng. and M.Eng. degrees in Electronics and Telecommunications from the University of Danang - University of Science and Technology in 2008 and 2011, respectively. She obtained the Ph.D. degree from the University of Leeds, United Kingdom, in 2018. She was a Visiting Researcher with the Arizona State University, Arizona, USA, in 2012. Her main research interests include image/video processing, machine learning, deep learning, anomaly detection, object detection and recognition.

Duc Hoang Xuan is currently a final year student at the Department of Electronics and Telecommunications, the University of Danang - University of Science and Technology, Vietnam. His research interests include Machine Learning and Image Processing.

Tuan Dao-Duy received the B.E. degree from the University of Danang - University of Science and Technology, Danang, Vietnam, in 2008, and the M.E. and Ph.D. degrees in computer science and information engineering from National Cheng Kung University, Taiwan, in 2013 and in 2019, respectively. He is currently an PhD Lecturer at the Department of Electronic and Telecommunication Engineering, the University of Danang - University of Science and Technology, Danang, Vietnam. His research interests include IP mobility management, wireless communications, mobile network protocols and vehicular network.

Hoang Le Uyen Thuc received her BE degree in electronics from the University of Danang - University of Science and Technology (DUT), Vietnam, in 1994 and her ME degree in communications from Hanoi University of Technology, Vietnam, in 1997. She then received her PhD degree in computer science under the joint program between Electronics and Telecommunications Engineering (ETE) department, DUT, Vietnam and Information Processing Lab, Electrical and Computer Engineering department, University of Washington, USA, in 2017. She cur-rently works as a senior lecturer of ETE department, DUT. Her research interests include signal processing, pattern recognition, and human behaviour recognition.