Masked face with facial expression recognition based on deep learning

Nawal Younis Abdullah[1], Ahmed Mamoon Fadhil Alkababji[2]
[1]Department of Computer Engineering Technology, Northern Technical University, Mosul, Iraq
[2]Department of Computer Engineering, Mosul University, Mosul, Iraq

Article Info

Article history:

Received Dec 29, 2021
Revised Apr 8, 2022
Accepted May 16, 2022

Keywords:

Convolutional neural networks
Face detection
Face expression
Haar classifier
Masked face

ABSTRACT

Wearing masks contributed to slowing the spread of the coronavirus disease (COVID-19) as the World Health Organization (WHO) recommended wearing face masks especially with the spreading of virus variants like omicron. Although people accept the idea of wearing these masks, it is still unknown the effect of covering parts of the face on social interaction among people in general and children in particular. Moreover, Social isolation affects emotional moods, which causes stress, sadness, and depression. In the current study, we have been exploring the emotional inferences on faces with and without a mask. The system can pick up the universal emotions: fear, disgust, anger, surprise, contempt, sadness, and happiness. The researchers in deep learning are concerned with global pandemic COVID-19 to enhance public health service. The proposed model is developed with a machine learning algorithm through the Haar feature-based cascade classifiers. The built model can detect people's emotions with mask and without a mask with high accuracy.

Corresponding Author:

Nawal Younis Abdullah
Department of Computer Engineering Technology, Northern Technical University
Mosul, Iraq
Email: nawal_younis@ntu.edu.iq

1. INTRODUCTION

Deep learning has achieved great success in a variety of real-world applications and fields. One of the most essential capabilities of deep learning is object detection [1], [2]. The rapid spread of the Corona virus has forced many countries of the world to issue preventive instructions requiring the wearing of face masks. Governments have adopted special strategies to maintain social distancing in order to reduce the number of infections among citizens and to preserve the life of medical personnel. Governments have implemented new restrictions requiring people to wear facemasks because the coronavirus spreads by airdrops and close contact. Face masks are used to limit the spread of the virus and prevent it from spreading between people. To prevent the virus from spreading, governments have sought the assistance of the army and police in many densely populated places, whose residents do not accept a commitment to social distancing, despite the danger of this matter on the lives of people and security forces. This research focus on detection and recognition of face emotion from a face masked images as shown in Figure 1. Starting from scratch, a deep learning model has been built to detect the region of masked face with the face expression of the input image. The main contribution of this paper is the development of a deep learning-based detector model. By using different datasets, this model can automatically detect masked faces with expression on an image.

Input images Output images

Figure 1. The result of the proposed system

2. RELATED WORK

Object detection and recognition from images is the most advanced aspect of computer vision, and it is widely employed in a variety of applications [3], [4]. Both supervised and unsupervised learning have been used in computer vision field to accomplish the object detection operations in images. This section includes many of related academic papers in the field of object detection for both face mask and expression detection. Most of papers focused on face mask detection based on traditional and deep learning techniques. Our work focusses on detecting face mask with expression to help in decreasing the spreading of coronavirus. Ghazal and Abdullah [5] a machine learning classifiers construction based on the traditional methods have been used for features extraction, Viola-Jones detector [6], Haar features method for face detection [7]-[9], while human detection is accomplished by a histogram of oriented gradients (HOG) feature extraction method in [10]-[12], facial expressions in [13]-[15], scale-invariant feature transform (SIFT) in [16] and local binary pattern (LBP) are the tools used for object detection and its application [17]. In the recent years, the deep learning-based detector proved an excellent performance in feature extraction [18], A convolutional neural network (CNN), which has a superior feature extraction capability, that make pattern recognition tasks relevant to computer vision more accurate [19]. CNN is a superior method in face recognition, as it uses convolution kernels to convolve with the source images or feature maps to excerpt higher-level characteristics [20], [21]. In this paper a CNN model has been proposed for face mask detection and expression recognition.

3. THE PROPOSED MODEL

The suggested face mask and expression recognition model is illustrated in block diagram in Figure 2. Before applying the data to the system some work is done to prepare the data and make it in a suitable size and form for the system. Then, using a Haar-like feature, a face detection stage is being used to detect the face in the image. Finally, the face expression of the input image is discovered with the help of CNN.

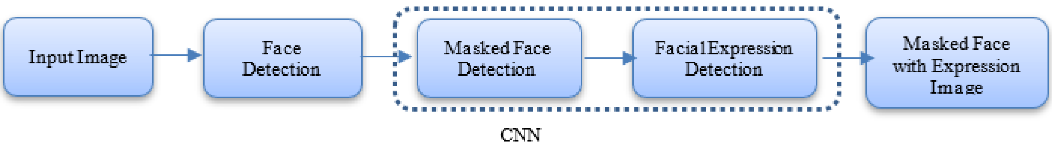


Figure 2. The proposed system

3.1. Datasets preparation

In this paper two public datasets have been used for experimental work. The first dataset is for Face Masks, published in (https://www.kaggle.com/ omkargurav/face-mask-dataset), which consists of 7553 images, divided into two groups: 3725 masked faces and 3828 without mask faces. Some samples are shown in Figure 3. The second dataset is a face expression recognition dataset published on Kaggle (https://www.kaggle.com/deadskull7/fer2013), which contains 35,887 grayscale images of faces with 48*48 pixels stored in CSV format. Angry, disgust, fear, happy, sad, surprise, and neutral are the seven categories. Some samples of fer2013 introduced in Figure 4.

Figure 3. Face mask dataset samples

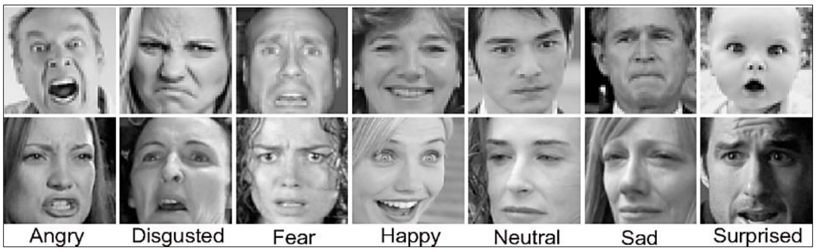


Figure 4. Facial emotions dataset samples

A pre-processing stage is needed to prepare the data to be fed in the proposed model. Cleaning and arranging raw data to make it acceptable for creating and training the proposed model is part of data preparation. The acquired datasets have been gathered from different sources, combined in a comma-separated values (CSV) file format, categorized according to Table 1, and annotated.

Table 1. Categories of expression

With mask							Without mask						
Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral	Angry	Disgust	Fear	Happy	Sad	Surprise	Neutral

3.2. Face detector

Haar feature-based cascade classifiers has been used to perform face detection in the proposed model. It is a machine learning-based approach for object detection in which a cascade function is developed from a large number of positive (with faces) and negative (without faces) images. A common Haar feature can be applied for face detection since the region of the eyes is darker than the region of the nose in the face image shown in Figure 5.

The difference between the summation of pixels in the dark and bright areas, as represented in (1), gives the Haar-like feature:

$$F(x, y) = \sum (F\ \text{white}) - \sum (F\ \text{black}) \quad (1)$$

where:
F (x, y)=is the Haar-like feature
∑ (F white)=the sum of pixels of bright area
∑ (F black)=the sum of pixels of dark area
The face detector will check each part of the image searching for a face, and classify it as (face) or (not face). The Haar-classifier can take data from an extensible markup language (XML) file to determine how each image location should be classified [22].

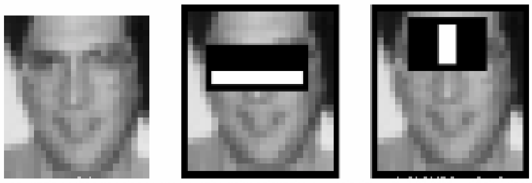


Figure 5. An image with Haar-like feature

3.3. Face mask and emotion detector

Convolutional neural networks (CNNs) are commonly used for object detection and recognition tasks because of their precise yet simple architecture [23]-[25]. In this work, an algorithm from scratch has been written based on CNN for face mask and emotion detection. A simple CNN architecture is shown in Figure 6. The CNNs include three types of layers:
- Convolutional layers.
- Pooling layers.
- Fully-connected layers.

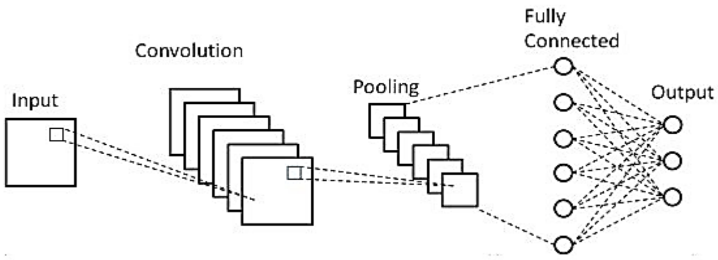


Figure 6. The basic architecture for CNN

From Figure 6. the CNN functionality is divided into four key areas: the input layer holds image pixel values, followed by the convolutional layer, which involves multiplying a set of weights (filter) with the input data. This multiplication produces a two-dimensional array of output values that indicate input filtering; this operation is known as a feature map. After constructing a feature map, each value can be passed through a nonlinearity, such as a rectified linear units (ReLU), in the same manner that the outputs of a fully connected layer are passed through a nonlinearity. The pooling layer will then just do down sampling along the spatial dimensions of the input, thus reducing the number of parameters within that activation. Finally, the fully-connected layers will use the activations to generate class scores, which will be used to classify the data.

The proposed structure of the CNN algorithm illustrated in Figure 7. Two sequential models have been built from scratch. The face mask detector made up of three convolution and max pooling layers respectively for features extraction then a softmax classifier for face mask classification with two outputs (With mask and Without mask). For the face emotion detection, a model made up of four convolution and max pooling layers respectively to extract all features in the image and then a softmax classifier for face emotion classification with seven outputs (angry, disgust, fear, happy, sad, surprise, and neutral).

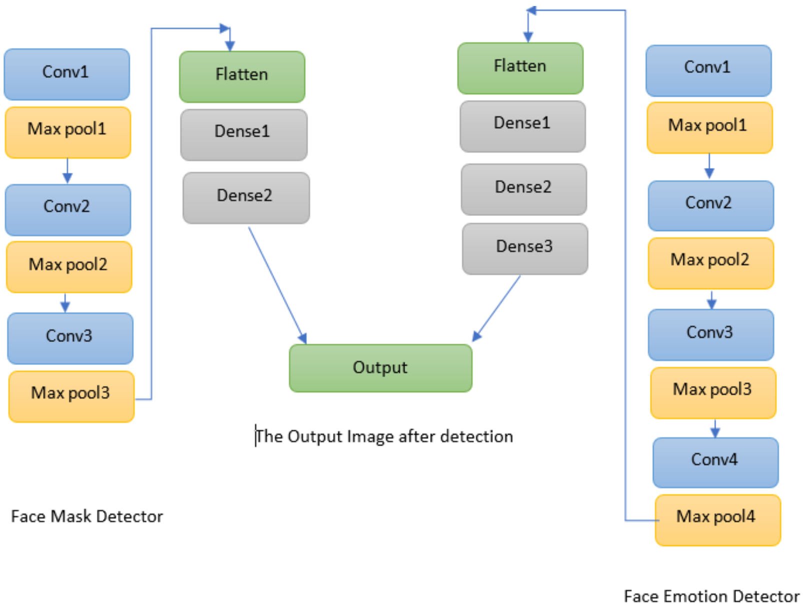
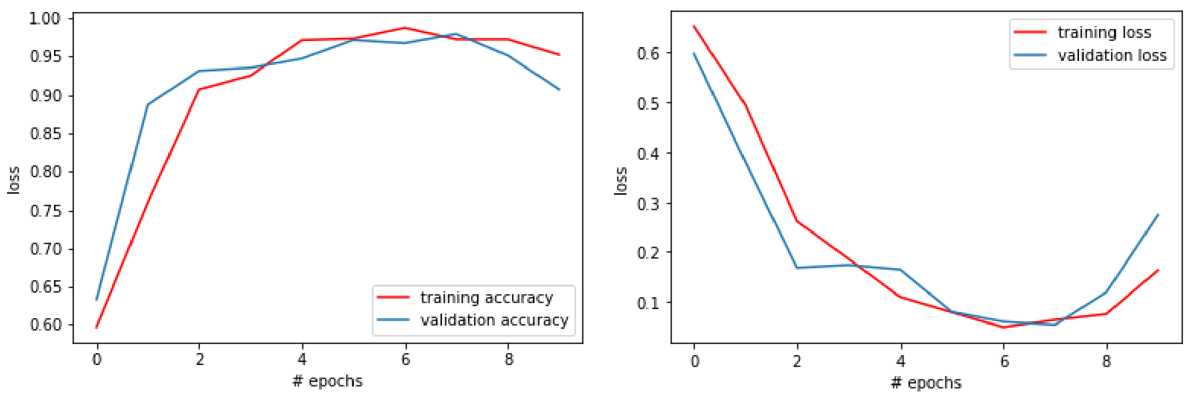


Figure 7. The structure of the detection system

4. EXPERIMENTAL WORK AND RESULTS

The experimental work for the proposed system was designed to train a CNN-based automatic facial mask and emotion detection system. At the pre-processing stage the untrainable images have been excluded, also, two different datasets are used to train the system, each of them has its specific specifications: the first is for face mask detection, the input images have been resized to 100*100 pixels, normalized, and split it in to 80% training, 10% test, and 10% validation before fed to the CNN model. The dataset trained with 10 epochs and 0.001 learning rate, which give an accuracy of 0.9565 and loss of 0.1829 as shown in Figure 8. For the second system of the face emotion detection the dataset is resized to 48*48 pixels, normalized, and split it in to 98% training and 2% test then fed to the CNN model. The dataset trained with 20 epochs and 0.01 learning rate, which result an accuracy of 0.9565 and loss of 0.1829 the normalized confusion matrix is shown in Figure 9. Adam optimizer and categorical cross-entropy have been used in both models based on TensorFlow environment, which resulted of high system performance. The system can detect faces with/ without mask and the human face emotion with/ without mask, some examples of detection system have been illustrated in Figure 10.



Training/Validation accuracy
The best train accuracy: 0.9565

Training/Validation loss
The best train loss: 0.1829

Figure 8. The system accuracy of the proposed face mask detection system

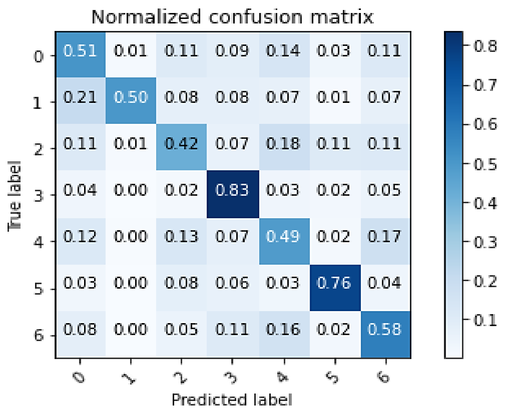


Figure 9. The confusion matrix of the proposed face emotion detection system

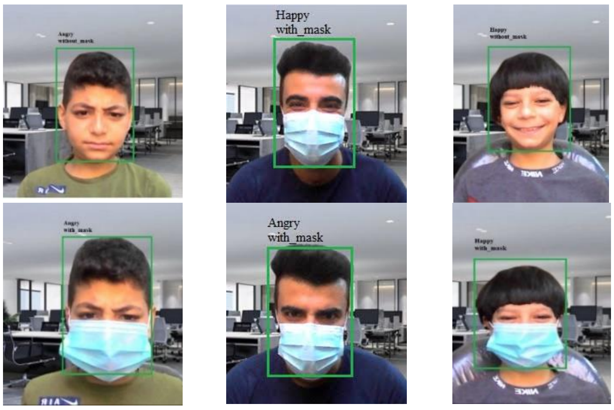


Figure 10. Samples of test result images

5. EVALUATION WITH RELATED WORK AND DISCUSSION

The detector model that has been proposed is a result of starting from scratch deep learning model trained with face mask and face expression datasets. This combined model has a good performance in medical face mask with expression detection. To improve the model Adam optimizer has been used. Indeed, most of the related work concentrate on the classification of the face mask only. The performance of the proposed model in compare with different methods is shown in Table 2. According to the results of our work the accuracy is equal to 95%.

Table 2. The performance of the proposed model in compare with other methods

Reference	Method	Classification	Detection	Result
Ref [19]	PCA	Yes	No	AC= 70%
Ref [1]	YOLOv2 with ResNet	Yes	Yes	AC=81%
Ref [20]	LLE-CNNs	Yes	Yes	AC=76%
Proposed	Start from scratch	Yes	Yes	AC=95%

6. CONCLUSION

This paper presents a combination of two CNN models starting from scratch to detect faces with and without masks to limit the spread of the Coronavirus. Moreover, a face emotion detection model is combined to the first model to help identify how children and autistic patients feel when wearing a mask. The research contributes on health care in the era that accompanied the Coronavirus. The proposed method's efficiency and performance are shown in the experimental work. It has also been noted that some emotions are difficult to estimate due to facial expression convergence. For the future work, authors will enhance the proposed model to improve the efficiency of the system.

REFERENCES

[1] M. Loey, G. Manogaran, M. H. N. Taha, and N. E. M. Khalifa, "Fighting against COVID-19: A novel deep learning model based on YOLO-v2 with ResNet-50 for medical face mask detection," *Sustainable Cities and Society*, vol. 65, p. 102600, Feb. 2021, doi: 10.1016/j.scs.2020.102600.

[2] M. Ghazal., N. Waisi, and N. Abdullah, "The detection of handguns from live-video in real-time based on deep learning," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 6, pp. 3026–3032, Dec. 2020, doi: 10.12928/TELKOMNIKA.v18i6.16174.

[3] N. Y. Abdullah, M. T. Ghazal, and N. Waisi, "Pedestrian age estimation based on deep learning," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 22, no. 3, pp. 1548–1555, Jun. 2021, doi: 10.11591/ijeecs.v22.i3.pp1548-1555.

[4] Y. Liu, P. Sun, N. Wergeles, and Y. Shang, "A survey and performance evaluation of deep learning methods for small object detection," *Expert Systems with Applications*, vol. 172, p. 114602, Jun. 2021, doi: 10.1016/j.eswa.2021.114602.

[5] M. T. Ghazal and K. Abdullah, "Face recognition based on curvelets, invariant moments features and SVM," *Telkomnika (Telecommunication Computing Electronics and Control)*, vol. 18, no. 2, pp. 733–739, Apr. 2020, doi: 10.12928/TELKOMNIKA.v18i2.14106.

[6] K. D. Ismael and S. Irina, "Face recognition using Viola-Jones depending on Python," *Indonesian Journal of Electrical Engineering and Computer Science*, vol. 20, no. 3, pp. 1513–1521, Dec. 2020, doi: 10.11591/ijeecs.v20.i3.pp1513-1521.

[7] P. Viola and M. Jones, "Rapid object detection using a boosted cascade of simple features," in *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, 2001, vol. 1, pp. I-511-I–518, doi: 10.1109/cvpr.2001.990517.
[8] 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.
[9] G. Zheng and Y. Xu, "Efficient face detection and tracking in video sequences based on deep learning," *Information Sciences*, vol. 568, pp. 265–285, Aug. 2021, doi: 10.1016/j.ins.2021.03.027.
[10] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proceedings - 2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, CVPR 2005*, 2005, vol. I, pp. 886–893, doi: 10.1109/CVPR.2005.177.
[11] P. Vijaykumar and J. K. Mani, "Face recognition with frame size reduction and DCT compression using PCA algorithm," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 22, no. 1, p. 168, Apr. 2021, doi: 10.11591/ijeecs.v22.i1.pp168-178.
[12] S. Nurmaini, A. Zarkasi, D. Stiawan, B. Y. Suprapto, S. D. Siswanti, and H. Ubaya, "Robot movement controller based on dynamic facial pattern recognition," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 22, no. 2, p. 733, May 2021, doi: 10.11591/ijeecs.v22.i2.pp733-743.
[13] I. Gangopadhyay, A. Chatterjee, and I. Das, "Face detection and expression recognition using Haar cascade classifier and fisherface algorithm," in *Recent Trends in Signal and Image Processing*, 2019, pp. 1–11, doi: 10.1007/978-981-13-6783-0_1.
[14] S. M. S. Abdullah and A. M. Abdulazeez, "Facial Expression recognition based on deep learning convolution neural network: A review," *Journal of Soft Computing and Data Mining*, vol. 02, no. 01, Apr. 2021, doi: 10.30880/jscdm.2021.02.01.006.
[15] T. Zhang, W. Zheng, Z. Cui, Y. Zong, J. Yan, and K. Yan, "A deep neural network-driven feature learning method for multi-view facial expression recognition," *IEEE Transactions on Multimedia*, vol. 18, no. 12, pp. 2528–2536, Dec. 2016, doi: 10.1109/TMM.2016.2598092.
[16] P. Felzenszwalb, D. McAllester, and D. Ramanan, "A discriminatively trained, multiscale, deformable part model," in *26th IEEE Conference on Computer Vision and Pattern Recognition, CVPR*, Jun. 2008, pp. 1–8, doi: 10.1109/CVPR.2008.4587597.
[17] T. Ojala, M. Pietikäinen, and D. Harwood, "A comparative study of texture measures with classification based on feature distributions," *Pattern Recognition*, vol. 29, no. 1, pp. 51–59, Jan. 1996, doi: 10.1016/0031-3203(95)00067-4.
[18] Z. Zou, Z. Shi, Y. Guo, and J. Ye, "Object detection in 20 years: A survey," 2019, *arXiv:1905.05055*.
[19] J. Deng, W. Dong, R. Socher, L.-J. Li, Kai Li, and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," in *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Jun. 2010, pp. 248–255, doi: 10.1109/cvpr.2009.5206848.
[20] M. Coskun, A. Ucar, O. Yildirim, and Y. Demir, "Face recognition based on convolutional neural network," in *Proceedings of the International Conference on Modern Electrical and Energy Systems*, Nov. 2017, vol. 2018-January, pp. 376–379, doi: 10.1109/MEES.2017.8248937.
[21] R. I. Bendjillali, M. Beladgham, K. Merit, and A. Taleb-Ahmed, "Illumination-robust face recognition based on deep convolutional neural networks architectures," *Indonesian Journal of Electrical Engineering and Computer Science (IJEECS)*, vol. 18, no. 2, pp. 1015–1027, May 2020, doi: 10.11591/ijeecs.v18.i2.pp1015-1027.
[22] T. Mantoro, M. A. Ayu, and Suhendi, "Multi-faces recognition process using Haar cascades and eigenface methods," in *International Conference on Multimedia Computing and Systems -Proceedings, May 2018*, vol. 2018-May, pp. 1–5, doi: 10.1109/ICMCS.2018.8525935.
[23] M. Ghazal, R. Albasrawi, N. Waisi, and M. A. Hammoshi, "Smart meeting attendance checking based on a multi-biometric recognition system," *Przeglad Elektrotechniczny*, vol. 98, no. 3, pp. 93–96, Mar. 2022, doi: 10.15199/48.2022.03.21.
[24] S. Albawi, T. A. Mohammed, and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proceedings of 2017 International Conference on Engineering and Technology, ICET 2017*, Aug. 2018, vol. 2018-January, pp. 1–6, doi: 10.1109/ICEngTechnol.2017.8308186.
[25] N. Z. Waisi, N. Y. Abdullah, and M. T. Ghazal, "The automatic detection of underage troopers from live-videos based on deep learning," *Przeglad Elektrotechniczny*, vol. 2021, no. 9, pp. 85–88, Sep. 2021, doi: 10.15199/48.2021.09.18.

BIOGRAPHIES OF AUTHORS

Nawal Younis Abdullah received the M.Sc. degree from Computer Engineering Technology, Northern Technical University/Mosul, Iraq in 2014. Her M.Sc. thesis entitled: "FPGA Based Video Scene Boundaries Detection Using Enhanced Sobel Filter" and her current research interests include the deep learning CNNs models and object detection. She is currently a Ph.D. student in Computer Engineering. She can be contacted at email: nawal_younis@ntu.edu.iq.

Ahmed Mamoon Fadhil Alkababji received the B.Sc., M.Sc. and PhD degree in Electrical Engineering from University of Mosul, Iraq in 1994. 1996 and 2007. Currently, an assistant professor at computer engineering department in University of Mosul, Iraq. Research interest in signal processing, realtime systems and biometric. He can be contacted at email: ahmedalkababji72@uomosul.edu.iq.