Face mask classification using convolutional neural networks with facial image regions and super resolution

Niwan Wattanakitrungroj[1], Wiphada Wettayaprasit[2], Peemakarn Rujirapong[2], Sasiporn Tongman[3]
[1]School of Information Technology, King Mongkut's University of Technology Thonburi, Bangkok, Thailand
[2]Computational Science Division, Faculty of Science, Prince of Songkhla University, Songkhla, Thailand
[3]Department of Biotechnology, Faculty of Science and Technology, Thammasat University, Pathum Thani, Thailand

Article Info

Article history:

Received Jul 29, 2023
Revised Dec 28, 2023
Accepted Jan 6, 2024

Keywords:

Computer vision
Convolutional neural networks
Face mask detection
Generative adversarial networks
Residual networks

ABSTRACT

Face mask classification is relevant to public health and safety, so an approach for face mask classification using multi-task cascaded convolutional networks (MTCNN) for face detection on image data, ResNet152 architecture for feature extraction, and super-resolution method, blind super-resolution generative adversarial networks (BSRGAN), for enhanced image quality was proposed. The classification model was trained by a fully connected layer of neural networks. The goal is to classify each facial image into three classes: the image with a mask, without a mask, or with an incorrectly worn mask. The performance of each classification model on two real-world datasets was evaluated by Accuracy, Precision, Recall, and F1 score for different sets of input patterns which were features extracted from the facial image regions including their combinations. Using multiple image regions, i.e. face, nose, and mouth, as resources for preparing input features showed the improved classification performance compared to using single image regions. In addition, the super-resolution technique applied to medium or large-sized images can improve the performance of the face mask classification model. Our findings may further guide the development for greater effective models and techniques on face mask classification contributing to practical scenarios.

Corresponding Author:

Wiphada Wettayaprasit
Computational Science Division, Faculty of Science, Prince of Songkhla University
Songkhla, Thailand
Email: wiphada.w@psu.ac.th

1. INTRODUCTION

One of computer vision tasks for identifying facial images of person wearing a mask is generally called as face mask detection. Many systems aim to accurately detect such masked face images, since it is necessary to prevent the spread of infectious diseases and to help us to easily manage security protocol of both private and public areas such as security camera system, surveillance CCTV during a COVID-19 outbreak, person authentication system, and so on [1]–[3]. Especially, during world public health emergency situation like COVID-19 pandemic, face mask detection system is one of many safety measures until today as a new normal way of life. Therefore, to automatically detect a person's face with or without mask using computer vision approaches, many previous techniques were developed [4]. Mostly, those techniques are supervised masked face classification from machine learning models applied to detect crowd or individuals wearing the mask. Their applications are not only for airborne disease protection, but also for high air toxification level indication. Thus the good face mask detection system should be robust and highly correct. This supervised image classification problem can be considered at two stages: face detection and face with or without mask identification. In the first stage, faces within an input image or video frame are located by face detection algorithm which returns image's regions that probably contain faces. Haar cascades [5] method, Viola-Jones algorithm [6], or deep learning-based approaches like multi-task cascaded neural networks (MTCNN) [7] or you only look once (YOLO) [8], these are some well-known machine learning object detection methods to effectively identify faces in the image. The output of this stage is usually a bounding box or a set of bounding boxes around the detected faces in each image. After the faces are detected, whether each face wearing a mask or not, this is decided by the consecutive stage which is classification. In the face classification stage, it is frequently implemented by training convolutional neural networks (CNNs) models [8]–[12] , one of successful machine deep learning models in computer vision tasks. These popular models, for instances, visual geometry group (VGG) [13], ResNet [14], MobileNets [15], [16], have shown high abilities in learning face image features in order to quite accurately separate between different classes of faces, such as two classes i.e. masked and unmasked faces, or three classes i.e. masked, unmasked, and improperly masked faces. CNNs models trained by transer learning approches for the face mask detection system involve the usage of the effectively pretrained models trained by feeding a large image dataset [17]–[21]. Although initial weights of these pretrained models, e.g. ImageNet, can be rapidly applied for face mask detection model training, it should be noted that the resulting models may show the suboptimal performance, not their best one. Thus, training CNNs model from the ground up in non-transfer learning process is another way for face mask detection model training [22]–[24]. By this process, the random weight initialization is performed, and then the features are learned from the provided data input data during model training procedure. It may be slower than using the pretrained model, but it offers a chance to reach the nearly best model performance when training the representative and large data for such a domain-specific problem. Hence, whether using pretrained models in the CNNs framework or not, this depends on several factors, e.g. resources of computation, the satisfied performance, and the existing set of training samples. Among many previous works proposed to solve face mask image classification problem, techniques designed for region-based feature extraction, such as locating the specific image regions and then performing relevant feature computation, usually helps to improve accuracy and automation of supervised classification models. Because the important features of such image regions related to the face with or without mask can be kept. Apart from that, enhancing the feature representation of the face, nose, and mouth image regions by applying preprocess and super-resolution GAN-based techniques [25]–[28] may be offer additionally useful features from medium or low-resolution images in order to become an input feature set of the training data samples. In this research, a comprehensive approach for face mask classification using CNNs incorporated with enhanced face-nose-mouth features is presented to improve model performance. The following sections are the proposed methodology in detail, the experimental results including discussion, and final summation of this work, respectively.

2. METHOD

One main idea is to make used of the extracted features from the nose and mouth image regions for improving model prediction performance. The classification model is trained for identifying each face image covered by either a mask or no mask. First of all, the MTCNN algorithm to locate the nose and mouth regions in each face image was employed. Besides features computed from the nose and mouth image regions, the relevant features from the whole face image were also extracted. Next, the residual neural networks in the form of ResNet152 architecture were used in order to perform feature extraction process. For each image as an input, the output from the average pool layer before the final output layer of ResNet152 was obtained as each input feature vector. By this way, all extracted features from each face image as well as its nose and mouse images were combined as one input feature vector. After that, the combined feature vectors were fed into a fully connected neural network to train a classification model capable of classifying whether each face image wearing a mask or not. The methodology employed in this study is described as the following details.

2.1. MTCNN for detecting nose and mouse regions

By taking an advantage of MTCNN [29] which is one of the cascaded neural networks, the face, nose, and mouth regions on each input image can be accurately detected and aligned to facilitate the efficeint feature extraction. The aligned face, nose, and mouth regions become various combianation of inputs to our classification framework. Initially, the face image was resized to 224×224 pixels. Then, MTCNN was applied to find the nose and mouth image regions with the sizes of $h \times h$ and $w \times h$, respectively, as follows. Let $(x_{\text{nose}}, y_{\text{nose}})$ be a pixel position of the nose. The nose region is defined as a rectangle with four corner coordinates:

$$(x_{\text{nose}} - {}^{h}\!/_{2}, y_{\text{nose}} - {}^{h}\!/_{2}),$$
$$(x_{\text{nose}} - {}^{h}\!/_{2}, y_{\text{nose}} + {}^{h}\!/_{2} - 1),$$
$$(x_{\text{nose}} + {}^{h}\!/_{2} - 1, y_{\text{nose}} - {}^{h}\!/_{2}) \text{ and}$$
$$(x_{\text{nose}} + {}^{h}\!/_{2} - 1, y_{\text{nose}} + {}^{h}\!/_{2} - 1).$$

Let $(x^{(l)}_{\text{mouth}}, y^{(l)}_{\text{mouth}})$ and $(x^{(r)}_{\text{mouth}}, y^{(r)}_{\text{mouth}})$ represent the left and right positions of the mouth, respectively. The center of the mouth is defined as $(x_{\text{mouth}}, y_{\text{mouth}})$ where $x_{\text{mouth}} = {}^{1}\!/_{2}(x^{(l)}_{\text{mouth}} + x^{(r)}_{\text{mouth}})$ and $y_{\text{mouth}} = {}^{1}\!/_{2}(y^{(l)}_{\text{mouth}} + y^{(r)}_{\text{mouth}})$. The mouth region is defined as a rectangle with four corner coordinates:

$$(x_{\text{mouth}} - {}^{w}\!/_{2}, y_{\text{mouth}} - {}^{h}\!/_{2}),$$
$$(x_{\text{mouth}} - {}^{w}\!/_{2}, y_{\text{mouth}} + {}^{h}\!/_{2} - 1),$$
$$(x_{\text{mouth}} + {}^{w}\!/_{2} - 1, y_{\text{mouth}} - {}^{h}\!/_{2}) \text{ and}$$
$$(x_{\text{mouth}} + {}^{w}\!/_{2} - 1, y_{\text{mouth}} + {}^{h}\!/_{2} - 1).$$

For instance, $h = 50$ and $w = 100$ are given. The position of the left and the right mouth corners including nose are obtained as $(x^{(l)}_{\text{mouth}}, y^{(l)}_{\text{mouth}})$, $(x^{(r)}_{\text{mouth}}, y^{(r)}_{\text{mouth}})$ and $(x_{\text{nose}}, y_{\text{nose}})$, respectively. The 100×50 mouth region and the 50×50 nose region are illustrated in Figure 1.

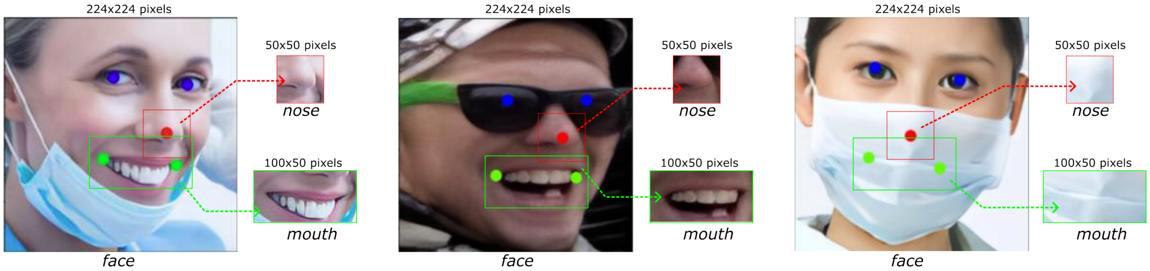


Figure 1. Nose and mouth image regions extracted by MTCNN

2.2. Image enhancement with super resolution

Super-resolution techniques can help to create an image with high-resolution from an original one with low-resolution. Approaches based on generative adversarial network (GAN) models are widley used, lately. For an example, super-resolution generative adversarial network (SRGAN) [25] has also been employed in face mask classification. Other related examples, methods that are later built upon SRGAN, e.g. enhanced super-resolution generative adversarial network ESRGAN) [27], ESRGAN+ [28], and blind super-resolution generative adversarial network (BSRGAN) [26], in order to to yield the better high-resolution images. Synthesizing a high-resolution face image using a low-resolution one by these super-resolution techniques, it can offer the more relevant features about mask or no-mask face images including nose and mouth images, aiding in training accurate classification model. BSRGAN was made upon the SRGAN and ESRGAN models to solve the challenge of blind super-resolution when the specific degradation process is unknown. The improved training process of BSRGAN can deal with more various real-world image problems. Some generated image results by using BSRGAN are shown in Figure 2.

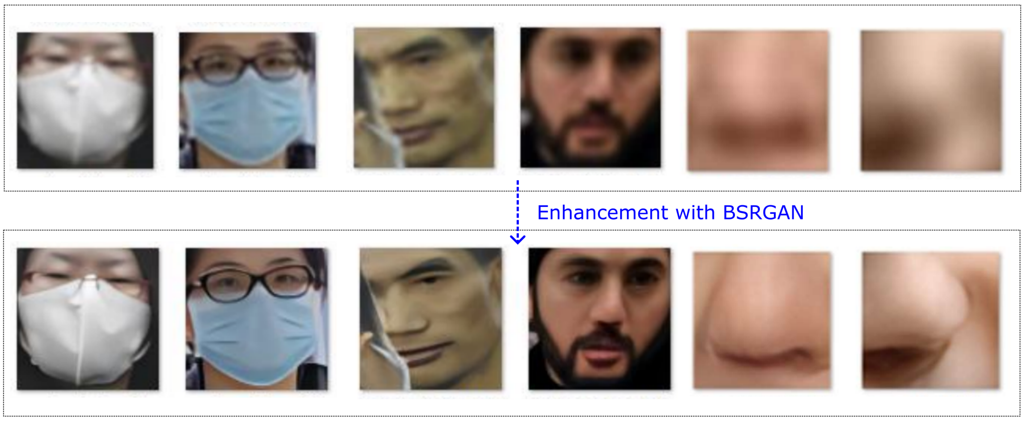


Figure 2. Enhanced images with super resolution using BSRGAN

2.3. ResNet152 for feature extraction

ResNet152 is one of the residual networks (ResNets) family which is a type of CNNs where 152 in ResNet152 stands for the number of layers in the network or its depth. ResNets [14] are widely used for many tasks associated with image feature extraction such as object detection, image classification, and segmentation. Their architechtures were designed to tackle the the disappearing gradient issue usually found in the considerably deep CNN architerchtures. The residual blocks are formed by multiple convolutional layers and skip connections which let the network to more effectively learn and also increase performance. ResNet152 was applied to image feature extraction in this work. An input image was fed into the network, and its features were captured by its layers. It performs gradual reducing image spatial dimensions as well as increasing the extracted feature size. In the final layer, the average pooling layer of ResNet152's convolutional layers gives us the feature vector representing this input image. In Figure 3, the RGB image size 224x224 pixels was processed resulting in a 2048-dimensional feature vector.

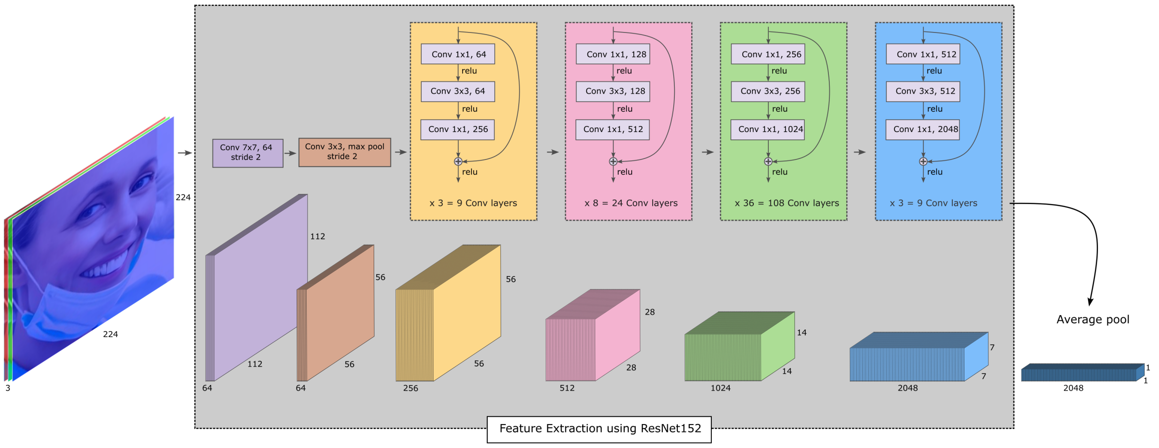


Figure 3. The architecture of ResNet152 used for feature extraction

2.4. Proposed face mask classification

Our proposed method for face mask classification brings MTCNN, BSRGAN, ResNet152, and a fully connected layer together in order to achieve precise and effective classification as depicted in Figure 4. Firstly, to enhance the quality of face images, BSRGAN wae utilized for ensuring optimal image analysis in the subsequent steps. Next, MTCNN is employed to extract the nose and mouth regions from the face images which are served as important areas for extracting helpful features in face mask classification. During the model training process, each of face, nose, and mouth images individually enters the network of ResNet152. The vector of high-level features from these images is obtained by the output of the ResNet152 average pooling layer. Each feature vector corresponding to the face, nose, and mouth images obtained from each ResNet152 is then concatenated into one input feature vector which allows the combination of the extracted features from different image regions. Finally, each combined input feature vector is passed through a fully connected layer of neural networks which functions as the final classification layer. The resulting classifier is responsible for accurate prediction, e.g. identifying the face image wearing, not wearing, or incorrectly wearing a mask in three-class face mask classificaion problem.

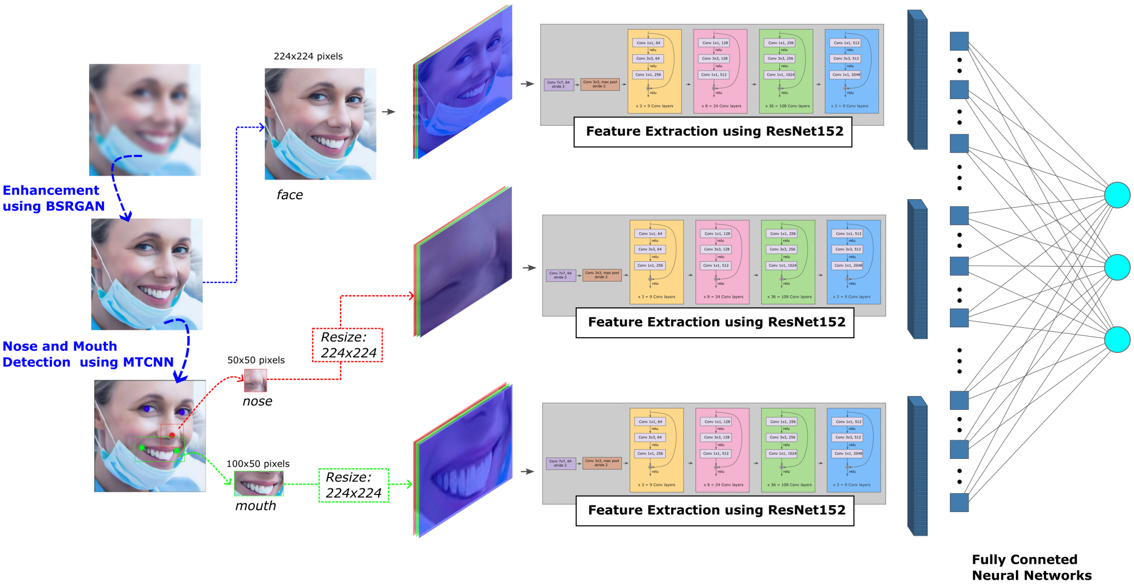
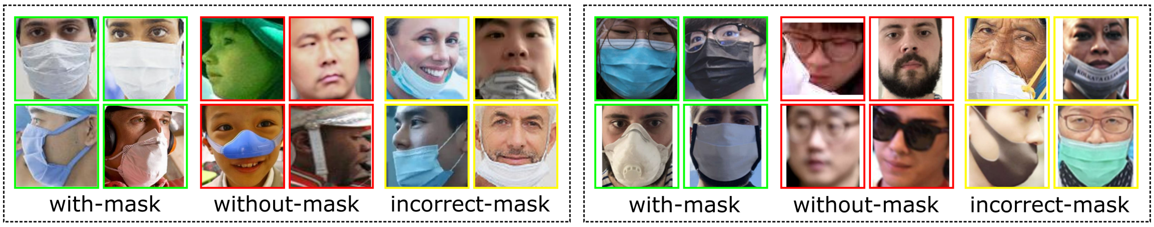

224x224 pixels
face
Enhancement using BSRGAN
Nose and Mouth Detection using MTCNN
50x50 pixels
nose
Resize: 224x224
100x50 pixels
mouth
Resize: 224x224
Feature Extraction using ResNet152
Feature Extraction using ResNet152
Feature Extraction using ResNet152
Fully Conneted Neural Networks

Figure 4. The proposed framework for face mask classification

2.5. Dataset preparation

Two datasets were used to evaluate the performance of models in various conditions. The first dataset is the face mask label dataset (FMLD) [22] which originally contains 41,937 images annotated with 63,072 face images. The subset of FMLD dataset is our first experimental dataset, called FMLD5K dataset, is comprised of 5,000 annotated face images which includes labels for the face with mask, without mask, and with mask worn incorrectly. The second experimental dataset called AFMDK dataset is the face mask detection dataset [30] which consists of 853 images annotated with 4,072 face images. Examples of face images from FMLD5K dataset and AFMDK dataset are illustrated in Figure 5(a) and Figure 5(b), respectively. An explanation of the two datasets is provided in Table 1.



(a) FMLD5K (b) AFMDK

Figure 5. Examples of face images from (a) FMLD5K and (b) AFMDK datasets

Table 1. The number of train and test samples in each class of FMLD5K and AFMDK datasets

Dataset		#Faces	#with-mask	#without-mask	#incorrect-mask
FMLD5K:	Train	3,500	1,750	875	875
	Test	1,500	750	423	327
	Total	5,000	2,500	1,298	1,202
AFMDK:	Train	2,850	2,250	511	89
	Test	1,222	982	206	34
	Total	4,072	3,232	717	123

2.6. Model evaluation

To estimate the model classification performance of our proposed method, Accuracy, Precision, Recall, and F1 score are computed as the percentage form. The whole correctness of the model's performance is measured by Accuracy. The number of correctly classified masked faces divided by the number of all predicted masked faces is represented by Precision. Recall or sensitivity is the number of correctly classified masked faces divided by all actual masked faces. F1 score is the harmonic mean of Precision and Recall which expressing a balanced measure of the model's performance. The following formulas are for two classes:

$$\begin{aligned}
\text{Accuracy} &= \frac{\text{TP}+\text{TN}}{\text{TP}+\text{TN}+\text{FP}+\text{FN}} \times 100\\
\text{Precision} &= \frac{\text{TP}}{\text{TP}+\text{FP}} \times 100\\
\text{Recall} &= \frac{\text{TP}}{\text{TP}+\text{FN}} \times 100\\
\text{F1} &= 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}} \times 100
\end{aligned}$$

When evaluating a classification model, it is important to understand the following terms. Suppose that there are two classes, namely, positive and negative classes. True Positive (TP) and True Negative (TN) refer to the number of instances that are correctly classified as positive class and negative class, respectively, by the model. False Positive (FP) is the number of instances that are actually negative class but incorrectly classified as positive class. While, False Negative (FN) is the number of instances that is actually positive class but incorrectly classified as negative class.

3. RESULTS AND DISCUSSION

The results from two experimental scenarios were showed and discussed. In the first scenario, a goal is to examine the performance of multiple input patterns which are the combination of feature vectors extracted from images of face, nose, and mouth comparing with the various single input patterns which are each feature vector extracted from only image of face, nose, or mouth. In the second scenario, its goal is to investigate the effect of enhanced images using the super-resolution technique for yielding the better performance of face mask classification model. The details are as follows.

3.1. Comparison on face mask classification with various input patterns

In this study, a comprehensive analysis of the model performance was conducted when the input patterns from images of face, nose, and mouth were variously combined as many sets of input feature vectors for both two-class and three-class classification tasks. The two-class classification task is distinguishing between facial images with and without masks while the three-class classification task is the extended version of the two-class classification that includes facial images with incorrect position of masks. The results are shown in Table 2. FMLD and AFMDK datasets were used for evaluating the performance of models with various input patterns including face, nose, mouth images, and their combinations. For two-class classification tasks on both datasets, the combined input patterns from face, nose, and mouth images are consistently outperformed the individual input patterns in terms of Accuracy, Precision, Recall, and F1 score. In the context of the three-class classification task, the performance of these different input patterns in classifying each facial image to one of three classes, i.e. an image with the mask, without the mask, or with incorrect mask position was further explored. For the FMLD dataset, the combination of face, nose, and mouth input patterns made its model performance reached the highest scores across all evaluation metrics. On the other hand, for the AFMDK dataset, the input patterns consisting of face and nose images gave us the most favorable results of the model performance although the combination of face, nose, and mouth input patterns also yielded acceptable performance.

Table 2. The model performance of various input patterns for two-class and three-class classification

Datasets	Input patterns	Accuracy	Precision	Recall	F1
FMLD5K (2 classes)	face	92.13	92.22	92.13	92.13
	nose	81.40	84.33	81.40	80.99
	mouth	75.40	79.09	75.40	74.60
	nose+mouth	81.87	84.25	81.87	81.55
	face+nose	92.67	92.74	92.67	92.66
	face+mouth	92.67	92.74	92.67	92.66
	face+nose+mouth	**92.87**	**92.93**	**92.87**	**92.86**
AFMDK (2 classes)	face	96.97	95.27	95.13	95.20
	nose	92.47	93.59	82.09	86.36
	mouth	91.24	90.07	80.86	84.41
	nose+mouth	92.96	94.40	83.03	87.30
	face+nose	**97.14**	**95.40**	**95.54**	**95.47**
	face+mouth	96.97	95.14	95.28	95.21
	face+nose+mouth	**97.14**	**95.40**	**95.54**	**95.47**
FMLD5K (3 classes)	face	84.00	81.99	81.34	81.12
	nose	68.27	63.47	58.95	59.54
	mouth	67.67	66.59	59.92	61.29
	nose+mouth	71.80	69.94	63.22	64.69
	face+nose	85.00	82.94	82.43	82.16
	face+mouth	84.73	82.75	82.18	81.97
	face+nose+mouth	**85.13**	**83.11**	**82.71**	**82.41**
AFMDK (3 classes)	face	96.40	91.24	83.03	**86.01**
	nose	91.41	75.67	61.24	65.80
	mouth	90.43	64.64	55.33	58.20
	nose+mouth	91.57	79.12	62.38	67.63
	face+nose	**96.48**	**91.38**	**83.06**	**86.10**
	face+mouth	96.24	90.70	80.94	84.28
	face+nose+mouth	96.40	91.02	82.08	85.23

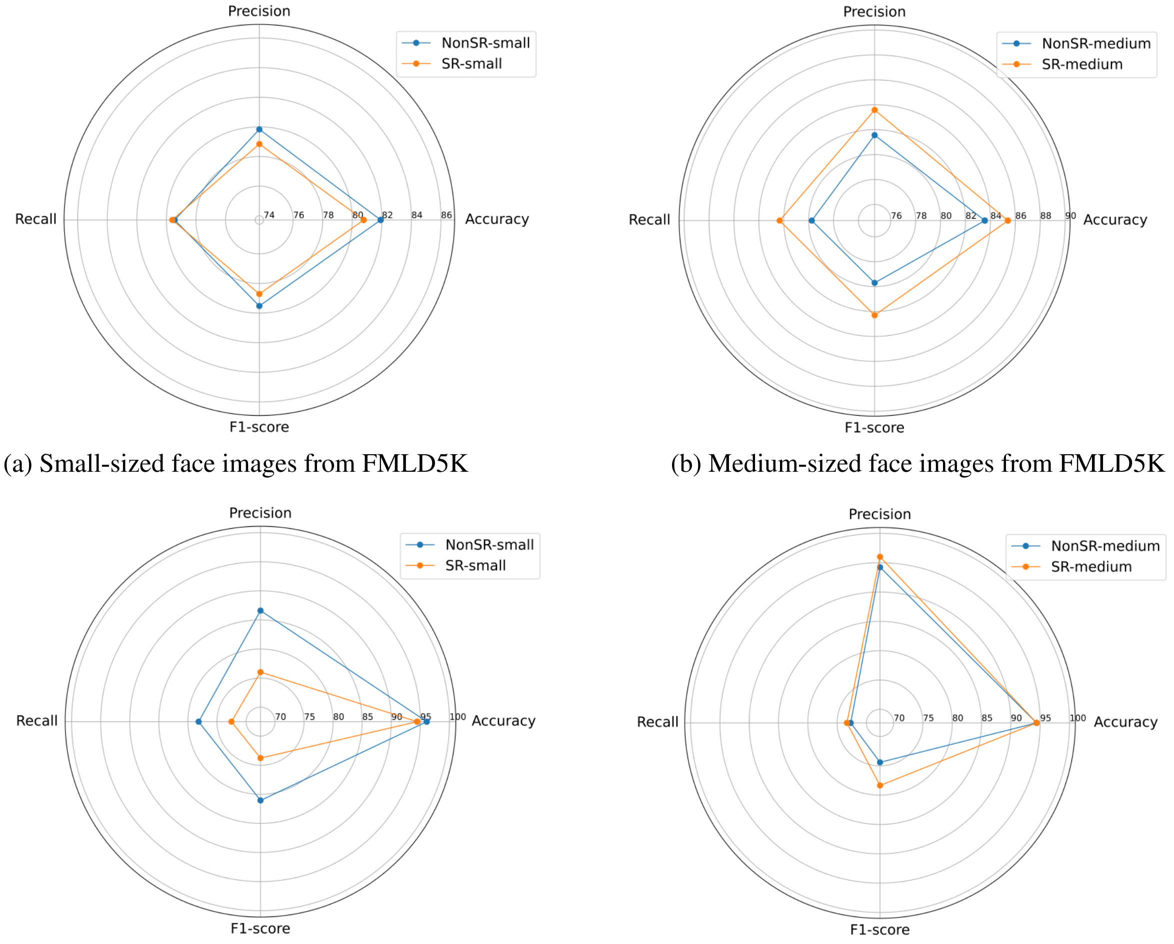
In conclusion, the results indicate that the input feature set (as input patterns) extracted from multiple facial image regions. For example, face, nose, and mouth regions generally brings out the improved classification performance in both two-class and three-class classification tasks. However, the specific combination of input patterns that yields the best results may depend upon the dataset as well as the number of classes.

3.2. Comparison on face mask classification with super resolution

To improve the performance of model, the images were ehnaced their resolution by using super-resolution (SR) GAN-based technique, i.e. BSRGAN, before feature extraction process. Based on our experimental results, the image size quite gives impacts the model's performance. This experimetal setup focused on analyzing the model's performance which was trained by two sets of image sizes: a small set (less than 2,500 pixels) and a medium set (no lesser than 2,500 pixels). The summarization of two data sets according to their sizes is provided in Table 3. Each combined input pattern set that included face, nose, and mouth image regions from the FMLD5K and AFMDK datasets was experimented to develop a model for classifying images into three categories, i.e. the image with a mask, without a mask, or with an incorrectly worn mask. The performance of the models when image enhancement was applied and not applied were measured for the small and medium face image sizes. The performance results on the FMLD5K dataset, i.e. small-sized face images and medium-sized face images, are illustrated in Figure 6(a) and Figure 6(b), respectively. Similarly, the performance results on the small-sized and the medium-sized face images from AFMDK dataset are shown in Figure 6(c) and Figure 6(d), respectively. For these two datasets, the application of image enhancement techniques, specifically super-resolution (SR) method like BSRGAN, is beneficial for building the improved model trained by medium-sized or larger face images. But, for small-sized face images, the image enhancement does not cause extra improvements in model performance. In brief, our experimental results indicate that the use of BSRGAN can particularly boost the performance of models when applied to the train set with medium-sized or larger face images.

Table 3. Details of image sizes of FMLD5K and AFMDK datasets

Dataset	#Faces	#Faces-small	#Faces-medium
FMLD5K:			
Train	3,500	320	3,180
Test	1,500	260	1,240
Total	5,000	580	4,420
AFMDK:			
Train	2,850	2,386	464
Test	1,222	1,866	36
Total	4,072	3,412	660



(a) Small-sized face images from FMLD5K

(b) Medium-sized face images from FMLD5K

(c) Small-sized face images from AFMDK

(d) Medium-sized face images from AFMDK

Figure 6. Four testing performance values of models trained by images with super resolution (SR) and models trained by image without super resolution (NonSR), denoting that subfigures (a) and (b) show results of small-sized and medium-sized face images from FMLD5K dataset, while subfigures (b) and (c) show results of small-sized and medium-sized face images from AFMDK dataset

4. CONCLUSION

In this research, the face mask classification method was offered. This method utilized the multi-task cascade convolutional neural networks (MTCNN) for locating nose and mouth regions on the face images and ResNet152 for feature extraction from these regions. The objective was to classify face images into three classes, i.e. the image with a mask, without a mask, or with an incorrectly worn mask. The effectiveness of our proposed method was displayed through experimental evaluation on FMLD5K and AFMDK datasets. The results demonstrate that the combined multiple features extraced from face, nose, and mouth images beated the features extraced from a sigle source due to the higher values of Accuracy, Precision, Recall, and F1 score. This emphasizes the importance of capturing intricate information as input feature patterns related to face mask presence by integrating multiple facial image regions. Furthermore, the impact of image enhancement based on super-resolution GAN-based techniques like BSRGAN on the classification performance was examined. The results reveal that applying image enhancement to non-small face images gives us the increased Accuracy, Precision, Recall, and F1 score comparing with non-enhanced images. This suggests that adjusting the visual quality of face images can bring better feature extraction and subsequently enable the higher classification performance. On the whole, the effectiveness of joining multiple feature sources and the benefits of the enhanced image for face mask classification were proposed. Besides, the specific considerations for small-sized face images, in future research, was also acknowledged.

REFERENCES

[1] I. Javed *et al.*, "Face mask detection and social distance monitoring system for COVID-19 pandemic," *Multimedia Tools and Applications*, vol. 82, no. 9, pp. 14135–14152, Apr. 2023, doi: 10.1007/s11042-022-13913-w.

[2] S. Sanjay, S. S. N. Soorya, R. Vengatesh, and K. C. S. H. Priya, "Security access control system enhanced with face mask detection and temperature monitoring for pandemic trauma," in *2022 2nd International Conference on Intelligent Technologies (CONIT)*, Jun. 2022, pp. 1–6, doi: 10.1109/CONIT55038.2022.9848266.

[3] H. K. Wong and A. J. Estudillo, "Face masks affect emotion categorisation, age estimation, recognition, and gender classification from faces," *Cognitive Research: Principles and Implications*, vol. 7, no. 1, p. 91, Oct. 2022, doi: 10.1186/s41235-022-00438-x.

[4] Y. Himeur, S. Al-Maadeed, I. Varlamis, N. Al-Maadeed, K. Abualsaud, and A. Mohamed, "Face mask detection in smart cities using deep and transfer learning: lessons learned from the COVID-19 pandemic," *Systems*, vol. 11, no. 2, p. 107, Feb. 2023, doi: 10.3390/systems11020107.

[5] A. B. Shetty, Bhoomika, Deeksha, J. Rebeiro, and Ramyashree, "Facial recognition using Haar cascade and LBP classifiers," *Global Transitions Proceedings*, vol. 2, no. 2, pp. 330–335, Nov. 2021, doi: 10.1016/j.gltp.2021.08.044.

[6] J. M. Al-Tuwaijari and S. A. Shaker, "Face detection system based Viola-Jones algorithm," in *6th International Engineering Conference "Sustainable Technology and Development" (IEC)*, Feb. 2020, pp. 211–215, doi: 10.1109/iec49899.2020.9122927.

[7] S. Sakshi, A. K. Gupta, S. Singh Yadav, and U. Kumar, "Face mask detection system using CNN," in *2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE)*, Mar. 2021, pp. 212–216, doi: 10.1109/ICACITE51222.2021.9404731.

[8] S. Abbasi, H. Abdi, and A. Ahmadi, "A face-mask detection approach based on YOLO applied for a new collected dataset," in *2021 26th International Computer Conference, Computer Society of Iran (CSICC)*, Mar. 2021, pp. 1–6, doi: 10.1109/CSICC52343.2021.9420599.

[9] H. Goyal, K. Sidana, C. Singh, A. Jain, and S. Jindal, "A real time face mask detection system using convolutional neural network," *Multimedia Tools and Applications*, vol. 81, no. 11, pp. 14999–15015, May 2022, doi: 10.1007/s11042-022-12166-x.

[10] K. Podbucki, J. Suder, T. Marciniak, and A. Dabrowski, "CCTV based system for detection of anti-virus masks," in *2020 Signal Processing: Algorithms, Architectures, Arrangements, and Applications (SPA)*, Sep. 2020, pp. 87–91, doi: 10.23919/SPA50552.2020.9241303.

[11] R. K. Kodali and R. Dhanekula, "Face mask detection using deep learning," in *2021 International Conference on Computer Communication and Informatics (ICCCI)*, Jan. 2021, pp. 1–5, doi: 10.1109/ICCCI50826.2021.9402670.

[12] J. R. V. Jeny, B. Shraddha, B. Ashritha, D. S. Sai, and M. Naveen, "Deep learning framework for face mask detection," in *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, Jun. 2021, pp. 1705–1712, doi: 10.1109/ICOEI51242.2021.9452930.

[13] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *3rd International Conference on Learning Representations, ICLR 2015 - Conference Track Proceedings*, 2015.

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

[15] A. G. Howard *et al.*, "MobileNets: Efficient convolutional neural networks for mobile vision applications," 2017, [Online]. Available: http://arxiv.org/abs/1704.04861.

[16] M. Sandler, A. Howard, M. Zhu, A. Zhmoginov, and L. C. Chen, "MobileNetV2: Inverted residuals and linear bottlenecks," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Jun. 2018, pp. 4510–4520, doi: 10.1109/CVPR.2018.00474.

[17] J. Gathani and K. Shah, "Detecting masked faces using region-based convolutional neural network," in *2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS)*, Nov. 2020, pp. 156–161, doi: 10.1109/ICIIS51140.2020.9342737.

[18] R. Liu and Z. Ren, "Application of Yolo on mask detection task," in *2021 IEEE 13th International Conference on Computer Research and Development (ICCRD)*, Jan. 2021, pp. 130–136, doi: 10.1109/ICCRD51685.2021.9386366.

[19] P. S. Reddy, M. Nandini, E. Mamatha, K. V. Reddy, and A. Vishant, "Face mask detection using machine learning techniques," in *2021 5th International Conference on Trends in Electronics and Informatics (ICOEI)*, Jun. 2021, pp. 1468–1472, doi: 10.1109/ICOEI51242.2021.9452826.

[20] S. Srinivasan, R. Rujula Singh, R. R. Biradar, and S. A. Revathi, "COVID-19 monitoring system using social distancing and face mask detection on surveillance video datasets," in *2021 International Conference on Emerging Smart Computing and Informatics (ESCI)*, Mar. 2021, pp. 449–455, doi: 10.1109/ESCI50559.2021.9396783.

[21] M. R. Karim Sujon, M. R. Hossain, M. J. Al Amin, C. Bepery, and M. M. Rahman, "Real-time face mask detection for COVID-19 prevention," in *2022 IEEE 12th Annual Computing and Communication Workshop and Conference (CCWC)*, Jan. 2022, pp.

341–346, doi: 10.1109/CCWC54503.2022.9720764.

[22] B. Batagelj, P. Peer, V. Štruc, and S. Dobrišek, "How to correctly detect face-masks for COVID-19 from visual information? " *Applied Sciences*, vol. 11, no. 5, p. 2070, Feb. 2021, doi: 10.3390/app11052070.

[23] L. Shuangyan and G. Huayong, "Lighter and faster face mask detection method based on YOLOv5," in *2023 IEEE 6th Information Technology,Networking,Electronic and Automation Control Conference (ITNEC)*, Feb. 2023, pp. 1016–1022, doi: 10.1109/ITNEC56291.2023.10082188.

[24] W. Vijitkunsawat and P. Chantngarm, "Study of the performance of machine learning algorithms for face mask detection," in *2020 - 5th International Conference on Information Technology (InCIT)*, Oct. 2020, pp. 39–43, doi: 10.1109/InCIT50588.2020.9310963.

[25] C.Ledig *et al.*, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Jul. 2017, pp. 105–114, doi: 10.1109/CVPR.2017.19.

[26] K. Zhang, J. Liang, L. Van Gool, and R. Timofte, "Designing a practical degradation model for deep blind image super-resolution," in *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, Oct. 2021, pp. 4771–4780, doi: 10.1109/ICCV48922.2021.00475.

[27] X.Wang *et al.*, "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proceedings of the European conference on computer vision (ECCV) workshops*, 2018, p. 0.

[28] N. C. Rakotonirina and A. Rasoanaivo, "ESRGAN+: Further improving enhanced super-resolution generative adversarial network," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, May 2020, pp. 3637–3641, doi: 10.1109/ICASSP40776.2020.9054071.

[29] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, Oct. 2016, doi: 10.1109/LSP.2016.2603342.

[30] A. Larxel, "Face Mask Detection,", June 2023, [Online]. Available: https://www.kaggle.com/datasets/andrewmvd/face-mask-detection

BIOGRAPHIES OF AUTHORS

Niwan Wattanakitrungroj received the Ph.D. in Computer Science and Information Technology from Chulalongkorn University, Thailand in 2018. She also received M.Sc. in Computer Science and B.Sc. in Mathematics from Prince of Songkla University, Thailand in 2007 and 2004, respectively. She is currently lecturing with the School of Information Technology, King Mongkut's University of Technology Thonburi, Thailand. Her research interests are Machine Learning, Neural Networks, and Pattern Recognition. She can be contacted at email: niwan.watt@mail.kmutt.ac.th or watta.niwan@gmail.com.

Wiphada Wettayaprasit is an assistant professor at Department of Computer Science, Division of Computational Science, Faculty of Science, Prince of Songkla University, Thailand. She got B.Sc., M.Sc., Ph.D. in Computer Science from Prince of Songkla University, University of Missouri-Columbia, USA, and Chulalongkorn University, respectively. Her research interests are Artificial Intelligence, Neural Networks, and Machine Learning. She can be contacted at email: wiphada.w@psu.ac.th.

Peemakarn Rujirapong holds a Bachelor of Engineering from Prince of Songkla University, Thailand in 2017. He is currently a master's student at Division of Computational Science, Faculty of Science, Prince of Songkla University. His research interests are computer vision and image analysis. He can be contacted at email: 6310220004@psu.ac.th.

Sasiporn Tongman completed her Ph.D. in Computer Science and B.Sc. in Biochemistry from Chulalongkorn University, Bangkok, Thailand, in 2017 and 2006, respectively. Currently, she is a lecturer at Department of Biotechnology, Thammasat University, Pathum Thani, Thailand. Her primary research fields of interest are about applying and improving computational techniques and machine learning approaches to gain insights from data, especially bio-data, and make progress in computer-based bioscience. She can be contacted at email: tongman.sas@gmail.com.