Deep Learning Approach for Masked Face Identification

Maad Shatnawi, Nahla Almenhali, Mitha Alhammadi, Khawla Alhanaee
Department of Electrical Engineering Technology, Higher Colleges of Technology, Abu Dhabi, UAE

***Abstract*—Covid-19 is a global health emergency and a major concern in the industrial and residential sectors. It has the ability to spread leading to health problems or death. Wearing a mask in public locations and busy areas is the most effective COVID-19 prevention measure. Face recognition provides an accurate method that overcomes uncertainties such as false prediction, high cost, and time consumption, as it is understood that the primary identification for every human being is his face. As a result, masked face identification is required to solve the issue of recognizing individuals with masks in several applications such as door access systems and smart attendance systems. This paper offers an important and intelligent method to solve this issue. We propose deep transfer learning approach for masked face human identification. We created a dataset of masked-face images and examined six convolutional neural network (CNN) models on this dataset. All models show great performance in terms of very high face recognition accuracy and short training time.**

Keywords—Masked face human identification; face recognition; deep transfer learning; convolutional neural networks

I. Introduction

Covid-19 is a global pandemic that began in the year 2020 and has triggered a global health crisis and impacted a wide range of businesses, including education, aviation, health care, tourism, luxury shopping, religion, and so on. It had a significant impact on people's daily lives as well. Wearing a face mask in public areas, according to the World Health Organization (WHO), is one of the most effective preventative measures to stop the spread of disease and save lives [1]. Furthermore, several public service providers restrict customers to utilize their services while wearing facemasks that meet specific standards. However, due to the face mask covering the majority of the crucial facial characteristics, such as the nose and mouth, conventional face recognition systems used for security have proven ineffective in the current circumstance making it exceedingly challenging to identify the person [2].

The unlocking techniques based on passwords or fingerprints are risky since the COVID-19 virus can be transmitted through contact. Without touching, face recognition makes it considerably safer, but when wearing a mask, existing face recognition technologies become unreliable. To address the current challenges, it is essential to enhance the current face recognition techniques, which primarily rely on all facial feature points, so that identity verification may still be carried out with reliability even when faces are only partially revealed. [3]These systems recognize people identities without the need to take off the mask which can be used in hospitals, offices, educational institutes, construction sites, manufacturing plants, airports, and in many other places. Masked-face recognition can also be used in attendance systems in schools, offices, and other working places.

Extensive studies enable the establishment of new datasets or masked face recognition. These studies significantly contribute since datasets of masked faces are correspondingly few and make it difficult to train suggested algorithms. There are several machine learning and deep learning techniques used to identify a face mask, some of which are based on enhanced pre-trained and existing models that all perform effectively [4].

Machine learning is a rapidly growing field of computational intelligence that seeks to emulate human learning from their environment. Computer vision, image recognition, speech recognition, natural language processing, and bioinformatics are some domains of machine learning. It is currently a primary competence of computer scientists. It includes computationally intensive methods, and is already widely utilized in the social sciences, marketing, systems engineering, and applied sciences. The degree of complexity of these systems can vary, and they may comprise multiple phases of complex human-machine interactions and decision making, which would naturally draw machine learning algorithms to enhance and automate various procedures. Numerous systems' efficiency and safety may be increased if machine learning algorithms had the potential to generalize from their current context and learn new tasks [5].

Deep learning is a type of machine learning that mimics the human brain's data processing to detect objects, recognize speech, translate languages, and make decisions. It is learning from both labeled and unlabeled data without the need for human interaction. Deep learning's capacity to handle large amounts of information makes it extremely powerful when dealing with unstructured data [6]. Deep learning has been effectively used to handle a broad variety of issues in the fields of image identification and natural language processing. It is based on neural network architectures with several layers of processing units [7].

Transfer learning is a type of machine learning in which a model is designed for one task and then utilized as the starting point for another task to be modified. People's ability to logically use previously acquired knowledge to solve new issues more quickly or effectively is what drives the study of transfer learning. Transfer learning is particularly beneficial in science as most real-world problems involve big amount of labeled data, demanding complicated models. Transfer learning can be used by developers to combine multiple applications into one. Developers can also train new models for difficult applications quickly. It's also a useful tool for improving the accuracy of computer vision models [8].

This paper proposes a deep transfer learning approach for masked face recognition. We trained and tuned six convolutional neural network (CNN) models based on a masked face dataset.

II. Pre-trained Networks

Pre-trained networks have various characteristics that are important to consider when choosing a network to perform a certain task. The most important performance measure of a network is accuracy, computational time, and memory size. Choosing a network is usually alternated between these functions. In this work, we have examined six pre-trained convolutional neural network models which are briefly explained in this section.

A. SqueezeNet

SqueezeNet is an eighteen-layer deep convolutional neural network that can classify images into 1000 different classes. The network has learned sophisticated function representations for a wide spectrum of pictures. The goal of utilizing SqueezeNet is to create a smaller neural network with fewer datasets that can be more readily integrated into computer memory and transmitted across a computer network [9]. SqueezeNet architecture is illustrated in Fig. 1.

B. GoogleNet

GoogleNet has twenty two layers with twenty seven levels dedicated to pooling. There are a total of nine initiation modules layered in a linear fashion. It also employs eleven convolution filters. The parallel network architecture significantly contributes to minimize computational time and memory requirements. GoogleNet architecture is illustrated in Fig. 2.

C. AlexNet

AlexNet is one of the most widely used neural network designs nowadays. It's been used to train and classify millions of object images. The network input is 227×227×3 RGB images. AlexNet solves the over-fitting issue by using dropout layers. AlexNet architecture is illustrated in Fig. 3.

D. ResNet-50

ResNet-50 is a 50-layer deep convolutional neural network that has been trained on over a million of images. The network has been pre-trained to classify images into 1000 different classes. The network input is 224×224×3 RGB images [10]. The network architecture is illustrated in Fig. 4.

E. VGG-16

VGG-16 is one of the very deep convolutional networks for large-scale image recognition. In ImageNet, a dataset with over 14 million pictures divided into 1000 classes, VGG-16 model achieves 92.7% classification accuracy. The network input is 224×224×3 RGB images [11]. The network architecture is illustrated in Fig. 5.

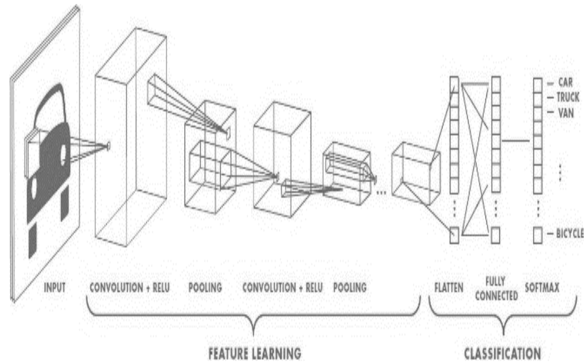

Fig. 1. SqueezeNet Architecture.

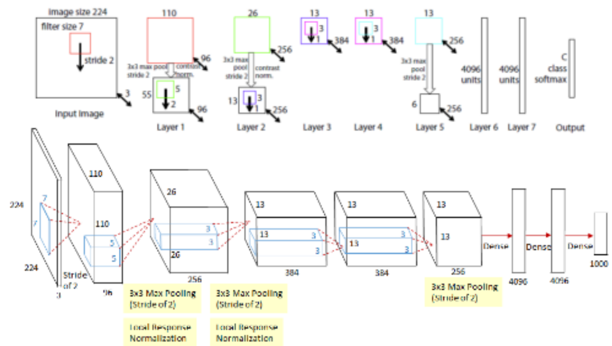

Fig. 2. GoogleNet Architecture.

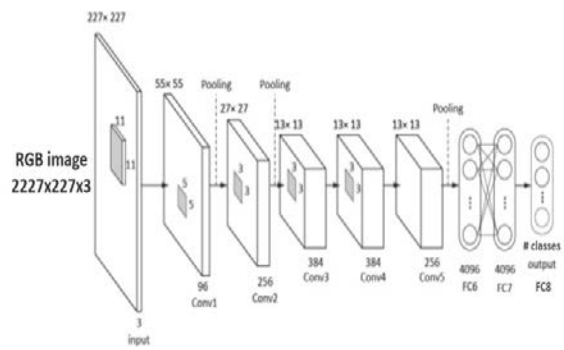

Fig. 3. AlexNet Architecture.

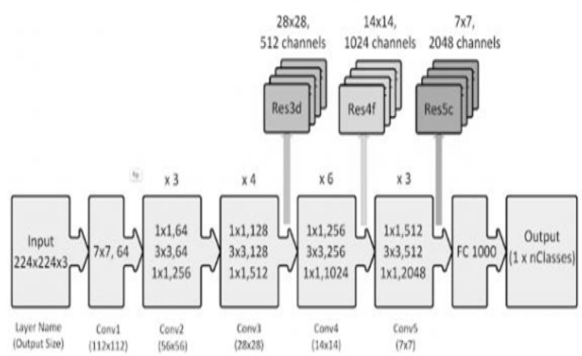

Fig. 4. ResNet-50 Architecture.

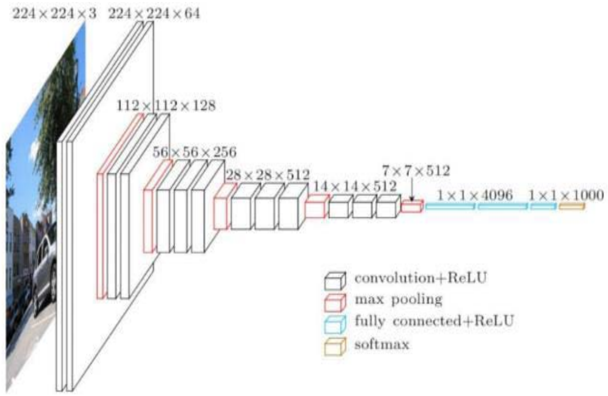

Fig. 5. VGG-16 Net Architecture.

F. *MobileNet-V2*

MobileNet-V2 was one of the first initiatives to create CNN architectures that could be readily implemented in mobile apps. Depth-wise separable convolutions, which are illustrated here, are one of the key advances. A separable convolution divides a single convolution kernel into two. Instead of a 3x3 kernel, we get a 3×1 and a 1×3 kernel, for example. It scales the image's input size from 224 to 128 pixels. You can train your MobileNet-V2 on 224×224 pictures and then use it on 128×128 images since it utilizes global average pooling instead of flattening [12]. The network architecture is illustrated in Fig. 6.

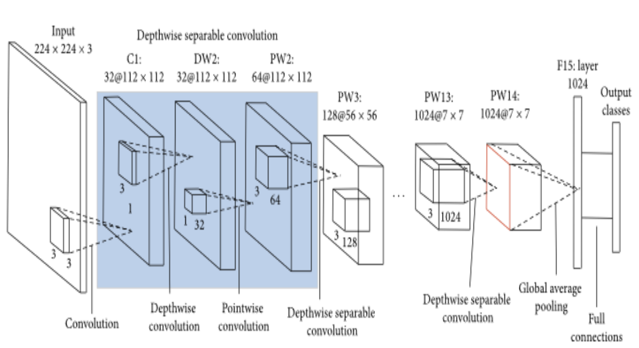

Fig. 6. MobileNet-V2 Architecture.

III. Current Approaches

The current face mask detection approaches are summarized and discussed in this section.

Cabani et al. [13] presented a dataset of correctly and incorrectly masked face images. Each set has 49 randomly sampled images, with 70,000 high-quality images in PNG and 1024 ×1024 resolution. The masked Face-Net dataset contained 139,646 images with 49% of correctly masked faces and 51% of incorrectly masked faces.

Mingjie et al. [14] proposed a Retina Face Mask method that uses transfer learning. The proposed method is a one-stage detector that combines a feature pyramid network with several feature maps to fuse high-level semantic information with a unique context attention module to focus on identifying face masks. The dataset has been split into a train, validation, and test set with 4906, 1226, and 1839 images individually; each experiment operates 250 epochs. The Advantage of the method MobileNet-V2 backbone shows that transfer learning can increase the detection performance by 3 to 4% in face and mask detection results. Also, feature extraction ability is enhanced by utilizing pre-trained weights.

Loey et al. [15] presented a hybrid model with deep transfer learning and classical machine learning for face mask detection. The machine learning approach uses decision trees and Support Vector Machines (SVM) to classify face masks. Three face masked datasets have been selected for investigation. The decision trees classifier achieved 96% of validation accuracy. The SVM classifier achieved 98% of validation accuracy with the least time consumed in the training process.

Chavda et al. [16] presents a CNN-based architecture for detecting instances of incorrect use of face masks. The system uses transfer learning technique with three nets which are MobileNetV2, DenseNet121, and NASNet. The system classifies images into three classes: correct face mask covering, incorrect face mask covering and no face mask. The dataset is used from a pre-trained model in Kaggle for quick generalization and stable detection. The results and accuracy percentage of each net are: MobileNetV2 is 99.12%, DenseNet121 is 99.40%, NASNet is 99.13%. The advantages are high accuracy while having using less training data, the inference time is significantly shorter, and the system can be integrated with an image or video capturing device. On the other hand, the disadvantages are having a motion blur, dynamic focus and frame transition may face in video feed analysis.

Hariri [17] presents a masked face recognition technique based on occlusion removal and deep learning-based features. The system utilizes three pre-trained deep Convolutional Neural Networks (CNNs) to extract deep features from the resulting regions which are: VGG-16, AlexNet, and ResNet-50. This system is using two datasets which are, Masked Face Recognition Dataset (RMFRD), and Simulated Masked Face Recognition Dataset (SMFRD) that are from an open source [17]. The advantages are that the system obtained high accuracy compared to other face recognizers due to the best features extracted from the last convolutional layers of the pre-trained models, and the high efficiency. Recognition performance on RMFRD dataset obtained 91.3% accuracy, and SMFRD dataset is 88.9%.

Nagrath et al. [18] presented a method that uses deep learning, TensorFlow, Keras, and OpenCV to detect face masks. The SSDMNV2 method employs the Single Shot Multibox Detector as a face detector and the MobilenetV2 architecture as a framework for the classifier. The dataset used in this method is from Kaggle’s Medical Mask Dataset (678 images) and the dataset available at PyImageSearch (1,376 images; wearing masks 690 pictures, and without wearing a mask, 686 pictures). The advantages are that SSDMNV2 is very lightweight and can even be used in embedded devices, it is useful method for real-time mask detection, and it achieved a high accuracy score of 92% which indicated a well-trained model.

Aswal et al. [19] presented two techniques that are used to recognize and identify masked faces. Using a single camera: A two-step process with a pre-trained one-stage feature pyramid detector network RetinaFace for localizing masked faces and VGGFace2 that generates facial feature vectors for efficient mask face verification; and (ii) a single-step pre-trained YOLO-face/trained YOLOv3 model on the set of known individuals. The dataset consists of 17 videos of individuals wearing face masks nearby. The dataset contains variations in orientation, scale, occlusions, illumination, and appearance. On their dataset, experimental results show that RetinaFace and VGGFace2 deliver state-of-the-art results in 1:1 face mask verification, with 92.7% overall performance, 98.1% face detection, and 94.5% face verification accuracy, respectively. The advantage of using transfer learning on YOLOv3 pre-trained the dataset is that the training process requires fewer images per class. And the YOLOv3 is a more advanced version of YOLO, Darknet-53, with shortcut connections for better GPU utilization, efficient evaluation, and faster performance. In comparison, the YOLO-face is to increases face detection performance at a fast detection speed.

Ding et al. [20] constructed two datasets created for masked faces recognition (MFR) for masked faces verification (MFV), which contains 400 pairs of 200 identities for verification, and masked faces identification (MFI), which contains 4,916 images of 669 identities for identification. A latent part detection (LPD) method is also developed for locating the latent facial part that is robust to mask wear. Extensive tests on the MFV, MFI and synthetic masked LFW data show that the LPD model outperforms alternative approaches. They use the LPD network, which has a level of accuracy of 94.34 %. There are two major drawbacks, when using these approaches to solve the MFR problem directly. First, the PFR techniques need the use of pre-defined partial face images as inputs, which are difficult to identify or define semantically in masked faces. Second, the PRF techniques disregard global signals such as chin contours in masked face images.

Most of the current approaches aims to detect whether an individual is wearing a face mask or not or to detect if the mask is worn in a correct way or not. There are few masked-face human identification approaches. This work proposes a deep transfer learning approach for masked face recognition that outperform existing approaches in terms of identification accuracy.

IV. METHOD

A. Dataset

We collected a dataset of 400 RGB images for ten human masked faces. Each individual class includes 30 images for the training and validation, and 10 images for the testing. Examples of the masked face data are illustrated in Fig. 7. The image sizes are in the range of 1.5 and 4 MB. We resized the images to match the required input dimensions of each of the six networks. SqueezeNet and AlexNet use 227 × 227 pixels, while GoogleNet, VGG16, and ResNet-50 use 224 × 224 pixels. However, MobileNet-V2 uses an input size of 128 to 224 pixels.

Data augmentation is a technique used in machine learning to expand the quantity of data by introducing slightly modified copies of available image data or newly created artificial data from real data. It functions as a preprocessing step and aids in the training of a machine learning model in order to reduce overfitting. Geometric transformation, flipping, color altering, cropping, rotation, noise injection, and random erasure are all examples of data augmentation techniques. We used data augmentation in our trained networks by acquiring many photos from various perspectives, surroundings and conditions, orientation, position, and brightness. The objective for this is to ensure that the output contains accurate predictions. We also applied several geometric transformations to the existing images to generate synthetic images. These transformations are random rotation in the range of -90° and +90° and random scaling in the range of 1 and 2.

B. CNN Model Training and Validation

We trained and evaluated six pre-trained CNN models which are SqueezeNet, GoogleNet, AlexNet, ResNet-50, VGG-16, and MobileNet-V2. The training was performed using a single CPU. The validation was done in a 10-iteration process to ensure that the system is trained well without overfitting.

SqeezeNet training takes 33 minutes and 46 seconds. The network achieved a validation accuracy of 97.78% as illustrated in Fig. 8. GoogleNet training takes 46 minutes and 59 seconds and achieved a validation accuracy of 100% as illustrated in Fig. 9.

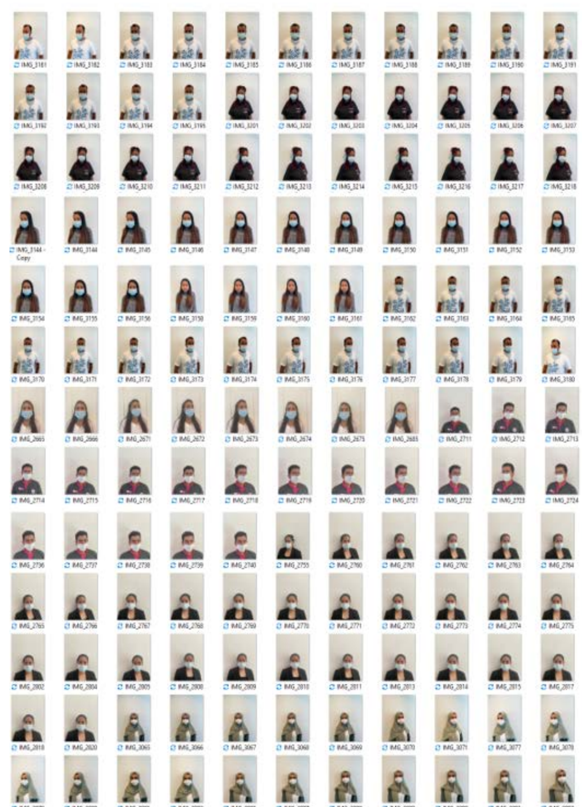

Fig. 7. Sample of Masked Face Dataset.

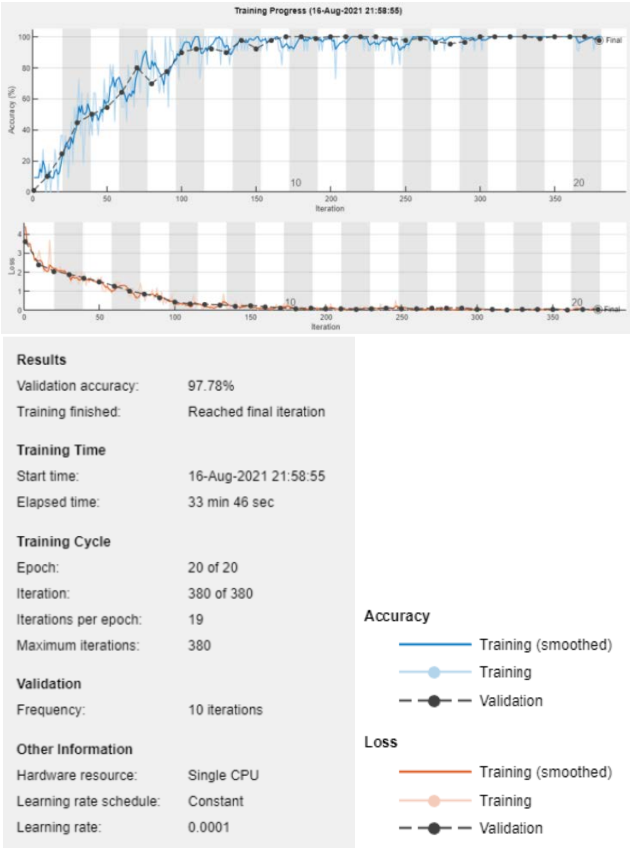

Fig. 8. SqueezeNet Training and Validation.

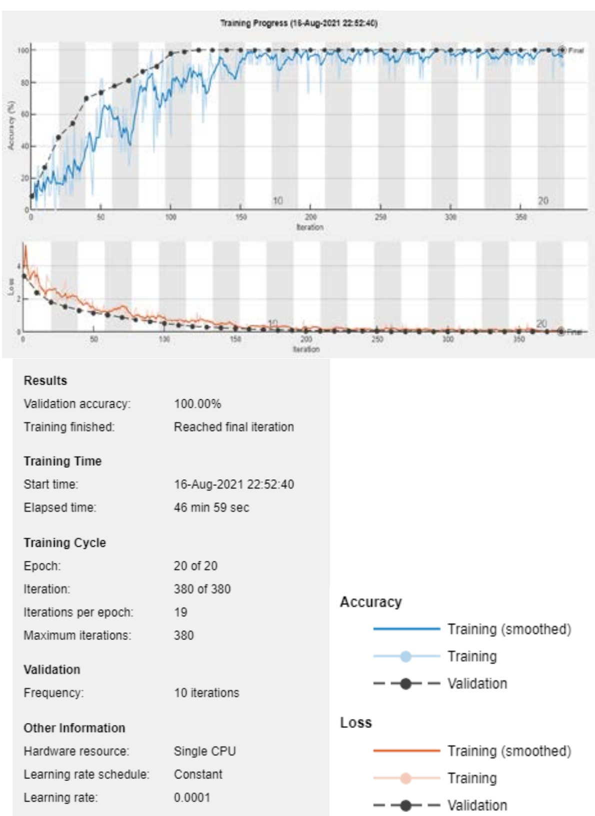

Fig. 9. GoogleNet Training and Validation.

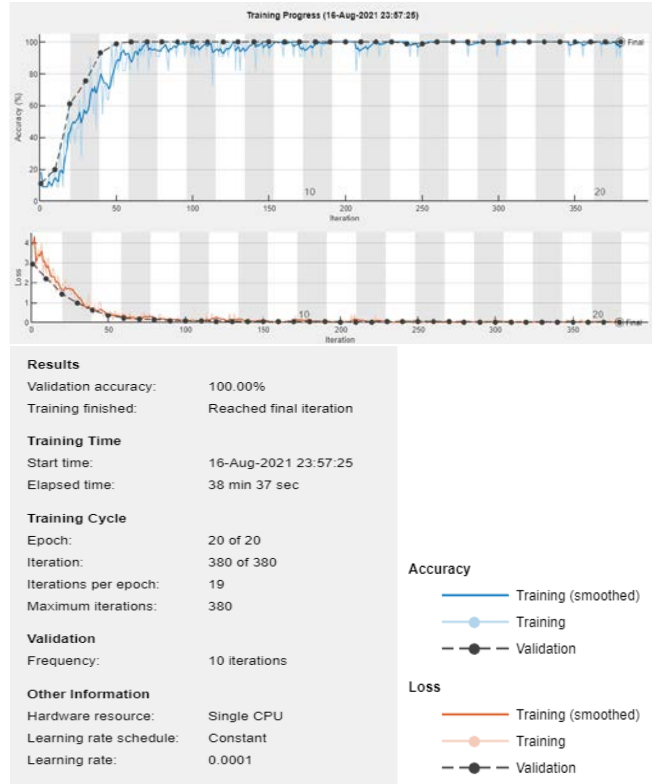

Fig. 10. AlexNet Training and Validation.

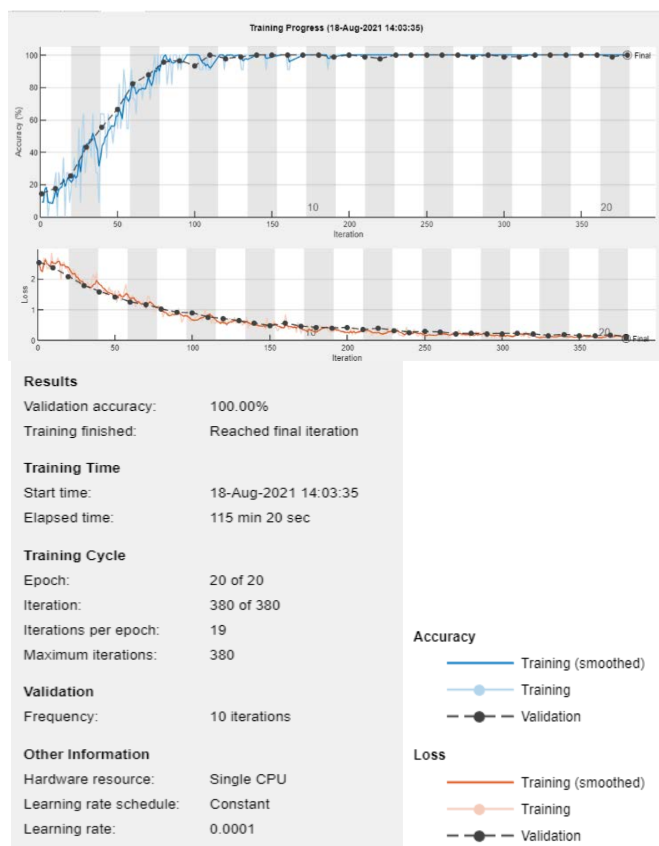

Fig. 11. ResNet-50 Training and Validation.

AlexNet training takes 38 minutes and 37 seconds and achieved a validation accuracy of 100% as illustrated in Fig. 10. Training the ResNet-50 network takes 115 minutes and 20 seconds and achieved a validation accuracy of 100% as illustrated in Fig. 11. Training the VGG-16 network takes 148 minutes and 50 seconds and achieved a validation accuracy of 100% as illustrated in Fig. 12. Training the MobileNet-V2 training takes 62 minutes and 20 seconds. The network achieved a validation accuracy of 100% as illustrated in Fig. 13.

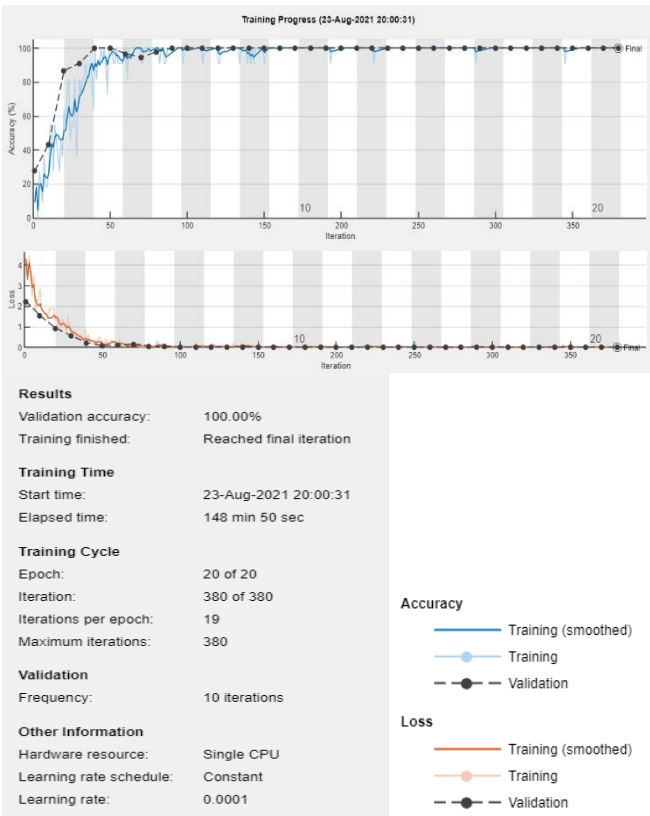

Fig. 12. VGG-16 Training and Validation.

The training performance of the six CNN is summarized in Table I. According to the results, most of the nets achieved a very high accuracy, but when the training time is considered, AlexNet has the shortest training time. GoogleNet is the second-best model which achieved an accuracy of 100% with a reasonable training time of 46 mins 59 sec. However, MobileNet-V2 is the third-best model which achieved a validation accuracy of 100% with a good training time which is 62 mins 20 sec. Moreover, the fourth-best network is ResNet-50 as it achieved validation accuracy of 100% but in a long training time of 115 mins 20 sec. The fifth network is VGG-16 that has achieved high validation accuracy of 100% but in a very long training time. The lowest network performance was observed is in SqueezeNet because it provides the lowest validation accuracy among the six networks, but it took a very good training time, which is 33 minutes and 46 seconds.

C. CNN Model Testing

The six CNN models were tested on unseen images from each of the ten classes; we tested the system on ten images for each class, the models output the predicted label and the confidence level of each prediction. Fig. 14 illustrates samples of testing results of each of the six models on three unseen images. All the six models have successfully identified all the images correctly with very high confidence level indicating that the six models achieve 100% recognition rate. The overall testing results are summarized in Table II. In the presence of the mask, the suggested method enhances the generalization of the face recognition process.

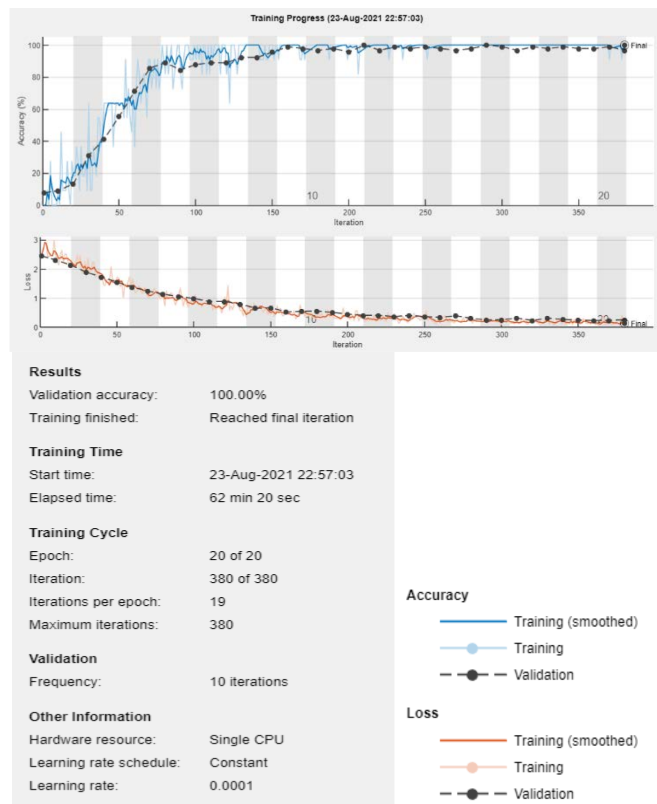

Fig. 13. MobileNet-V2 Training and Validation.

TABLE I. COMPARISON OF CNN TRAINING PERFORMANCE

Model	Learning rate	Epochs	Validation Accuracy	Elapsed Time	Hardware Resource	Max. iterations
SqueezeNet	*0.0001*	*20*	*97.78%*	*33 mins 46 sec*	*Single CPU*	*380*
GoogleNet	*0.0001*	*20*	*100%*	*46 mins 59 sec*	*Single CPU*	*380*
AlexNet	*0.0001*	*20*	*100%*	*38 mins 37 sec*	*Single CPU*	*380*
ResNet-50	*0.0001*	*20*	*100%*	*115 mins 20 sec*	*Single CPU*	*380*
VGG-16	*0.0001*	*20*	*100%*	*148 mins 50 sec*	*Single CPU*	*380*
MobileNet-V2	*0.0001*	*20*	*100%*	*62 mins 20 sec*	*Single CPU*	*380*

(a) SqueezeNet

(b) GoogleNet

(c) AlexNet

(d) ResNet-50

(e) VGG-16 Net

(f) MobileNet-V2

Fig. 14. Samples of Testing Results.

TABLE II. TESTING RESULTS

Model	**Name**	**Results**			**Testing Accuracy**	**Percentage Confidence level**		
						Max	***Min***	***Average***
SqueezeNet	Annet	100	100	100	100%	100	100	100
	Ella	96.8	98.8	99.4	100%	99.4	96.8	98.33
	Krisdna	100	99.7	97.4	100%	100	97.4	99.03
	Mouza	100	99.9	100	100%	100	99.9	99.97
	Priya	100	84.2	96.4	100%	100	84.2	93.53
	Ronalie	100	100	100	100%	100	100	100
	Saddam	99.1	100	100	100%	100	99.1	99.70
	Salah	84.8	63.3	99.2	100%	99.2	63.3	82.43
	Wafa	100	100	100	100%	100	100	100
	Zeeshan	97.6	97.6	98.9	100%	98.9	97.6	98.03
GoogleNet	Annet	99.9	93	99.7	100%	99.9	93	97.53
	Ella	66.4	94	99.4	100%	99.4	66.4	86.60
	Krisdna	80.9	99.6	98.8	100%	99.6	80.9	93.10
	Mouza	99.6	99.7	99.9	100%	99.9	99.6	99.73
	Priya	89.6	97.5	97.5	100%	97.5	89.6	94.87
	Ronalie	99.3	100	100	100%	100	99.3	99.77
	Saddam	97.2	99.7	98.4	100%	99.7	97.2	98.43
	Salah	40.9	99.8	97.9	100%	99.8	40.9	79.53
	Wafa	95.6	98	99.6	100%	99.6	95.6	97.73
	Zeeshan	98.9	97	98.8	100%	98.9	97	98.23
AlexNet	Annet	100	100	100	100%	100	100	100
	Ella	46.4	100	99.9	100%	100	46.4	82.10
	Krisdna	90.7	99.9	100	100%	100	90.7	96.87
	Mouza	100	99.5	100	100%	100	99.5	99.83
	Priya	99.7	100	42.6	100%	100	42.6	80.77
	Ronalie	99.9	100	99.9	100%	100	99.9	99.93
	Saddam	99.9	99.9	100	100%	100	99.9	99.93
	Salah	99.6	77.7	100	100%	100	77.7	92.43
	Wafa	99.8	99.7	96.8	100%	99.8	96.8	98.77
	Zeeshan	63.3	98.4	99.8	100%	99.8	63.3	87.17
ResNet-50	Annet	96.9	93.2	80.9	100%	96.9	80.9	90.33
	Ella	76	99.9	96.9	100%	99.9	76	90.93
	Krisdna	76.4	85	79	100%	85	76.4	80.13
	Mouza	91.2	98	92.4	100%	98	91.2	93.87
	Priya	78.4	81.3	95	100%	95	78.4	84.90
	Ronalie	99.3	98.8	98.6	100%	99.3	98.6	98.90
	Saddam	67.2	95.4	95.7	100%	95.7	67.2	86.10
	Salah	92.6	50.6	96	100%	96	50.6	79.73
	Wafa	96.8	92.6	93	100%	96.8	92.6	94.13
	Zeeshan	49.4	63	79	100%	79	49.4	63.80
VGG-16	Annet	100	99.8	100	100%	100	99.8	99.93
	Ella	74.3	99.8	86.5	100%	99.8	74.3	86.87
	Krisdna	99.9	100	100	100%	100	99.9	99.97

	Mouza	100	100	100	100%	100	100	100
	Priya	99.9	100	100	100%	100	99.9	99.97
	Ronalie	100	100	100	100%	100	100	100
	Saddam	100	100	100	100%	100	100	100
	Salah	42.1	100	100	100%	100	42.1	80.70
	Wafa	100	100	100	100%	100	100	100
	Zeeshan	97.9	100	100	100%	100	97.9	99.30
MobileNet-V2	Annet	95.1	97.4	96.2	100%	97.4	95.1	96.23
	Ella	56.9	99.3	85.4	100%	99.3	56.9	80.53
	Krisdna	90.9	83.3	71.6	100%	90.9	71.6	81.93
	Mouza	87.2	88.2	93.3	100%	93.3	87.2	89.57
	Priya	75.4	90.2	90.9	100%	90.9	75.4	85.50
	Ronalie	81.8	86.3	85.8	100%	86.3	81.8	84.63
	Saddam	77.9	91.7	95.5	100%	95.5	77.9	88.37
	Salah	47	86.9	72.1	100%	86.9	47	68.67
	Wafa	66	90.5	81.7	100%	90.5	66	79.40
	Zeeshan	82.7	92.8	95.2	100%	95.2	82.7	90.23

V. Conclusion

We propose a deep transfer learning approach for masked face recognition. We were able to find an appropriate pre-trained network models and utilize them as a starting point for fine-tuning and developing a deep neural network for a new image classification job using our training masked face data. We successfully utilized transfer learning to speed up training and increase the deep learning model's performance. We examined six convolutional neural network (CNN) models on this dataset. All models show great performance in terms of very high face recognition rate and short training time. The models are SqueezeNet, GoogleNet, AlexNet, ResNet-50, VGG-16, and MobileNet-V2.

These six models achieved validation accuracy between 97.8% and 100%. When tested on unseen data, all the models were able to identify all masked faces correctly with very high confidence level indicating that the six models achieve 100% recognition rate. The proposed approach is critical because it detects a person faster than the human eye, it is more accurate, and provides complete security. In the presence of the mask, the suggested method enhances the generalization of the face recognition process. The proposed approach can be applied in smart access systems and intelligent attendance systems in schools, hospitals, and various industries. The work can be expanded by evaluating more CNN models and by including more human masked-face image data.

The research opens up exciting new areas for study. The suggested method is not restricted to mask detection only and can be incorporated into any high-resolution video surveillance systems. A face mask may be used with the model to recognize facial landmarks for biometric applications. In future work, we anticipate being able to create our own network and provide it with the tools it needs to be well-trained to suggest a more extensive work.

References

[1] K. Suresh, M. Palangappa and S. Bhuvan, "Face Mask Detection by using Optimistic Convolutional Neural Network," in 2021 6th International Conference on Inventive Computation Technologies (ICICT), 2021, pp. 1084--1089.

[2] X. Li, "Masked Face Detection and Calibration with Deep Learning," IOP Science, vol. 2196, p. 10, 2022.

[3] Zhongyuan Wang; Guangcheng Wang; Baojin Huang; Zhangyang Xiong; Qi Hong; Hao Wu; Peng Yi; Kui Jiang;, "Masked Face Recognition Dataset and Application," Arxiv , no. 2, p. 3, 2020.

[4] R. Golwalkar and N. Mehendale, "Masked-face recognition using deep metric learning and FaceMaskNet-21," 2022. [Online]. Available: https://doi.org/10.1007/s10489-021-03150-3.

[5] E. Naqa, Issam, Murphy and M. J, What is machine learning?, Springer, 2015, pp. 3--11.

[6] M. Hargrave, "Deep Learning," Investopedia, 17 May 2021. [Online]. Available: https://www.investopedia.com/terms/d/deep-learning.asp.

[7] B. Antonio, F. Boris, T. David and C. Borja, "A Systematic Review of Deep Learning Approaches to Educational Data Mining," Complexity, vol. 2019, no. 1306039, p. 22, 2019.

[8] K. Alhanaee, M. Alhammadi, N. Almenhali and M. Shatnawi, "Face Recognition Smart Attendance System using Deep Transfer Learning," Procedia Computer Science, vol. 192, pp. 4093--4102, 2021.

[9] T. M. Ayyar, "A practical experiment for comparing LeNet, AlexNet, VGG and ResNet models with their advantages and disadvantages.," [Online]. Available: https://tejasmohanayyar.medium.com/a-practical-experiment-for-comparing-lenet-alexnet-vgg-and-resnet-models-with-their-advantages-d932fb7c7d17.

[10] Mathworks, "ResNet-50 convolutional neural network," [Online]. Available: https://www.mathworks.com/help/deeplearning/ref/resnet50.html.

[11] Neurohive, "VGG16 – Convolutional Network for Classification and Detection," [Online]. Available: https://neurohive.io/en/popular-networks/vgg16/.

[12] C. Lepelaars, "The Evolution Of Mobile CNN Architectures," 2020. [Online]. Available: https://wandb.ai/carlolepelaars/mobile_architectures/reports/The-Evolution-Of-Mobile-CNN-Architectures--VmlldzoyMDQ0ODQ.

[13] A. Cabani, K. Hammoudi, H. Benhabiles and M. Melkemi, "MaskedFace-Net--A dataset of correctly/incorrectly masked face images in the context of COVID-19," Smart Health, vol. 19, no. Elsevier, p. 100144, 2021.

[14] J. Mingjie, F. Xinqi and Y. Hong, "Retina facemask: A face mask detector," arXiv preprint arXiv:2005.03950, vol. 2, p. 2020.

[15] M. Loey, G. Manogaran, M. H. N. Taha and N. E. M. Khalifa, "A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic," Measurement, vol. 167, no. Elsevier, p. 108288, 2021.

[16] A. Chavda, J. Dsouza, S. Badgujar and A. Damani, "Multi-stage cnn architecture for face mask detection," in 2021 6th International Conference for Convergence in Technology (I2CT), IEEE, 2021, pp. 1—8.

[17] W. Hariri, "Efficient masked face recognition method during the covid-19 pandemic," rXiv preprint arXiv:2105.03026, 2021.

[18] P. Nagrath, R. Jain, A. Madan, R. Arora, P. Kataria and J. Hemanth, "SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2," Sustainable cities and society, vol. 66, p. 102692, 2021.

[19] V. Aswal, O. Tupe, S. Shaikh and N. N. Charniya, "Single Camera Masked Face Identification," in 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA), IEEE, 2020, pp. 57--60.

[20] F. Ding, P. Peng, Y. Huang, M. Geng and Y. Tian, "Masked face recognition with latent part detection," in Proceedings of the 28th ACM International Conference on Multimedia, 2020, pp. 2281--2289.