Contents lists available at ScienceDirect

Journal of King Saud University – Computer and Information Sciences

journal homepage: www.sciencedirect.com

A novel DeepMaskNet model for face mask detection and masked facial recognition

Naeem Ullah [a], Ali Javed [b,*], Mustansar Ali Ghazanfar [c], Abdulmajeed Alsufyani [d], Sami Bourouis [e]

[a] Department of Software Engineering, University of Engineering and Technology, Taxila 47050, Pakistan
[b] Department of Computer Science, University of Engineering and Technology, Taxila 47050, Pakistan
[c] School of Architecture, Computing and Engineering, University of East London, UK
[d] Department of Computer Science, College of Computers and Information Technology, Taif University, 11099, Taif 21944, Saudi Arabia
[e] Department of Information Technology, College of Computers and Information Technology, Taif University, 11099, Taif 21944, Saudi Arabia

ARTICLE INFO

ABSTRACT

Coronavirus disease (COVID-19) has significantly affected the daily life activities of people globally. To prevent the spread of COVID-19, the World Health Organization has recommended the people to wear face mask in public places. Manual inspection of people for wearing face masks in public places is a challenging task. Moreover, the use of face masks makes the traditional face recognition techniques ineffective, which are typically designed for unveiled faces. Thus, introduces an urgent need to develop a robust system capable of detecting the people not wearing the face masks and recognizing different persons while wearing the face mask. In this paper, we propose a novel DeepMasknet framework capable of both the face mask detection and masked facial recognition. Moreover, presently there is an absence of a unified and diverse dataset that can be used to evaluate both the face mask detection and masked facial recognition. For this purpose, we also developed a largescale and diverse unified mask detection and masked facial recognition (MDMFR) dataset to measure the performance of both the face mask detection and masked facial recognition methods. Experimental results on multiple datasets including the cross-dataset setting show the superiority of our DeepMasknet framework over the contemporary models.

1. Introduction

Coronavirus disease (COVID-19) (Wang et al., 2020) was caused by the propagation of a novel coronavirus (SARS-CoV-2) that began in Wuhan, China and was quickly labelled a global pandemic by the World Health Organization (WHO). Since the emergence of COVID-19, 220,563,227 COVID positive cases and 4,565,483 deaths have been reported globally till September 6, 2021 (WHO, 2021). The arrival of vaccines initially brought some relief among the people around the world, however, cases of COVID-19 infections in vaccinated people raised an alarming situation and importance of following the SOPs (i.e., wearing face masks, social distancing, etc.)

recommended by the WHO. Before this pandemic, face masks were commonly used by the people to protect themselves from air pollution (World Health Organization, 2020; Al-Ramahi et al., 2021) or paramedical staff in the hospitals or to hide their identities for committing the crimes, etc. During the pandemic, however, individuals all across the world must wear face masks in public areas to prevent the spread of COVID-19. Presently, COVID-19 dispersal control is a huge challenge for WHO policy makers and the entire humanity. The predominance of evidence from WHO (Greenhalgh et al., 2020), analysis (Balotra et al., 2019) and research (McDonald et al., 2020) reveals that wearing face mask reduces the expansion of COVID-19 by decreasing the probability of transmission of contaminated respiratory (virus-laden) droplets (Jayaweera et al., 2020; Eikenberry et al., 2020; Jayaweera et al., 2020). Therefore, many countries require people to wear face masks in public places to prevent the spread of COVID-19. Manually inspecting persons in public locations for wearing face masks is a difficult task. As a result, there is a need to create automated systems for detecting the face masks.

At one side, governments have to enforce their population to wear the face masks to avoid the rapid spread of COVID-19,

* Corresponding author.
  E-mail address: ali.javed@uettaxila.edu.pk (A. Javed).

Production and hosting by Elsevier

whereas, on the other side, wearing face mask also introduces new challenges for the traditional face recognition applications, which are typically designed for unveiled faces. These facial recognition applications installed at different checkpoints (Mundial et al., 2020; Damer et al., 2021) offer degraded performance in the case of masked faces due to the loss of significant information of face segments such as nose, lips, chin, cheeks, etc., (Du et al., 2021). The failure of facial recognition methods in case of face masks has raised significant challenges for the verification/authentication applications such as mobile payments, public safety examination, unlock phones and attendance, etc., (Kumar et al., 2021; Meenpal et al., 2019; Proença and Neves, 2018). For example, in the public security checkpoints such as railway and bus stations, the entry gates are equipped with installed cameras and rely on the conventional face recognition methods that are unable to recognize the persons with face masks. The ongoing COVID pandemic has discouraged the use of various traditional biometric-based methods such as fingerprint recognition, facial recognition, etc., as these methods can cause the spread of COVID-19 virus among the users. Automated user verification systems capable of recognizing the persons in the presence of face mask seems an effective solution in this pandemic. However, the existing facial recognition methods are not much robust for masked faces. Therefore, this pandemic situation demands an urgent need to develop robust facial recognition systems that are able to recognize any person while wearing the face mask. It is significant to improve the present face recognition methods that rely on all facial feature points, so that the reliable authentication can be performed in case of masked faces with less face exposure.

Recently, few studies have been presented for masked face recognition (Du et al., 2021; Ding et al., 2020; Geng et al., 2020), but still this research problem is less explored. One potential reason for such limited work in this domain is the unavailability of a standard masked face recognition dataset. Although, there exists multiple largescale and diverse unmasked facial recognition datasets (Trillionpairs, 2018; Guo et al., 2016; Cao et al., 2018; Yi et al., 2014) containing the face images of thousands of people. Still, to the best of our knowledge, there doesn't exist any standard and large-scale masked face recognition dataset. Unlike unmasked facial recognition datasets, the preparation of masked face datasets also requires the images of people with different masks on the face beside other diverse parameters i.e., variations in face angles, race, age, skin tone, lightning, and gender, etc. This also highlights the potential research gap of the unavailability of a robust and diverse masked face recognition dataset. *Moreover, according to the best of our knowledge, there doesn't exist any unified dataset that can be utilized to assess both the face mask detection and masked facial recognition applications.* This paper also addresses this issue by developing a unified mask detection and masked facial recognition (MDMFR) dataset to fill this existing research gap. Our in-house unified MDMFR dataset contains both the masked and unmasked facial images of 226 persons of different gender, age, skin tone, and face angles.

In existing research, the most frequently used methods for face mask detection and masked face recognition are transfer learning of pre-trained deep learning models and support vector machine (SVM). But SVM takes long training time and huge memory requirements for a large dataset (Zhang et al., 2005). Whereas in transfer learning, overfitting and negative transfer are the most alarming limitations (Zhao, 2017). Also, due to the lack of accurate labelled datasets, masked facial recognition is very challenging. To resolve these issues of transfer learning and SVM, we proposed a novel deep learning based model that can be used to reliably detect face masks (binary classification) and recognize masked faces (multiclass classification).

Existing research works on face mask detection and masked facial recognition have certain limitations, i.e., the absence of a unified method and dataset to tackle both problems, low accuracy of masked face recognition, less uncovered face exposure that makes it difficult to capture enough facial landmarks, etc. To address the aforementioned limitations of existing methods, we proposed a novel Deepmasknet model that is capable of reliable face mask detection and masked facial recognition. Our novel end-to-end Deepmasknet model automatically extracts the most discriminating features for reliable face mask detection and masked facial recognition. The proposed technique is robust to variations in face angles, lightning conditions, gender, skin tone, age, types of masks, occlusions (glasses), etc. We develop a largescale and diverse MDMFR dataset that can be used to measure the performance of both the face mask detection and masked facial recognition methods. Rigorous experiments were performed including the cross corpora evaluation to compare the performance of our proposed model with 9 deep transfer learning models for face mask detection and 8 models for masked facial recognition to prove the efficacy of our model.

The remaining paper is organized as follows. The related work is described in Section 2. We presented the motivation for the planned work and a description of it in Section 3. The details of our in-house dataset and experimental results were described in Section 4. Finally, Section 5 concludes our discussion.

2. Related work

In the era of the COVID-19 epidemic, face mask detection and masked facial recognition deployment face various operational challenges. As a result, we've recently seen a few research efforts focused on face mask detection and masked facial recognition.

Existing works (Geng et al., 2020; Ud Din et al., 2020; Li et al., 2020; Venkateswarlu et al., 2020; Qin and Li, 2020) on face mask detection can be categorized into conventional machine learning (ML) methods, deep learning (DL) based methods, and hybrid methods. Hybrid approaches contain algorithms that use both the deep learning and traditional machine learning (ML) based methods. The deep learning-based methods were utilized in the majority of studies on face mask detection, while conventional ML-based methods (Nieto-Rodríguez et al., 2015) are limited. In Nieto-Rodríguez et al. (2015), an automated system was designed that activates an alarm in the operational room when the healthcare workers do not wear the face mask. This system used the Viola and Jones face detector for face detection and Gentle AdaBoost for face mask detection. In Vijitkunsawat and Chantngarm (2020), two traditional ML classifiers i.e., KNN and SVM and one DL algorithm i.e., Mobilenet were evaluated to find the best model for detecting the face masks. The results confirmed that MobileNet achieves better performance over the KNN and SVM.

The significance of deep learning in computer vision has motivated the researchers to use it for face mask detection. In Chowdary et al. (2020), the InceptionV3 deep learning model was used to automate the mask detection process. To classify the images into mask and un-mask, the pre-trained InceptionV3 model was fine-tuned. The Simulated Masked Face dataset was used for the training and testing of the InceptionV3 model (SMFD). In Nagrath et al. (2021), SSDMNV2 deep learning based model was presented for face mask detection. The SSDMNV2 model used a single shot multibox detector for face detection and MobilenetV2 model for the classification of the mask and un-masked images. In Militante and Dionisio (2020), a real-time end-to-end network based on VGG-16 was presented for face mask detection. In Bu et al. (2017), a cascaded CNN framework comprises of three CNN

models was proposed for masked face detection. The first CNN model was comprised of five layers, whereas, second and third CNN models used seven layers each. The disadvantage of the three cascaded CNNs is the higher computational cost of this method.

Apart from the standalone deep learning models, existing works have also employed the hybrid models comprising of both the traditional ML and DL-based techniques. Loey et al. (2021) presented a hybrid model based on both the traditional ML and DL-based methods for face mask detection. This hybrid model employed the resnet50 algorithm for feature extraction and used them to train the SVM, ensemble algorithm, and decision tree for classification of images into mask and un-mask. In Bhattacharya (2021), a HybridFaceMaskNet model was proposed for detection of face masks. This model was based on deep learning, handcrafted feature extractor and classical machine learning classifiers. Due to the limited data, both deep learning (CNN) and handcrafted feature extraction (LBP, textural Harlick feature and Hue moments) techniques were applied to extract more robust features. Then features selection was performed through the Principal Component Analysis (PCA). Finally, the classification was accomplished using a random forest classifier.

Although, wearing the face masks are important to control the transmission of COVID-19, however, we also observe that people are often reluctant to cover both the lips and nose regions of the face. Thus, it reduces the effectiveness of face mask due to improper placement of the mask on the face. The research community also presented a few studies to monitor the placement of masks on the face. In Qin and Li (2020), an automated method was presented to identify the placement of the face mask, i.e., improper mask wearing, proper mask wearing, or no mask wearing. By integrating image super-resolution and classification networks, the scientists developed the SRCNet model, which enumerates a three-category classification issue based on unconstrained 2D facial pictures (SRCNet). In Inamdar and Mehendale, 2020, a Facemasknet model was proposed to verify the correct placement of the mask on the face. Face mask images were divided into three categories using the Facemasknet model: no mask, improperly worn mask, and properly worn mask. The model was trained using a total of 35 images, including both masked and non-masked faces.

Facial recognition is a commonly employed technique for user authentication-based applications. Deep face recognition has attained remarkable performance because of two reasons: first the availability of large-scale datasets e.g., VGGFace2 (Cao et al., 2018) and CASIA-Webface (Yi et al., 2014) and second the availability of sophisticated deep learning methods e.g., DeepFace (Taigman et al., 2014) and Center Loss-based CNN (Wen et al., 2016), etc. However, deep learning models used these datasets to learn the features from the complete uncovered facial images. Unlike the conventional unmasked recognition (Simonyan and Zisserman, 2014), masked facial recognition is very challenging due to the availability of limited uncovered facial exposure. In Sabbir Ejaz and Rabiul Islam (2019), the Google FaceNet network was employed to extract the features from the masked images and later SVM was used to recognize the candidate faces. In Ejaz et al. (2019), authors have used the PCA technique whereas the work in Deng et al. (2021) have used the Facenet deep learning-based technique for both the masked and unmasked facial recognition. It was observed in (Ejaz et al., 2019; Deng et al., 2021) that the performance drops by 23% and by a range of 4% to 11% respectively for masked facial recognition as compared to unmasked face recognition. According to Deng et al. (2021), masked facial recognition accuracy is 72% on average, while non-masked face image recognition accuracy is 95% on average. According to Deng et al. (2021), masked facial recognition accuracy is best (~97%), while non-masked face image recognition accuracy varies between 86% and 93%.

Occlusion is considered one of the key limitations of face recognition approaches. In the pre-pandemic era, most of the obstacles were glasses, clothes, helmet and other devices. However, in the post-pandemic era, masks are the most widely used obstacles in front of the face. Masked facial recognition has become a significant research problem recently and researchers are working in this domain. The traditional ML approaches, as well as deep learning approaches, have been employed for occluded face recognition. In Oh et al. (2008), the authors used a local approach based on PCA to detect the occlusions and Selective Local Non-Negative Matrix Factorization (S-LNMF) technique was employed for occluded face recognition. In (Min et al. (2011), the authors first detected the existence of occlusion i.e., scarf or sunglasses, etc., by using the Gabor wavelets, PCA and SVM, and later used the block-based local binary patterns to perform the recognition of non-occluded facial parts. Deep learning-based techniques automatically ignored the occluded parts and just highlights on the un-blocked or un-occluded ones. CNN is very effective for facial recognition applications specifically after the arrival of Alexnet (Krizhevsky et al., 2012). Alexnet has been found very useful for facial recognition in the presence of various occlusions (He et al., 2019). In Song et al. (2019), pairwise differential Siamese network (PDSN) based on the CNN was designed to counter the occlusion (i.e., identifies and removes corrupted feature elements) and only concentrates on the non-occluded facial regions. Firstly, the PDSN was used to generate a mask dictionary by altering the differences between the top characteristics of occlusion free and occluded face pairings. The items of the dictionary called Feature Discarding Mask (FDM) captured the correspondence between corrupted feature elements and occluded facial areas. When dealing with an occluded face image, its FDM was generated by combining the appropriate dictionary items.

Existing literature on face mask detection and masked facial recognition are unable to attain better classification and recognition performance because of different varieties of face masks beside other diverse parameters, i.e., variations in face angles, race, age, skin tone, lightning, and gender, etc. So, to address the limitations of both the face mask detection and masked facial recognition, a robust system and a unified dataset are urgently required.

3. Methodology

The proposed work comprises of two main phases: first phase includes the data collection and dataset preparation, while the second phase presents a novel Deepmasknet model construction for face mask detection and masked facial recognition. We proposed a novel Deepmasknet model that can effectively be used for both the face mask detection and masked facial recognition. Fig. 1 demonstrates the basic workflow of the proposed method. The model consists of 10 learned layers, i.e., six convolutional and four fully connected layer.

3.1. Motivations

Alexnet is the most widely used CNN architecture proposed by Krizhevsky and Sutskever (Krizhevsky et al., 2012). Alexnet has done a decent job at distinguishing persons who hide their identities by wearing scarves over their faces (Abhila and Sreeletha, 2018). Inspired from the Alexnet architecture, we proposed a deepmasknet model for face mask detection and masked facial recognition in this study. The goal of this research was to create a unified model that can detect the face masks and recognize the masked faces with higher accuracy. Our proposed model is based on the two most widely used CNNs scaling techniques i.e., scaling network depth and input image resolution. Depth scaling is used by
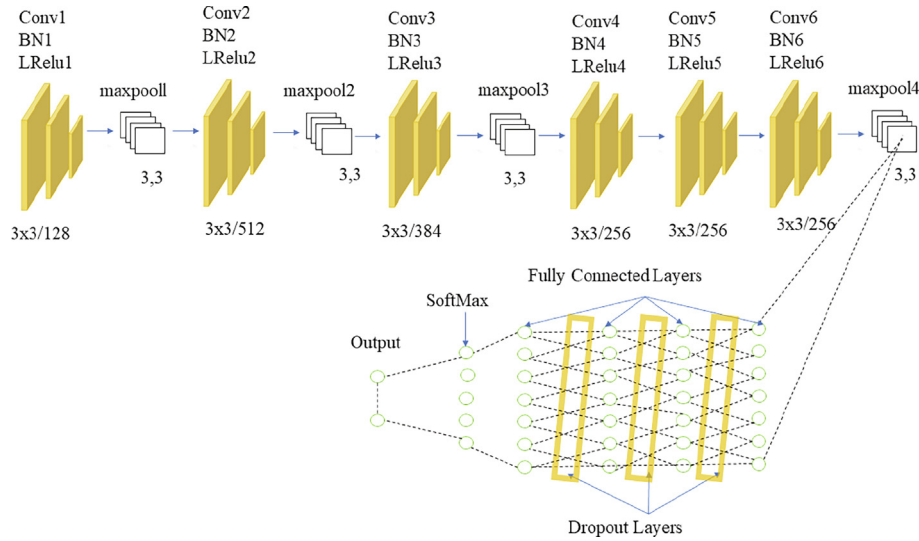
Fig. 1. Flow Diagram of face mask detection and masked face recognition.

many CNNs to achieve better accuracy (He et al., 2016; Huang et al., 2017). The obvious perception is that deeper CNNs capture more complex features and improves the classification performance of the network. However, as the network's depth grows, the computational complexity also increases without guaranteeing the improvement in accuracy for all cases. Also, CNNs capture more fine-grained features with higher resolution input images to achieve better performance.

CNNs accept input images of different resolutions i.e., 224 $\times$ 224, 297 $\times$ 297 and 299 $\times$ 299, but the architectures employing higher resolution tend to result in improved recognition performance (Szegedy et al., 2016). Our Deepmasknet model is ten layers deep and accepts an input image of 256 $\times$ 256 resolution for processing. The architecture of our Deepmasknet model and input image size is selected according to the available computing machine requirements. As the distribution of each layer's input changes during the training when the model is trained predominantly on one type of images, but the dataset contains slightly different images in a very low proportion. For example, in case of face masks, people often use surgical masks, so the dataset contains more surgical masks as compared to other types of masks i.e., N95, KN95, Cloth, etc. This causes the parameter training to be tremendously tedious and needs better initialization. To deal with this covariate shifting, we used batch normalization in DeepMaskNet architecture. And used the LeakyRelu (Maas et al., 2013) activation function instead of Relu activation function (Nair and Hinton, 2010) to overcome the dying relu problem. To add the non-linearity, Relu activation function is widely used between the layers to deal with more complex datasets. But Relu has certain issues. Firstly, gradient at x = 0 cannot be computed for Relu, which effects the training performance slightly. Secondly, all negative values (x < 0) are set to zero by the Relu activation function, so all the neurons with negative values are set to zero, i.e., neurons die and hence the issue is referred to as the dying Relu problem. The network stops learning in case of a dying Relu problem. Leaky relu is used in our Deepmasknet model to address these issues. The Leaky Relu activation function permits a minor, non-zero gradient when the unit is inactive. So, it would continue learning without reaching to the dead end.

3.2. DeepMaskNet architecture details

Assume that X is a collection of data samples (masked images) as illustrated below:

$$X = [X_1, X_2, X_3, \ldots, X_K] \tag{1}$$

In the image space $R^{I \times J}$, X_K signifies a sample datum, and K denotes the number of sample data. Assume that Y is a set of labels that correspond to X in the following way:

$$Y = [y_1, y_2, y_3, \ldots, y_K] \tag{2}$$

And in R^C, Y_K is X_K's label vector, where C is the number of classes.

$$Y_k^c = \begin{cases} 1 \ if \ X_K \in C \\ 0 \ if \ X_K \notin C \end{cases} \tag{3}$$

The purpose of image classification and recognition is to create a mapping function that maps image set X to label set Y using a training set and then categorizes fresh data points based on the learned mapping function. We present a novel deep learning-based Deepmasknet model to address the problem of image classification and recognition.

The architecture of our DeepMaskNet model is revealed in Fig. 2. The Deepmasknet model comprises ten learnable layers, i.e., six convolutional layers H^2, H^3, H^4, H^5, H^6, $H^{7,}$ and four fully connected layers as revealed in Table 1. The input layer H^1 is the first layer in Deepmasknet model, the size of the input layer H^1 is equal to the dimension of the input features, and it has $I \times J$ units. Our model accepts the input images of size 256 $\times$ 256 for processing.

To create the feature maps, the convolutional layer performs convolutional operations on image data and the output feature map is computed as follows:

$$h_{ij}^k = \sum ((w^k \times x) + b_k) \tag{4}$$

where k denotes the kth layer, h denotes the feature's value, (i,j) denotes pixel coordinates, w^k denotes the current layer's convolution kernel, and b_k denotes the bias.

When convolution operations are performed on an image of size $h \times$ w, with a kernel size of k, padding p, and stride size s, the output is of size $\frac{h-k+2p}{(s+1)} \times \frac{w-k+2p}{(s+1)}$. When the kernels are convolved with the image, they operate as feature detectors, resulting in a set of convolved features.

Activation functions are usually followed by the convolutional layers. In the past, the most widely used activation functions were *sigmoid* and *tanh*. However, because of its disadvantages, researchers have developed different activation functions like the rectified

Fig. 2. Architecture of our proposed DeepMaskNet model.

Table 1
The DeepMaskNet architecture.

Sr. No	Layer	Filters	Size	Stride	Padding
1	Input				
2	Convolutional-1 (Batch Normalization + LeakyRelu)	128	3×3	4×4	0×0
3	Max pooling		3×3	2×2	0×0
4	Convolutional-2 (Batch Normalization + LeakyRelu)	512	3×3	1×1	2×2
5	Max pooling		3×3	2×2	0×0
6	Convolutional-3 (Batch Normalization + LeakyRelu)	384	3×3	1×1	1×1
7	Max pooling		3×3	2×2	0×0
8	Convolutional-4 (Batch Normalization + LeakyRelu)	256	3×3	1×1	1×1
9	Convolutional-5 (Batch Normaslization + LeakyRelu)	256	3×3	1×1	1×1
10	Convolutional-6 (Batch Normalization + LeakyRelu)	256	3×3	1×1	1×1
11	Max pooling		3×3	2×2	0×0
12	Fully Connected + LeakyRelu + Dropout				
13	Fully Connected + LeakyRelu + Dropout				
14	Fully Connected + LeakyRelu + Dropout				
15	Fully Connected				
16	softmax				
17	classification				

linear unit (ReLU) and its derivatives (leaky ReLU, Noisy ReLU, ELU), which are now employed in most deep learning applications. The activation function specifies how a node in a layer converts the weighted sum of the input into an output. Here, we have used leaky ReLU to specify the ReLU activation function as an extremely small linear component of *x* instead of declaring it as 0 for negative values of inputs (*x*). This activation function is computed as follows:

$$f(x) = \max(0.01 \times x, x) \tag{5}$$

This method returns *x* if the input is positive, but a very small value if the input is negative, 0.01 times *x*. As a result, it also outputs the negative values. Additionally, down-sampling procedures are performed utilizing max-pooling layers to minimize the spatial size.

$$f(x) = \max(x_1, x_2, x_3, \cdots, x_k) \tag{6}$$

where $f(x)$ is optimized feature vector. Max-pooling is a down-sampling procedure that uses a kernel (k) and a stride (s) to extract the largest value from the current neighborhood (based on kernel size) on an image of size $h \times w$. The operation generates a file with the following dimensions: $\frac{h-k}{s+1} \times \frac{w-k}{s+1}$. The key purpose for inserting a max-pooling layer between the convolutional layers is to gradually decrease the size of the spatial representation, i.e., *h* and *w*, lower-

ing the number of parameters to train and the computations in the network as a whole. This also helps to prevent the issue of overfitting.

The first convolution layer of the proposed model extracts the features from a $256 \times 256 \times 3$ input image by applying 128 filters (kernels) of size 3-by-3 with a stride (shift) of 4 pixels at a time. The output of the first convolutional layer is sent to the second convolutional layer (after applying normalization and pooling). The subsequent convolutional layer applies 512 kernels of 3-by-3 with a stride (shift) of 1 pixel at a time. The third convolutional layer filters the inputs by applying 384 filters of size 3-by-3 which use a stride value of 1 pixel. The next two convolutional layers i.e., fourth and fifth are not followed by the pooling layers. On the input feature map, the fourth, fifth, and sixth convolutional layers apply 256 filters of size 3-by-3 with a default stride of 1 pixel. All six convolutional layers are proceeded with the batch normalization, leaky ReLU (nonlinear activation function), and maximum pooling layers. The architecture has four maximum pooling layers, the first three and last convolutional layers are followed by the maximum pooling layers.

Each fully linked layer's node is connected to all of the upper layer's nodes. To synthesize the features retrieved from the image and translate the two-dimensional feature map into a one-dimensional feature vector, fully linked layers are used. The sample

label space is mapped to the distributed feature representation via the fully connected layers. Equation (7) expresses the fully connected operation as:

$$a_i = \sum_{j=0}^{m \times n-1} w_{ij} \times x_i + b_i \tag{7}$$

where i represents the index of the fully connected layer's output; m, n, and d represent the width, height, and depth of the feature map derived from the last layer, respectively. Moreover, w and b represent the shared weights and bias, respectively. The first three fully connected layers are followed by leaky ReLU and dropout layers (to reduce overfitting or overlearning during training), whereas the last fully connected layer is proceeded by the softmax and classification layers. The output of the last fully connected layer is used as an input to a 2-way softmax in case of face mask detection (masked and unmasked face) and 226-way softmax in case of masked face recognition (226 different persons/subjects).

3.3. Training parameters

We used the trial-and-error based approach, i.e., experiments are performed using different values of the parameters to find the optimum value for each parameter. The details of the selected parameters are presented in Table 2. We performed the training of our Deepmasknet model by stochastic gradient descent (SGD) with final learning rate of 0.01 and a minibatch size of 10 images. The model is trained on 14 epochs for both the detection and recognition considering the occurrence of overfitting.

4. Results and discussion

We offer an in-depth explanation of the results of various experiments meant to assess the effectiveness of our DeepmaskNet model. This section also contains additional information about our in-house MDMFR dataset. Performance of our method is assessed using three publicly available Kaggle datasets (Facemask (Smansid, 2020), Facemask Detection Dataset 20,000 Images (Jain and Singaraju, 2020), and Facemask dataset (Shah, 2020), as well as our custom MDMFR dataset with cross-dataset settings.

4.1. Dataset

The unavailability of a unified standard dataset for face mask detection and masked facial recognition motivated us to develop an in-house MDMFR dataset (MDMFR, 2022) to measure the performance of face mask detection and masked facial recognition methods. Both of these tasks have different dataset requirements. Face mask detection requires the images of multiple persons with and without mask. Whereas, masked face recognition requires multiple masked face images of the same person. Our MDMFR dataset consists of two main collections, 1) face mask detection, and 2) masked facial recognition. There are 6006 images in our MDMFR dataset. The face mask detection collection contains two categories of face images i.e., mask and unmask. Our detection

Table 2
Parameters of DeepMaskNet model.

Parameter	Value
learning rate	0.01
Maximum Epochs	14
Shuffle	Every epoch
Validation frequency	30
Iterations per epoch	42
Verbose	false

database consists of 3174 with mask and 2832 without mask (unmasked) images. To construct the dataset, we captured multiple images of the same person in two configurations (mask and without mask). The masked facial recognition collection contains a total of 2896 masked images of 226 persons. More specifically, our dataset includes the images of both male and female persons of all ages including the children. The images of our dataset are diverse in terms of gender, race, and age of users, types of masks, illumination conditions, face angles, occlusions, environment, format, dimensions, and size, etc. Before being fed to our DeepMaskNet model, all images are scaled to a width and height of 256 pixels. All images have a bit depth of 24. We prepared the images of our dataset for the proposed DeepMaskNet model during preprocessing where images are cropped in Adobe-Photoshop to exclude the extra information like neck and shoulder. As the input size of our Deepmasknet model was 256-by-256, so images were resized to 256-by-256 in publicly available Plastiliq Image Resizer software (Plastiliq, 2022). Some representative images of our MDMFR dataset are shown in Fig. 3, and statistical details of our MDMFR dataset are provided in Table 3.

4.2. Evaluation metrics

To assess the effectiveness of our Deepmasknet model for face mask detection and masked facial recognition, we used precision, recall, F1-score and accuracy metrics. We computed these metrics as follows:

$$Accuracy = \frac{(TN + TP)}{TS} \tag{8}$$

$$Precision = \frac{TP}{(TP + FP)} \tag{9}$$

$$Recall = \frac{TP}{(TP + FN)} \tag{10}$$

$$F1 - score = 2 \times \frac{Precision \times Recall}{(Precision + Recall)} \tag{11}$$

True positive, total samples, false positive, true negative, and false negative are represented as TP, TS, FP, TN, and FN, respectively.

4.3. Experimental setup

All of our experimentations were carried out on a computer with an Intel (R) Core (TM) i5-5200U processor and 8 GB of RAM. We utilized the R2020a version of MATLAB to complete the project. Each dataset is divided into training and testing sets for all experiments. More specifically, we used 80% of the dataset for model training and 20% for testing. The performance of our proposed DeepMaskNet model for both the face mask detection and masked facial recognition is evaluated through a series of experiments.

4.3.1. Performance evaluation on face mask detection

The key purpose of this experiment is to assess the performance of our proposed model for face mask detection. For this experiment, we used all the 6006 images of our dataset, where 4805 images (2539 masked and 2266 unmasked) were used for training and rest 1201 images (635 masked and 566 unmasked) for testing. We used the training set to train our DeepMaskNet model on the same experimental settings mentioned in Table 2 for the classification of masked and unmasked face images. The training of our DeepMaskNet model took 192 min and 4 s for face mask detection. Our method achieved the optimal accuracy, precision, recall, and F1-score of 100% that shows the reliability of our method for face

Fig. 3. Sample images of our MDMFR dataset.

Table 3
Mask detection mask face recognition dataset details.

Samples	Male	Female	Types of masks used			Images details	
With mask	2923	251	Surgical	cloths	N95 and KN95	Bit depth	Type
Without mask	2625	206	2892	121	164	24	RGB
Total	5548	457					

mask detection. These results are due to the fact that our DeepMaskNet model successfully extracts the most discriminating features to represent the masked and unmasked faces for accurate classification.

To further assess the performance and verify the robustness of our DeepMaskNet model for face mask detection, we validated our model on the three standard and diverse Kaggle datasets (Smansid, 2020; Jain and Singaraju, 2020; Shah, 2020). Each dataset contains two types of images i.e., images with masks and images without masks. The details of these datasets and obtained results in terms of accuracy, precision, recall, and F1-score are given in the Table 4. From these results (Table 4), we can see that we achieved the optimal 100% accuracy, precision, recall, and F1-score on our MDMFR, facemask (Smansid, 2020) and Facemask Detection Dataset 20,000 Images (Jain and Singaraju, 2020) datasets. Whereas, we also achieved remarkable results on the Face-Mask Dataset (Shah, 2020). Regardless of the limited and imbalanced data in the FaceMask dataset, we achieved an accuracy of 98.51%. Therefore, we argue that the proposed DeepMaskNet model can reliably be used for face mask detection under diverse conditions.

4.3.2. Masked detection comparison with state-of-the-art deep learning models

The purpose of this experiment is to assess the effectiveness of the DeepMaskNet model for face mask detection over the contemporary deep learning models. For this purpose, we compared the performance of DeepMaskNet model with these contemporary models i.e., GoogleNet (Szegedy et al., 2015), SqueezeNet (Iandola et al., 2016), DenseNet (Huang et al., 2017), ShuffleNet (Zhang et al., 2018), ResNet (He et al., 2016), MobileNetv2 (Sandler et al., 2018), InceptionresNetv2 (Szegedy et al., 2017), DarkNet 19 (Redmon, 2022), and DarkNet 53 (Simonyan and Zisserman, 2014). All of these comparative deep learning models were used in a transfer learning setup trained on millions of images of the ImageNet database. All pre-trained versions of these networks can classify the images into 1000 different categories. The final layer was fine tuned to divide the images into two groups: with mask and without mask. The input image size of networks differs from model to model, like the image input size of VGG19 is 224-by-224, 227-by-227 for squeezenet, and so on. We tuned all of these models with the same experimental settings (mentioned in Table 2) as used for our model. The results are

Table 4
Performance evaluation on face mask detection.

Dataset	With mask images	Without mask images	Type	Accuracy %	Precision %	Recall %	F-score %
MDMFR	3174	2832	RGB	100	100	100	100
Facemask	690	686	RGB	100	100	100	100
Facemask Detection Dataset 20,000 Images	10,000	10,000	Gray Scale	100	100	100	100
FaceMask Dataset	208	131	RGB	98.51	100	97.56	98.76

presented in Table 5. Based on these findings, we observed that our model outperformed the other nine models by achieving 100% accuracy for face mask detection. The Mobilenetv2 model achieved the second-best accuracy of 99.82%, whereas, InceptionresNetv2 obtained the lowest accuracy of 97.43% among all models. Whereas, Resnet18 and Mobilenet outperformed other models in terms of true positive rate by achieving a recall of 100%. It is to be noted that six comparative models (DenseNet, ShuffleNet, ResNet, MobileNetv2, DarkNet 19, and DarkNet 53) achieved an accuracy of more than 99%. It is also worth noting that our proposed method obtained the optimal 100% accuracy, recall, precision, and F1-score. The Mobilenetv2 model achieved the second-best accuracy because Mobilenetv2 uses the concept of depth wise convolution and pointwise convolutions having 1×1 convolution which is used to capture the most important features. The linear bottleneck is used between the layers to prevent nonlinearities from abolishing greater loss of information. Inception-resNetv2 model achieved the lowest accuracy because of the use of ReLU activation function. ReLU activation function sets all the negative values (x < 0) to zero, so all the neurons with negative values are set to zero, thus giving no guarantee that all the neurons will be active throughout and causes the dying ReLU problem. The network does not learn through the optimization algorithm in this situation. Dying ReLU problem is undesirable because it gradually makes a big portion of the network inactive. The proposed DeepMaskNet model achieved the best results because of the ability to extract more descriptive, discriminative and robust deep features for face mask detection. These comparative results illustrate the superiority of our DeepMaskNet model over comparative models for face mask detection.

4.3.3. Face mask detection comparison with existing methods

The key purpose of this experiment is to validate the efficacy of proposed DeepMaskNet model over existing face mask detection methods. For this experiment, we compared our method with these approaches (Chowdary et al., 2020; Song et al., 2019). In Loey et al. (2021), the authors used the resnet50 model to extract the features and employed the SVM for the classification of masked and unmasked facial images. We tested this method on our own MDMFR dataset and achieved an accuracy of 98.78%. In Chowdary et al. (2020), inceptionV3 deep learning algorithm was presented for face mask detection. We also evaluated this method on our MDMFR dataset and attained an accuracy of 98.69%. In comparison to (Chowdary et al., 2020; Loey et al., 2021), our method achieved 100% accuracy.

This comparative analysis also demonstrates the effectiveness of our DeepMaskNet model over these approaches (Chowdary et al., 2020; Loey et al., 2021). It is important to mention that these approaches are computationally more complex than our proposed approach, because they employ deeper networks, which can easily lead to overfitting. These results evidently demonstrate the effectiveness of the suggested approach along with the supplementary

benefits, i.e., computationally efficient. As our proposed DeepMaskNet model comprises of only ten layers followed by the batch normalization operation and Leaky ReLU activation function, Thus, all biases of the CNN layers are not activated. Therefore, we can conclude that our approach is more efficient.

4.3.4. Performance evaluation on masked facial recognition

The purpose of this experiment is to assess the performance of proposed DeepMaskNet model for recognizing different persons while wearing the face mask. For this experiment, 2330 images of 226 classes of our MDMFR dataset were used. Among these image samples, 1864 were utilized for the training and 466 images were used for testing. We used the training set to train and testing set to validate DeepMaskNet model on the same experimental settings mentioned in Table 2 for the recognition of masked faces. The training process took 165 min and 43 s for masked face recognition. Our method achieved the accuracy of 93.33% for masked facial recognition. It is worthy to mention that although the masked facial images contain a limited visible portion of the face, yet our method is capable of extracting the distinctive traits from the uncovered facial image and recognizing the person with better accuracy. The proposed DeepMaskNet model was able to achieve the best results because of the ability to extract more discriminative, descriptive and robust deep features.

4.3.5. Masked facial recognition comparison with state-of-the-art models

The purpose of this experiment is to assess the effectiveness of the proposed DeepMaskNet model for masked face recognition over contemporary models. For this, we compared the masked facial recognition performance of our Deepmasknet model with eight state-of-the-art deep learning models i.e., DenseNet (Huang et al., 2017), ShuffleNet (Zhang et al., 2018), ResNet (He et al., 2016), MobileNetv2 (Sandler et al., 2018), DarkNet 19 (Redmon, 2022), DarkNet 53 (Redmon, 2022), VGG19 (Simonyan and Zisserman, 2014) and Alexnet (Krizhevsky et al., 2012). For this purpose, 2330 image samples of 226 classes were used. All of these comparative deep learning models were used in a transfer learning setup trained on the ImageNet database. To recognize the masked faces, last layers of all networks were fine tuned to classify images into 226 categories. We resized the input image size to meet the requirement of each comparative deep learning model. We tuned all of these models with the same experimental settings (mentioned in Table 2) as used for our model. The experimental findings are presented in Table 6. Based on our findings from these results, we observed that our model significantly outperforms the eight comparative deep learning models. More precisely, we achieved an accuracy gain of 2.42% from the second-best performing model (VGG-19) and 9.13% from the worst performing model (Resnet-18) for masked facial recognition. VGG-19 achieved the second-best results because the architecture uses very small (3×3)

Table 5
Detection performance results.

Model	Accuracy %	Precision %	Recall %	F1_score %
Deepmasknet	100	100	100	100
GoogleNet	98.08	98.51	99.25	98.87
Squeezenet	98.29	98.28	99.75	99.00
Densenet201	99.12	99.83	98.52	99.17
Shufflenet	99.82	100	99.66	99.82
Resnet18	99.12	98.33	100	99.15
Mobilenetv2	99.82	99.66	100	99.82
Inceptionresnetv2	97.43	98.50	98.50	98.50
Darknet19	99.75	97.72	100	98.84
Darknet53	99.51	98.48	97.02	97.74

Table 6
Recognition comparison with state-of-the-art models.

Model	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
Darknet19	90.00	90.00	91.5	90.74
Darknet53	89.1	87.96	87.43	87.69
Resnet18	84.20	83.65	85.31	84.47
VGG19	90.91	90.00	93.00	90.47
Shufflenet	88.57	88.5	88.6	88.54
Alexnet	88.89	90.00	90.00	90.00
Mobilenetv2	86.76	88.9	86.78	87.82
Densenet201	88.89	89.00	89.00	89.00
Deepmasknet (Proposed)	93.33	93.00	94.5	93.74

Table 7

Cross dataset validation results.

Training dataset	Testing dataset	Accuracy	Precision	Recall	F-score
Facemask Detection Dataset 20,000 Images	MDMFR	100	100	100	100
MDMFR	Facemask Detection Dataset 20,000 Images	99.99	99.99	100	99.99
MDMFR	FaceMask	99.99	99.99	100	99.99
FaceMask	MDMFR	100	100	100	100
FaceMask	Facemask Detection Dataset 20,000 Images	100	100	100	100
Facemask Detection Dataset 20,000 Images	FaceMask	100	100	100	100

convolution filters, which extract the most important features and shows significant improvement in accuracy for masked facial recognition. ResNet18 model achieved the lowest accuracy because of the use of ReLU activation function which gradually makes a big portion of the network inactive.

The reason behind achieving the best performance is that our proposed DeepMaskNet model effectively extracts more distinctive features from the face mask images. We applied the small filter of size 3×3 which guarantees the extraction of capturing detailed features. The batch normalization technique used in the feature map of the proposed model standardizes the inputs to a layer for each mini-batch, provides regularization, and reduces the generalization error. Also, the dropout technique used in the classification unit of our model provides regularization by dropping out a percentage of outputs from the previous layer to prevent the overfitting and improve generalization. These results show the effectiveness of our method for masked facial recognition.

4.3.6. Cross-dataset validation

The key purpose of this experiment is to test the generalizability of our model face mask detection over cross-dataset scenarios. The fundamental goal of cross-dataset validation is to determine our model's generalization potential and demonstrate its applicability for real-world scenarios. We used our MDMFR dataset, "Facemask Detection Dataset 20,000 Images" and "FaceMask Dataset" (Iandola et al., 2016; Huang et al., 2017) for this purpose. The "Facemask Detection Dataset 20,000 images" has 10,000 images in each class i.e., with mask, and without mask, whereas the "Face-Mask Dataset" contains 208 images with masks and 131 images without masks. To perform the cross-dataset validation process, we acquired the following six scenarios: (a) training on the "Facemask Detection Dataset 20,000 Images" and testing on our MDMFR dataset, (b) training on our MDMFR dataset and testing on "Facemask Detection Dataset 20,000 Images" dataset, (c) training on the MDMFR dataset and testing on the "facemask" dataset, (d) training on the "facemask" dataset and testing on our MDMFR dataset, (e) training on the "facemask" dataset and testing on the "Facemask Detection Dataset 20,000 Images" dataset, (f) training on the "Facemask Detection Dataset 20,000 Images" dataset and testing on the "facemask dataset". The results of our cross-dataset validation experiment are reported in Table 7. Our proposed model attained the testing accuracy of 100% for scenarios (a), (d), (e) and (f), whereas, 99.99%, for scenarios (b) and (c). These remarkable results validate the generalizability of our proposed model for face mask detection in real-world diverse scenarios. Because of the unavailability of standard large scale masked facial recognition datasets, we are unable to evaluate our framework on cross dataset for the masked facial recognition task.

5. Conclusion

This paper has presented a Deepmasknet framework for accurate face mask detection and masked facial recognition. Moreover, an in-house unified MDMFR dataset has also been developed to assess the performance of the proposed method. We have created a largescale and diverse face images dataset to evaluate both the face mask detection and masked facial recognition. The accuracy of 100% for face mask detection and 93.33% for masked facial recognition have confirmed the superiority of our Deepmasknet model over the contemporary techniques. Furthermore, experimental results on the three standard Kaggle datasets and our MDMFR dataset have verified the robustness of the proposed Deepmasknet model for face mask detection and masked facial recognition under diverse conditions i.e., variations in face angles, lightning conditions, gender, skin tone, age, types of masks, occlusions (glasses), etc. The proposed method can be utilized for a variety of purposes including the security and safety of the people i.e., recognition of robbers wearing a face mask, etc. In the future, we intend to further expand our dataset in terms of both the quantity and diversity.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

References

Abhila, A.G., Sreeletha, S.H., 2018. A deep learning method for identifying disguised faces using AlexNet and multiclass SVM. Int. Res. J. Eng. Technol. 5 (07).

Al-Ramahi, M., Elnoshokaty, A., El-Gayar, O., Nasralah, T., Wahbeh, A., 2021. Public discourse against masks in the COVID-19 era: infodemiology study of Twitter data. JMIR Public Health Surveillance 7 (4), e26780.

Balotra, A., Maity, M., Naik, V., 2019. An empirical evaluation of the COTS air pollution masks in a highly polluted real environment. In: 2019 11th International Conference on Communication Systems & Networks (COMSNETS). IEEE, pp. 842–848.

Bhattacharya, I., 2021. HybridFaceMaskNet: A Novel Face-Mask Detection Framework Using Hybrid Approach.

Bu, W., Xiao, J., Zhou, C., Yang, M., Peng, C., 2017. A cascade framework for masked face detection. In: 2017 IEEE International Conference on Cybernetics and Intelligent Systems (CIS) and IEEE Conference on Robotics, Automation and Mechatronics (RAM). IEEE, pp. 458–462.

Cao, Q., Shen, L., Xie, W., Parkhi, O.M., Zisserman, A., 2018, May. Vggface2: A dataset for recognising faces across pose and age. In 2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018) (pp. 67-74). IEEE.

Chowdary, G.J., Punn, N.S., Sonbhadra, S.K., Agarwal, S., 2020. Face mask detection using transfer learning of inceptionv3. In: International Conference on Big Data Analytics. Springer, Cham, pp. 81–90.

Damer, N., Boutros, F., Süßmilch, M., Kirchbuchner, F., Kuijper, A., 2021. Extended evaluation of the effect of real and simulated masks on face recognition performance. IET Biometrics 10 (5), 548–561.

Deng, J., Guo, J., An, X., Zhu, Z., Zafeiriou, S., 2021. Masked face recognition challenge: the insightface track report. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 1437–1444.

Ding, F., Peng, P., Huang, Y., Geng, M., Tian, Y., 2020. Masked face recognition with latent part detection. In: Proceedings of the 28th ACM International Conference on Multimedia, pp. 2281–2289.

Du, H., Shi, H., Liu, Y., Zeng, D., Mei, T., 2021. Towards NIR-VIS masked face recognition. IEEE Signal Process Lett. 28, 768–772.

Eikenberry, S.E., Mancuso, M., Iboi, E., Phan, T., Eikenberry, K., Kuang, Y., Kostelich, E., Gumel, A.B., 2020. To mask or not to mask: Modeling the potential for face mask use by the general public to curtail the COVID-19 pandemic. Infect. Dis. Model. 5, 293–308.

Ejaz, M.S., Islam, M.R., Sifatullah, M., Sarker, A., 2019. Implementation of principal component analysis on masked and non-masked face recognition. In: 2019 1st International Conference on Advances in Science, Engineering and Robotics Technology (ICASERT). IEEE, pp. 1–5.

Geng, M., Peng, P., Huang, Y., Tian, Y., 2020. Masked face recognition with generative data augmentation and domain constrained ranking. In: Proceedings of the 28th ACM International Conference on Multimedia, pp. 2246–2254.

Greenhalgh, T., Schmid, M.B., Czypionka, T., Bassler, D., Gruer, L., 2020. Face masks for the public during the covid-19 crisis. BMJ 369.

Guo, Y., Zhang, L., Hu, Y., He, X., Gao, J., 2016. Ms-celeb-1m: A dataset and benchmark for large-scale face recognition. In: European conference on computer vision. Springer, Cham, pp. 87–102.

He, L., Li, H., Zhang, Q.i., Sun, Z., 2019. Dynamic feature matching for partial face recognition. IEEE Trans. Image Process. 28 (2), 791–802.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.

Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q., 2017. Densely connected convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4700–4708.

Iandola, F.N., Han, S., Moskewicz, M.W., Ashraf, K., Dally, W.J. and Keutzer, K., 2016. SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and< 0.5 MB model size. arXiv preprint arXiv:1602.07360.

Inamdar, M., Mehendale, N., 2020. Real-Time Face Mask Identification Using Facemasknet Deep Learning Network. Available at SSRN 3663305.

[dataset] Lavik Jain, Jathin Pranav Singaraju, 2020. Facemask Detection Dataset 20,000 Images. https://www.kaggle.com/luka77/facemask-detection-dataset-20000-images.

Jayaweera, M., Perera, H., Gunawardana, B., Manatunge, J., 2020. Transmission of COVID-19 virus by droplets and aerosols: a critical review on the unresolved dichotomy. Environ. Res. 188, 109819.

Jayaweera, M., Perera, H., Gunawardana, B., Manatunge, J., 2020. Transmission of COVID-19 virus by droplets and aerosols: a critical review on the unresolved dichotomy. Environ. Res. 188, 109819.

Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. Adv. Neural Inf. Process. Systems 25, 1097–1105.

Kumar, R.S., Rajendran, A., Amrutha, V., Raghu, G.T., 2021, March. Deep Learning model for face mask based attendance system in the era of the Covid-19 pandemic. In 2021 7th International Conference on Advanced Computing and Communication Systems (ICACCS) (Vol. 1, pp. 1741-1746). IEEE.

Li, C., Wang, R., Li, J., Fei, L., 2020. Face detection based on YOLOv3. In: Recent Trends in Intelligent Computing, Communication and Devices. Springer, Singapore, pp. 277–284.

Loey, M., Manogaran, G., Taha, M.H.N., Khalifa, N.E.M., 2021. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement 167, 108288.

Loey, M., Manogaran, G., Taha, M.H.N., Khalifa, N.E.M., 2021. A hybrid deep transfer learning model with machine learning methods for face mask detection in the era of the COVID-19 pandemic. Measurement 167, 108288.

Maas, A.L., Hannun, A.Y., Ng, A.Y., 2013, June. Rectifier nonlinearities improve neural network acoustic models. In Proc. icml (Vol. 30, No. 1, p. 3).

McDonald, F., Horwell, C.J., Wecker, R., Dominelli, L., Loh, M., Kamanyire, R., Ugarte, C., 2020. Facemask use for community protection from air pollution disasters: an ethical overview and framework to guide agency decision making. Int. J. Disaster Risk Reduct. 43, 101376.

MDMFR dataset. [Online]. Available: https://drive.google.com/drive/folders/1YgHEQzb4yL2FigYPfLKblabmTRJR6rew.

Meenpal, T., Balakrishnan, A., Verma, A., 2019. Facial mask detection using semantic segmentation. In: 2019 4th International Conference on Computing, Communications and Security (ICCCS). IEEE, pp. 1–5.

Militante, S.V., Dionisio, N.V., 2020. Real-time facemask recognition with alarm system using deep learning. In: 2020 11th IEEE Control and System Graduate Research Colloquium (ICSGRC). IEEE, pp. 106–110.

Min, R., Hadid, A. and Dugelay, J.L., 2011. Improving the recognition of faces occluded by facial accessories. In: 2011 IEEE International Conference on Automatic Face & Gesture Recognition (FG) (pp. 442-447). IEEE.

Mundial, I.Q., Hassan, M.S.U., Tiwana, M.I., Qureshi, W.S., Alanazi, E., 2020. Towards facial recognition problem in COVID-19 pandemic. In: 2020 4rd International Conference on Electrical, Telecommunication and Computer Engineering (ELTICOM). IEEE, pp. 210–214.

Nagrath, P., Jain, R., Madan, A., Arora, R., Kataria, P., Hemanth, J., 2021. SSDMNV2: A real time DNN-based face mask detection system using single shot multibox detector and MobileNetV2. Sustainable Cities Soc. 66, 102692.

Nair, V., Hinton, G.E., 2010. Rectified linear units improve restricted boltzmann machines. Icml 2010, 807–814.

Nieto-Rodríguez, A., Mucientes, M., Brea, V.M., 2015, June. System for medical mask detection in the operating room through facial attributes. In Iberian Conference on Pattern Recognition and Image Analysis (pp. 138-145). Springer, Cham.

Oh, H.J., Lee, K.M., Lee, S.U., 2008. Occlusion invariant face recognition using selective local non-negative matrix factorization basis images. Image Vis. Comput. 26 (11), 1515–1523.

Plastiliq Image Resizer. https://download.cnet.com/Plastiliq-Image-Resizer/3000-2192_4-75416095.html

Proença, H., Neves, J.C., 2018. A reminiscence of "Mastermind": iris/periocular biometrics by "In-Set" CNN Iterative analysis. IEEE Trans. Inf. Forensics Secur. 14 (7), 1702–1712.

Qin, B., Li, D., 2020. Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19. Sensors 20 (18), 5236.

Qin, B., Li, D., 2020. Identifying facemask-wearing condition using image super-resolution with classification network to prevent COVID-19. Sensors 20 (18), 5236.

Redmon, Joseph. Darknet: Open-Source Neural Networks in C. https://pjreddie.com/darknet.

Md. Sabbir Ejaz, Md. Rabiul Islam, 2019. Masked face recognition using convolutional neural network. In: International Conference on Sustainable Technologies for Industry, 2019.

Sandler, M., Howard, A., Zhu, M., Zhmoginov, A., Chen, L.C., 2018. Mobilenetv 2: Inverted residuals and linear bottlenecks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4510–4520.

[dataset] Sushant Shah, 2020. Facemask. https://www.kaggle.com/sushantshah/facemask22?select=without+mask.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556.

[dataset] Smansid., 2020. FaceMask Dataset. https://www.kaggle.com/sumansid/facemask-dataset.

Song, L., Gong, D., Li, Z., Liu, C., Liu, W., 2019. Occlusion robust face recognition based on mask learning with pairwise differential siamese network. In: Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 773–782.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9.

Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z., 2016. Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2818–2826.

Szegedy, C., Ioffe, S., Vanhoucke, V., Alemi, A.A., 2017. February. Inception-v4, inception-resnet and the impact of residual connections on learning. Thirty-first AAAI Conference on Artificial Intelligence.

Taigman, Y., Yang, M., Ranzato, M.A., Wolf, L., 2014. Deepface: Closing the gap to human-level performance in face verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1701–1708.

Trillionpairs. http://trillionpairs.deepglint.com/overview 2018.

Ud Din, N., Javed, K., Bae, S., Yi, J., 2020. A novel GAN-based network for unmasking of masked face. IEEE Access 8, 44276–44287.

Venkateswarlu, I.B., Kakarla, J., Prakash, S., 2020. Face mask detection using mobilenet and global pooling block. In: 2020 IEEE 4th Conference on Information & Communication Technology (CICT). IEEE, pp. 1–5.

Vijitkunsawat, W., Chantngarm, P., 2020. Study of the performance of machine learning algorithms for face mask detection. In: 2020-5th International Conference on Information Technology (InCIT). IEEE, pp. 39–43.

Wang, C., Horby, P.W., Hayden, F.G., Gao, G.F., 2020. A novel coronavirus outbreak of global health concern. Lancet 395 (10223), 470–473.

Wen, Y., Zhang, K., Li, Z., Qiao, Y., 2016. A discriminative feature learning approach for deep face recognition. In: European Conference on Computer Vision. Springer, Cham, pp. 499–515.

The WHO COVID-19, Dashboard, 2021. [Online]. Available: https://covid19.who.int/.

World Health Organization, 2020. Advice on the Use of Masks in the Context of COVID-19: Interim Guidance. World Health Organization.

Yi, D., Lei, Z., Liao, S. and Li, S.Z., 2014. Learning face representation from scratch. arXiv preprint arXiv:1411.7923.

Zhang, J.P., Li, Z.W., Yang, J., 2005. A parallel SVM training algorithm on large-scale classification problems. In: 2005 International Conference on Machine Learning and Cybernetics. IEEE, pp. 1637–1641.

Zhang, X., Zhou, X., Lin, M., Sun, J., 2018. Shufflenet: An extremely efficient convolutional neural network for mobile devices. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6848–6856.

Zhao, W., 2017, August. Research on the deep learning of the small sample data based on transfer learning. In AIP Conference Proceedings (Vol. 1864, No. 1, p. 020018). AIP Publishing LLC.