Masked Face Detection And Identification By Using Deep Learning Technology

Ruaa M Al-Alaf[1*],Dr. Fadwa Al-Azzo[2],Dr. ThabatThabet

Abstract

In the wake of the global health disaster brought on by the globally circulating COVID-19 coronavirus. It is currently a research topic in many fields, especially those interesting such as artificial intelligence and new information technologies. Many regulatory agencies now require wearing face masks, particularly in crowded areas involving regular and large-scale human interaction, like inside overcrowded transit facilities, where everyone must wear masks. It is challenging to identify the identity of a person using conventional facial recognition techniques, so it needs developed technology with high accuracy. The paper presents a new system by utilizing the advanced MobileNetV2 network to recognize the person's identity without the need to take off the face mask. The proposed system has trained by using different eight classes of regular people's faces (without wearing face masks) under diverse environmental conditions. The performance of the proposed system demonstrated high efficiency in identifying the identity of the person accurately up to 100%. The recognition process was achieved using Keras with TensorFlow in terms of accuracy and detection speed.

KeyWords: Face identification, MobileNetV2, Transfer learning (TL) , ImageNet , Confusion Matrix .

DOINumber:10.14704/nq.2022.20.8.NQ44778 **NeuroQuantology2022;20(8):7547-7555**

Introduction

Face recognition (FR) is a computer vision problem that involves recognizing and authenticating a person from an image of their face. (FR) Systems are used in facial image processing tasks, and their prominence as a research subject has recently increased. These systems are most typically used and selected for people and security cameras, and they can be used for a range of security purposes, such as crime reduction, video surveillance, and human verification[1]. Computer vision is a form of machine learning applied to the science of vision[2], and it is an essential aspect of the artificial intelligence field. The aim of digital vision is really to collect images or videos, analyze the images or videos, and then extract the necessary information [3].

Computer vision is currently widely employed in a variety of applications, including video surveillance, This enzyme is that the mainkey to the transfer automatic driving, medical attention, and consumption, and can be studied from the standpoints of object recognition and space vision[4]. The goal of object perception is to identify the type of item, whereas the goal of space vision is to determine the object's position and shape. In the discipline of computer vision, several problems such as image categorization, FR, and object identification are currently being worked on to solve problems. FR has always been a significant project among them.

Correspondingauthor: Ruaa M Al-Alaf
Address: 1*Department of Computer Engineering, Technical Engineering College Mosul, Northern Technical University, Mosul, Iraq.
2 Lecturer, Department of Computer Engineering, Technical Engineering College Mosul, Northern Technical University, Mosul, Iraq.
3Lecturer, Department of Computer Engineering, Technical Engineering College Mosul, Northern Technical University, Mosul, Iraq.
1*E-mail:ruaa.motasm@ntu.edu.iq , 2fadwaalezzo@ntu.edu.iq; 3thabet.tfy@ntu.edu.iq

Relevantconflictsof interest/financialdisclosures:

Basically, Attendance, security, and finance are the most common applications for FR technology, but it is also used in logistics, retail, smartphones, aviation, education, real estate, public administration, entertainment marketing, network information security, and other areas[5]. It is a huge step forward in artificial intelligence technology, meaning the need for more precise, adaptive, and speedier recognition technology. FR is the process of recognizing a previously identified face. It might be defined as recognizing people from our connected user database, who may be known or unknown[6]. The tasks of face recognition include "verification" and" identification". Face verification is a one-to-one match that compares a face image to a database of the available facial image to match a person's personality. Face recognition is a one-to-one problem in which a query image of the face is matched to an image in the face of the database. When the question face may or may not be in the avail faces databank. In this state, a similarity score can be computed based on the data. FR technologies have always been the most important component of identification, which has nearly failed due to new major challenges for face recognition-based authentication applications such as public entrance and exit, facial identity verification, facial participation, and facial access control, among others. Gateways at public checkpoints rely on a traditional face recognition system, which is incapable of acknowledging the Masked Faces during the COVID-19 outbreak [7]. The MobileNetV2 model was used, which is a pre-trained model on a set of images.

In this proposed work, a face recognition MobileNetV2 on custom data set will be created. This is intended to train the model on images that have not been seen before. Eight classes were created and these classes were divided into 85% For training and 15% for verification. The proposed system goal to define the person's identity even while still wearing a face mask, the input dataset has been collected from eight classes of normal people's faces of different (ages and gender).

Related work

proposed a simplified method for detecting and notifying individuals who had not wearing face masks. The system model had been trained and tested using Kaggle datasets. The mask has been created from real-time public faces and fed into a convolutional neural network as an input (CNN)[8]. proposed a hybrid face recognition algorithm that integrates Principal Component Analysis (PCA) and Linear Discriminant Analysis (LDA), two face recognition methods. The Eigenvector required by PCA and LDA algorithms is computed using the Jacobi approach. A Raspberry Pi 3 board-based embedded system with a face recognition system. The Raspberry Pi 3 module was used to implement the face recognition system on an embedded system that had a 97 percent accuracy rate.[9]. developed and designed a face recognition-based security system using machine vision. It was suggested to use Principal Component Analysis (PCA) in conjunction with a classifier for Euclidean distance to recognize faces. The study employed a training database made up of 20 students. The change in pose orientation is largely responsible for the 86 percent success rate and 14% mistake rate of the face recognition system on fixed face images. Additionally, it has been discovered that the threshold value has a crucial role in how well facial recognition works.[10]. proposed the image super-resolution and classification networks to create a special facemask-wearing condition recognition technique (SRCNet), The suggested technique has four basic components: facemask-wearing condition recognition, facial detection, and cropping, image super-resolution, and image pre-processing. The Medical Masks Dataset which consists of 3835 images and 671 images of people not wearing face masks, 134 images of people wearing face masks incorrectly, and 3030 images of people wearing face masks correctly. This data had been served as the training and evaluation for this approach. Finally, the suggested SRCNet beat conventional end-to-end picture classification techniques employing deep learning and obtained 98.70 percent accuracy[11]. studied the performance of drop support vector machines (SVM) by training it on the deep features of similar pre-trained networks, as well as the applicability of MobileNet V2 deep convolutional neural networks on palmprint identification. Datasets from the Hong Kong Polytechnic University of Science and Technology were used to conduct the studies. The collection consists of 6000 128 x 128-pixel grayscale images of 500 different palms. The strategy of MobileNet V2-based features with an SVM classifier was outperformed the best results previously published achieving the best average testing and validation accuracy rate of 100%[12].

Methodology

The proposed system shown in figure 1 aims to determine the person's identity even while still wearing a face mask. Many studies require the input dataset from a mixture of masked and unmasked faces to achieve during the training process to identify the person's identity[13]. Whereas the proposed work in this paper, the input dataset has been collected from eight classes of normal people's faces (without wearing face masks) of different ages and gender. Then the labeling technology has been used for drawing a bounding box (ground truth) around each person's face within eight classes of the input dataset.

The faces labeled will be an input to the MobilenetV2 network to be trained and extract the vital features of the face. Eventually, the proposed system will be tested for identification and identification of the person, even if (he or she) is wearing a face mask.

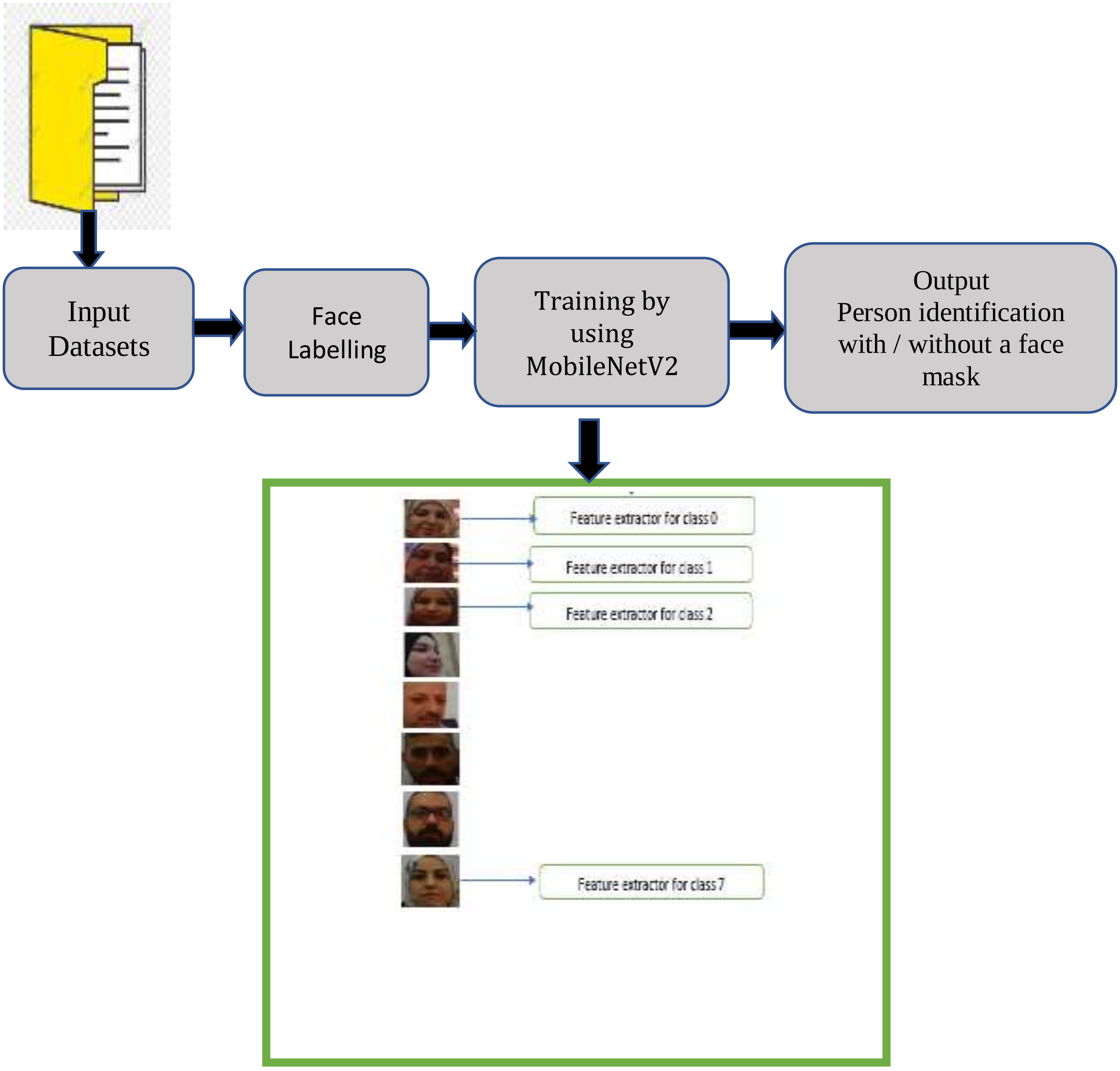


Figure 1. The block diagram for identifying the masked face

Collecting Datasets

The proposed system is evaluated on custom datasets involving eight classes of human faces belonging to different people undertaken on conditions such as various angles, gender (male, female), and camera low resolution. The datasets have been divided into a training set and validation set of 85% and 15% respectively to evaluate the candidate model efficiency performance. The primary stage is drawing a bounding box (ground truth) around only the face regardless of the rest of the other objects in the same image by using "haarcascade_frontalface_default.xml"[14]. As samples, figure 2 indicates samples of datasets for two persons under different conditions (gender, angles, and resolutions) by drawing a bounding box. Therefore, the proposed system understands that this person inside the box belonged to the person class. Afterward, all files are collected in a labels folder.

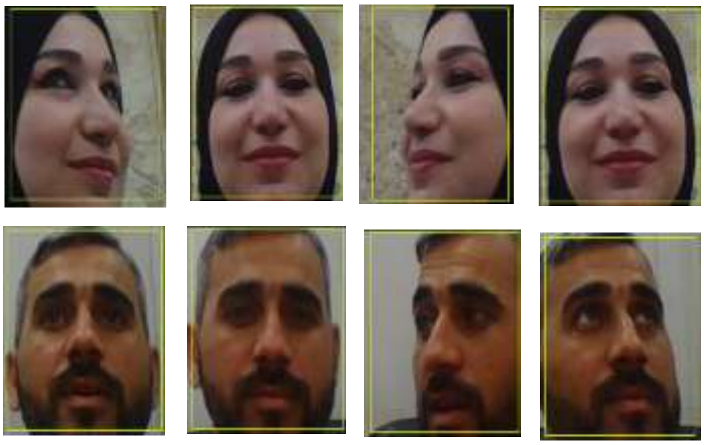


Figure 2: Samples of datasets for two persons of under different conditions (gender, angles and resolutions) by drawing bounding box.

MobileNetV2

In MobileNetV2, ‹there are two "types of blocks. The first one is a residual block‹with a stride of 1. Another, one is block with a stride of 2 for downsizing. There are three layers for both, type of blocks the first layer is "1×1" convolution with ReLU6. The depth wise convolution is the second layer. The third' layer is a 1x1 convolution once more, but this time there is no non- linearity[15] .

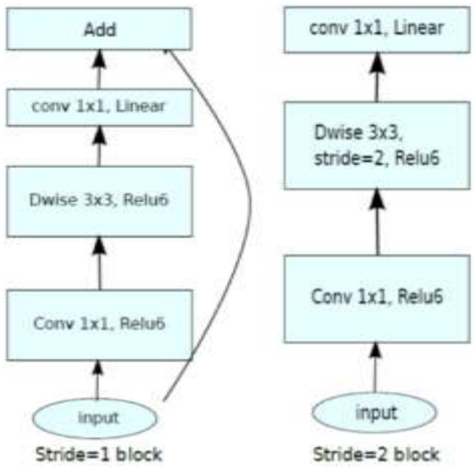


Figure 3: MobileNetV2 Convolutional Block

Table1 : Architectural details of the MobileNetV2 model

Input	Operator	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	3	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	-	

where t represents expansion factor, n: repeating number c: number of output channels, and s: stride. 3×3 "kernels" are used for spatial convolution[16].

Learning Rate (LR)

Learning -Rate (LR) is a crucial hyperparameter to adjust for efficient deep neural network training (DNNs)[17]. It is difficult to choose an appropriate constant rate for training a DNN where efficient learning rates require multi-step adjustment of LR values at different training process phases[18]. Therefore, Accuracy will decrease along with training speed if the learning rate is too high also training speed and application accuracy both increase as the learning rate declines.

Transfer learning (TL)

Transfer learning (TL) is a method that aims to apply knowledge gained from solving one problem to another that is unrelated but still present. Consider the scenario when a system needs to identify human faces in an image. Related issues can be resolved without a lot of data by using models that have previously been trained on millions of faces, often called pre-trained models. Figure 4 illustrates three typical ways that transfers may improve learning measures. The first performance in the goal task that can be achieved with only the transferred knowledge is compared to the initial performance of an ignorant agent. The second factor is how much quicker it is to fully learn the target task using the transferred information than it is to start from scratch. The third is the difference between the final performance level that can be achieved in the target task and the final level without transfer. A negative transfer has taken place if a transfer method actually makes the performance worse.

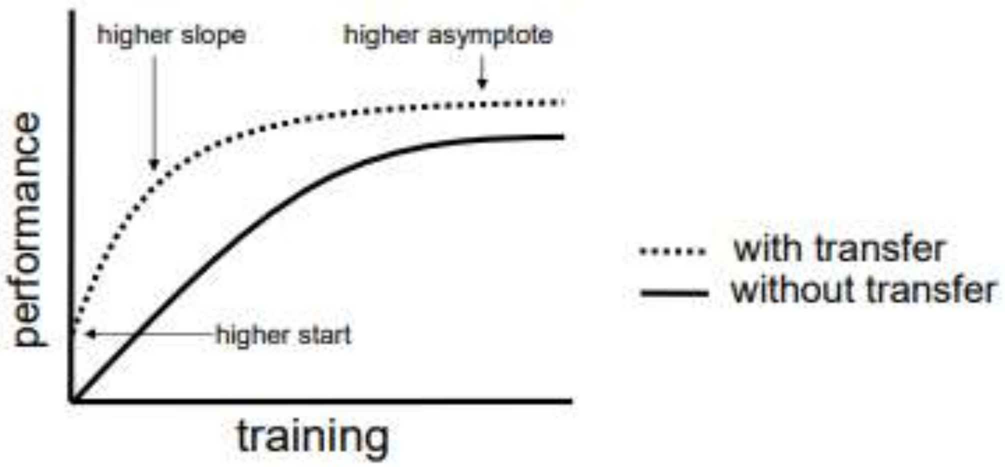


Figure 4: Three ways in which transfer improves learning[19]

Experimental Results

In this paper, the dataset composed of employees who worked in the Nineveh Governorate was collected using the "haarcascade_frontalface_default.xml" to run the face detector on each image and produce a bounding box around each face. All platforms use a deep open-source TensorFlow learning system. TensorFlow is an artificial intelligence learning system that calculates logarithmic graphs [20]. The platform analyzes and processes (AI) neural network models and is easy to use.

Training the Proposed Model

The proposed model employs the MobileNetV2 to personal identity detection even ‹wearing a face mask. MobileNetV2 is used to load pre-trained weights from ImageNet[21]. The proposed model is fine-tuned and the weights are saved. Fine-tuning is an operation to take the weights of a network pattern that had been before trained and used it as starting for a new model of a second identical task. It might be utilized to rate up the training operation and to decrease the possibility of overfitting in case the expand of the dataset is not enough. For this reason, using pre-trained models saves time by allowing to apply of current weight training while keeping already learned features. In

NeuroQuantology|July2022|Volume20|Issue8| Page 7547-7555|doi:10.14704/nq.2022.20.8.NQ44778

addition to employing pre-trained MobileNetV2 able of an inaccessible put image of size (224 x 224 x 3). The model needs less time to train because RELU has no complex math and it is sparsely" activated. It is demonstrated that the use of pretrained MobileNetV2 with fine-tuned fully connected layers as a classifier could be capable to produce a very surprising result. It has less execution time and memory usage while preserving the classification accuracy high. The fine-tuned pretrained model achieves 100% accuracy for 8 classes within 10 epochs.

Training results

Our findings revealed that the number of epochs had significant effects on the accuracy of the test. The system started to train gradually until it reached 10 epochs. It was noticed that with the increase in the number of epochs, the accuracy of the system increased as the model achieved 100% and the loss (error rate) started decreasing gradually until it approached zero. The number of periods represents the number of times the entire training data set was used to learn the algorithm. The loss is a number representing how accurate the model is to predict and the effects of the number of epochs on the loss are illustrated in Figure 5 and Table 2.

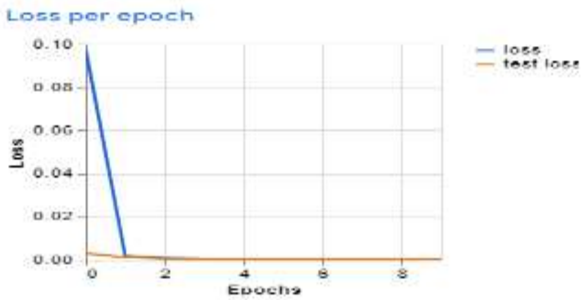
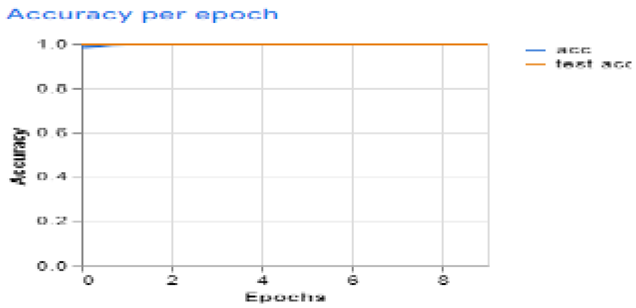
No of epoch	Loss per epoch	Accuracy per epoch
10		

Figure 5: The accuracy and loss curve for epoch **10** with the highest accuracy.

Table 2 Recognition evaluation at Epoch 10		
Epoch	Loss	Accuracy %
1	0.08533	97.88
2	0.001381	99.97
3	0.000574	100
4	0.000325	100
5	0.000236	100
6	0.000191	100
7	0.000165	100
8	0.000165	100
9	0.000138	100
10	0.000130	100

Testing results

The proposed system performance has been tested by taking a person's face image who is wearing a face mask. As a result, the person's identity will be identified because (his or her) features are located within the trained database. Vice versa the system will indicate that is found the person is unknown (a strange person) to the system because there are (his or her) no features within the system database. Accordingly, the proposed system performance will be highly efficient in the detection process of a person's face even if wearing a face mask, so that the detection accuracy reaches up to 100 percent.

Figure 6, observes that although the training was carried out on normal images (without a mask), then the system was able to recognize a person's identity while still wearing a face mask.

Figure 6: Sample of the proposed model prediction for a person's identity detection

Confusion Matrix

The confusion matrix is exceedingly used in deep learning for supervised classification of the behavior of the classification model. The square construction of a confusion matrix is represented through columns and rows where columns are the predicted classes and rows are the actual classes of the instances[22]. Figure 7 depicts the confusion matrix for the proposed DL model's precision and recall after fine-tuning the hyperparameters.

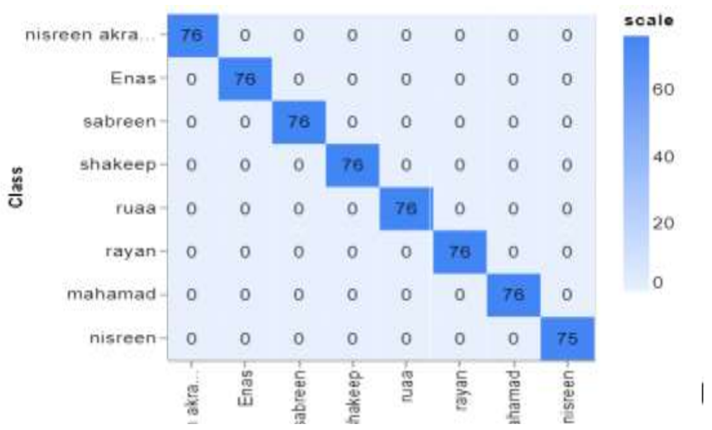


Figure7:confusion matrix for the proposed system

1(nisreen akra)	TP(1)	FN(1)	FN(1)	FN(1)	FN(1)	FN(1)	FN(1)	FN(1)
2(Enas)	FP(1)	TN(1) TP(2)						
3(sabreen)	FP(1)		TN(1) TP(3)					
4(shakeup)	FP(1)			TN(1)				
5(ruaa)	FP(1)				TN(1)			
6(rayan)	FP(1)					TN(1)		
7(mahamad)	FP(1)						TN(1)	
8(nisreen)	FP(1)							TN(1) TP(8)

Figure8 showing the position of TP, FP, and FN for each class in the proposed system

Table 3: Total of TP, FP, and FN for the proposed system

	TP	FP	FN
1(nisreen akra)	76	0	0
2(Enas)	76	0	0
3(sabreen)	76	0	0
4(shakeup)	76	0	0
5(ruaa)	76	0	0
6(rayan)	76	0	0
7(mahamad)	76	0	0
8(nisreen)	75	0	0
Total	607	0	0

True Positives (TP): There is only one cell (highlighted in green) where the true and predicted labels are both "(1) nisreenakra". True Positives will be represented by the number within this cell.

False Positives (FP): These are all the instances when the label was expected to be "(1) nisreenakra" but the real label was something different. Except for the cell with the TP, these are all the cells within the same columns as the true positives (highlighted blue). Therefore, the values in the blue area add up to represent false positives.

False Negatives (FN): All of these are instances where the making prediction is something different even if the real label was "(1) nisreenakra".Except for the cell with TP, all of the cells are within the same rows as the positive instances (highlighted in pink). All those cells add up to False Negatives.

Figure 8 shows the (TP, FP, FN) and table 3 shows these numbers in a table, and sums them up into totals.

	$Precision = \frac{TP}{TP+FN'}$ *100%	Recall= $\frac{TP}{TP+FN}$ *100%	$Accurcy = \frac{(TP+TN)}{[(TP+FP)+(TN+FN)]}$ *100%
1(nisreen akra)	100	100	100
2(Enas)	100	100	100
3(sabreen)	100	100	100
4(shakeup)	100	100	100
5(ruaa)	100	100	100
6(rayan)	100	100	100
7(mahamad)	100	100	100
8(nisreen)	100	100	100

Figure9: precision ,Recall and Accuracy for eight classes of the proposed

Conclusion

This work provided a new approach to identifying the person's identity even still wearing the face mask because many places are required to wear face masks, especially in crowded places where there is frequent and intense human interaction, such as inside overcrowded transportation facilities. The dataset was trained with eight different classes of regular people's faces (without wearing face masks) under diverse environmental conditions. MobileNetV2 is used as the base model with haarcascade_frontalface_default.xml. The number of epochs had a significant effect on the system performance during the dataset training which achieved high accuracy when the number of epochs was set to 10. The proposed system improved facial identity recognition speed with high performance and significant accuracy up to 100% in less loss. Transfer learning and fine-tuning technique were used in this work. The recognition process was achieved using Keras with TensorFlow in terms of accuracy and detection speed.

Acknowledgement

We would like to thank Nineveh Governorate employees for permitting us to usage their images in creating the dataset.

ConflictofInterest

None.

FundingSource

None.

References

F. Schroff and J. Philbin, “FaceNet: A Unified Embedding for Face Recognition and Clustering.”

J. Bai et al., “Real-Time Driver Awareness Detection System Real-Time Driver Awareness Detection System,” 2020, doi: 10.1088/1757-899X/745/1/012053.

N. Sharma, R. Sharma, and N. Jindal, “Machine Learning and Deep Learning Applications-A Vision ☆,” Glob. Transitions Proc., vol. 2, no. 1, pp. 24–28, 2021, doi: 10.1016/j.gltp.2021.01.004.

G. M. Rozhbayani, A. Tuama, and F. Al-azzo, “Real-Time Human Detection and Tracking Based on Deep Learning Technique,” vol. 20, no. 6, pp. 2084–2093, 2022, doi: 10.14704/nq.2022.20.6.NQ22202.

U. H. Salman, K. M. Z. Othman, and A. A. Abdullah, “Security and Integument Preserving Top-k inquiry transforming to Two-Tiered sensor system,” no. October 2018, 2022.

[6] S. S. R et al., “Face Recognition Using Haar - Cascade Classifier for Criminal Identification,” no. 6, pp. 1871–1876, 2019.

A. K. Sharadhi, V. Gururaj, S. P. Shankar, M. S. Supriya, and N. Sanjay, “Face mask recogniser using image processing and computer vision approach,” vol. 3, no. April, pp. 67–73, 2022, doi: 10.1016/j.gltp.2022.04.016.

K. Suresh, M. B. Palangappa, and S. Bhuvan, “Face Mask Detection by using Optimistic Convolutional Neural Network,” Proc. 6th Int. Conf. Inven. Comput. Technol. ICICT 2021, pp. 1084–1089, 2021, doi: 10.1109/ICICT50816.2021.9358653.

N. R. Borkar and S. Kuwelkar, “Real-Time implementation of face recognition system,” Proc. Int. Conf. Comput. Methodol. Commun. ICCMC 2017, vol. 2018-Janua, no. July 2017, pp. 249–255, 2018, doi: 10.1109/ICCMC.2017.8282685.

F. S. A. Saad, M. V. Lee, and S. Basah, “Face recognition using machine vision for Security system,” IOP Conf. Ser. Mater. Sci. Eng., vol. 705, no. 1, 2019, doi: 10.1088/1757-899X/705/1/012005.

Q. Bosheng and D. Li, “Identifying Facemask-Wearing Condition Using Image Super-Resolution with Classification Network,” Mdpi Sens., pp. 1–23, 2020.

A. Michele, V. Colin, and D. D. Santika, “ScienceDirectScienceDirectMobileNet Convolutional Neural Networks and Support Vector Machines for Palmprint Recognition,” ProcediaComput. Sci., vol. 157, pp. 110–117, 2019, doi: 10.1016/j.procs.2019.08.147.

R. Bharadwaj, G. Jaswal, A. Nigam, and K. Tiwari, “Mobile based Human Identification using Forehead Creases : Application and Assessment under COVID-19 Masked Face Scenarios,” pp. 3693–3701.“https://raw.githubusercontent.com/opencv/opencv/master/da
ta/haarcascades/haarcascade_f rontalface_default.xml.”

M. Akay et al., “Deep Learning Classification of Systemic Sclerosis Skin Using the MobileNetV2 Model,” vol. 2, pp. 104–110, 2021, doi: 10.1109/OJEMB.2021.3066097.

B. Khasoggi, “Efficient mobilenet architecture as image recognition on mobile and embedded devices,” vol. 16, no. 1, pp. 389–394, 2019, doi: 10.11591/ijeecs.v16.i1.pp389-394.

S. R. Awad, B. T. Sharef, A. M. Salih, and F. L. Malallah, “DEEP LEARNING-BASED IRAQI BANKNOTES CLASSIFICATION SYSTEM,” pp. 31–38, 2022, doi: 10.15587/1729-4061.2022.248642.

Y. Wu et al., “Demystifying Learning Rate Policies for High Accuracy Training of Deep Neural Networks,” Proc. - 2019 IEEE Int. Conf. Big Data, Big Data 2019, pp. 1971–1980, 2019, doi: 10.1109/BigData47090.2019.9006104.

R. Indraswari, “ScienceDirectScienceDirect Melanoma based on on MobileNetV2 network Melanoma image image classification classification based MobileNetV2 network,” ProcediaComput. Sci., vol. 197, pp. 198–207, 2022, doi: 10.1016/j.procs.2021.12.132.

[20] P. Barham et al., “TensorFlow : A system for large-scale machine learning,” pp. 265–284.

J. Deng, W. Dong, R. Socher, L. Li, K. Li, and L. Fei-fei, “ImageNet : a Large-Scale Hierarchical Image Database ImageNet : A Large-Scale Hierarchical Image Database,” no. May 2014, 2009, doi: 10.1109/CVPR.2009.5206848.

M. Hasnain, M. F. Pasha, I. Ghani, M. Imran, M. Y. Alzahrani, and R. Budiarto, “Evaluating Trust Prediction and Confusion Matrix Measures for Web Services Ranking,” pp. 90847–90861, 2020, doi: 10.1109/ACCESS.2020.2994222,