



# VGG16-random fourier hybrid model for masked face recognition

O. K. Sikha<sup>1</sup> · Bandla Bharath<sup>1</sup>

Accepted: 7 May 2022 / Published online: 10 July 2022

© The Author(s), under exclusive licence to Springer-Verlag GmbH Germany, part of Springer Nature 2022

## Abstract

With the recent COVID-19 pandemic, wearing masks has become a necessity in our daily lives. People are encouraged to wear masks to protect themselves from the outside world and thus from infection with COVID-19. The presence of masks raised serious concerns about the accuracy of existing facial recognition systems since most of the facial features are obscured by the mask. To address these challenges, a new method for masked face recognition is proposed that combines a cropping-based approach (upper half of the face) with an improved VGG-16 architecture. The finest features from the unoccluded facial region are extracted using a transfer learned VGG-16 model (Forehead and eyes). The optimal cropping ratio is investigated to give an enhanced feature representation for recognition. To avoid the overhead of bias, the obtained feature vector is mapped into a lower-dimensional feature representation using a Random Fourier Feature extraction module. Comprehensive experiments on the Georgia Tech Face Dataset, Head Pose Image Dataset, and Face Dataset by Robotics Lab show that the proposed approach outperforms other state-of-the-art approaches for masked face recognition.

**Keywords** Masked face recognition · Occlusion face recognition · Cropping based face recognition

## 1 Introduction

COVID-19 is currently wreaking havoc on the world. The uncontrolled coronavirus disease (COVID-19) is a viral infection caused by severe acute respiratory syndrome (SARS-CoV-2). With its deadly spread to 222 countries and territories worldwide, the COVID-19 has caused a global crisis. A total of 352, 234, 810 confirmed cases of COVID-19 that originated in Wuhan, China, have been reported, with a death toll of 5,615,082 as of January 2022. People become infected by coughing, sneezing, and/or talking in close proximity to an infected person's respiratory droplets. Furthermore, the virus can be spread by touching a virus-infected surface or object and then touching the mouth, nose, or eyes. Wearing a mask and social distancing are the most effective ways to combat this pandemic as shown in Fig. 1. Implementing these safety measures, particularly masking the lower half of the face,

has a significant impact on the current security systems based on facial recognition that have already been implemented by a number of corporations and government agencies.

In comparison to traditional methods such as PIN, secret password, finger impression (Multimodal fingerprint spoof detection using white light 2016; Schroff et al. 2015), etc., face recognition has gained a lot of attention as a remarkable biometric authentication procedure across all automatic personal authentication systems (Singh et al. 2021) (Hong et al. 2021; Nefian and Hayes 2000). Many organizations and government bodies rely on this technique to protect their assets and to secure public places such as airports, bus stands, and railway stations etc. With the rapid advancements and expansion of machine learning and deep learning techniques, challenges in the face recognition domain have also been well addressed. Machine learning-based face detection and verification models necessitate manual feature extraction and learning, whereas deep learning models do not. Deep learning architectures, specifically Convolutional Neural Networks (CNN), can learn valuable features from training images automatically. CNNs perform a series of Convolution operations and nonlinear activations on the image, therefore, making them more suitable for working with image data than generic

---

✉ O. K. Sikha  
ok\_sikha@cb.amrita.edu

<sup>1</sup> Department of Computer Science & Engineering Amrita  
School of Engineering, Coimbatore Amrita Vishwa  
Vidyapeetham, Coimbatore, India



**Fig. 1** Chance of Covid 19 transmission with/without social distancing and face mask

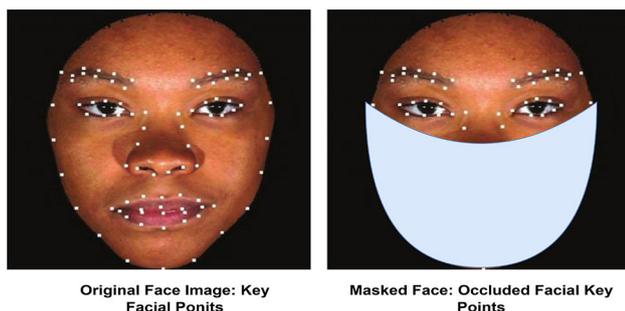
Artificial Neural Network (ANN) architectures. Several studies on the application of CNN-based models for face recognition in various poses and disguises have been published in the literature, with promising accuracies on various open benchmark datasets. The traditional Eigen-face algorithm's recognition accuracy on the LFW dataset is approximately 60% (Lfw recognition results). Whereas the most advanced deep learning face recognition models such as Facenet (Priya and Banu 2014) reported a recognition accuracy of 99% and beyond. Majority of modern machine learning/deep learning face recognition models with high recognition accuracy require labeled un-occluded face datasets (Deng et al. 2019; Li et al. 2021) for recognition. Models trained on normal un-occluded face images learn key facial points for recognition such as face edges, lips, and eyes, but they may fail with masked images since the mask covers most of the key facial points as illustrated in Fig. 2. According to a study conducted by NIST (Nist finds facial recognition has trouble with face masks), the failure rate of state-of-the-art face recognition models ranged between 20% and 50% for masked input images.

Recently certain studies have been carried out to solve the problem of **Masked Face Recognition (MFR)** which can be categorized primarily into three classes:

Occlusion removal-based models, deep learning-based models, and face reconstruction/face completion-based models. This paper proposes an occlusion removal based deep learning architecture for masked face recognition which involves two tasks: (1) face mask detection and (2) masked face recognition. Haar cascade model is used to detect the mask and the masked region is removed for further processing. A hybrid VGG16- Random Fourier deep learning architecture is then used to extract enhanced facial features for recognition. The major contributions of the presented work are as follows:

- Unavailability of masked face images to be trained with is the primary challenge for a masked face recognition model. The proposed model uses existing databases for training, therefore, negating the need for the creation of a masked face database.
- Colour of mask causes uncertainty in the accuracy of a model trained on masked images. The proposed model considers only the part of the face which is visible and discards the part of the face which is covered by the mask hence invariant to mask color. Optimal cropping for MFR is also explored.
- A hybrid VGG16-Random Fourier deep learning model is proposed to extract enhanced features from the upper half of the face for recognition.

The paper is arranged as follows: Related works are detailed in Sect. 2, and 3 describes the proposed model in detail. The experimental setup is detailed in Sect. 4 and 5 describes the results. Finally, the paper concludes with Sect. 6.



**Fig. 2** Facial Key points on original (Singh and Mary 2016) and masked face images

## 2 Related works

Occlusion is a major challenge faced by most of the existing Face Recognition (FR) systems which are generally caused due to the presence of objects or ornaments worn by a person which covers a part of the face such as glasses, hats, masks, bandages, etc. Masks differ from glasses and hats by causing a huge loss in the visual region of the face (Kumar et al. 2019; Simonyan and Zisserman 1409), hence most of the state-of-the-art face recognition models fails when masked faces are given as input. There are three major problems with masked face recognition compared to other occlusion face recognition. To begin with, there is a scarcity of large face datasets with masks. Second, the mouth and nasal features have been significantly harmed, and the effective characteristics have been substantially diminished (refer to Fig. 2). Finally, a mask-wearing face is difficult to detect. During this new normal where the mask has become an inevitable aspect, researchers started to explore techniques to perform face recognition with a mask. Masked face recognition models reported in the literature can be classified into three major categories: occlusion removal approaches, restoration approaches, and deep learning-based approaches.

### 2.1 Occlusion removal

Occlusion removal-based approaches initially detect the occluded face part and discard them completely during recognition. Priya and Banu (Neha and Nithin 2018) divided the face into smaller regions and used SVM classifier to remove the occlusion. The global masked projection method is used by Alyuz et al (Deng et al. 2019) to remove the occluded region.

Partial Gappy PCA is then applied to reconstruct the face using Eigenvectors. To detect occluded regions and eliminate them during the recognition phase, Andres et al. (Andrés et al. 2014) calculated the difference between occluded and non-occluded images of the same person. Most of the occlusion removal-based MFR models reported in the literature spend more time on the detection and removal of occlusion. The proposed model reduces the overhead of mask detection using optimal cropping of face images.

### 2.2 Image restoration

Restoration-based models reconstruct the entire face for recognition (Dolhansky and Ferrer 2018; Wu 2021). A 3D face restoration model was proposed by Bagchi et al. (Din et al. 2020). The authors first detected the occluded face part by thresholding the depth map, Principal Component

Analysis (PCA) is then used to reconstruct the face. Cen et al. (Cen and Wang 2019) introduced a robust occlusion FR classification system based on depth dictionary representation, which used a convolutional neural network as a feature extractor and then linearly encoded the extracted depth data using the dictionary. For occlusion FR, Du et al. (Du and Hu 2019) presented Nuclear Norm-based Adapted Occlusion Dictionary Learning (NNAODL), which used a dictionary of occluded pictures to build a unique reconstruction model. The image reconstruction algorithms have improved the occlusion FR process, yet the major drawback is the computational complexity as they try to reconstruct the original face image.

### 2.3 Deep learning

Deep learning has seen a lot of success in FR in recent years. Deep features have outperformed classical features and are widely used in occlusion FR. The Dynamic Feature Matching (DFM) technique, which combines FCN with SRC to recognize partial faces of any size, was proposed in He et al. (2016). This work explored the effectiveness of the deep features after the last pooling layer to represent the database images. An end-to-end BoostGAN model was proposed by Duan et al. (Duan and Zhang 2020) in which non-occluded face images were synthesized from the input occluded image for refined face recognition. An open-source tool called MaskTheFace was introduced by Anwar et al. (Anwar and Raychowdhury 2008), which initially generates a masked face image dataset from existing unmasked face images. The generated masked faces were then used for training a deep face recognition model. Walid et al. (Gourier et al. 2004) used a pre-trained CNN model to extract features from the upper half of the face (after mask removal). The quantized bag of feature representation was further extended for facial recognition. Table 1 summarizes various state-of-the-art Masked Face Recognition models.

Inspired by the high performance and accuracy of deep CNN-based models robust to facial expression, facial occlusion, and illumination, in this paper we propose a discard occlusion-based hybrid deep mask face recognition model.

## 3 Proposed system

The proposed model comprises four key modules: (1) Face alignment and pre-processing (2) Mask removal (3) Feature Extraction (4) Face recognition. Fig. 3 demonstrates the architecture of the proposed model.

**Table 1** Summary of various state-of-the-art masked face recognition models

Paper	Model Used	Dataset	Methodology	Other requirements	Accuracy(%)
Din et al. (2020)	GANs	CelebA	Map and editing modules	VGG-19	
Alzu'bi et al. (2016)	Pre-trained CNNs	RMFRD	Comparative study on CNNs	VGGFace, facenet, openface, deepface	68.17
Kumar et al. (2021)	MTArcFace	LFW, CFP, Agedb	Combination of arcface loss and mask-usage classification loss	ArcFace	99.78
Lucena et al. (2017)	VGG-16 and facenet	Custom dataset	Learning cosine distance		100
Hariri (2015)	VGG-16, AlexNet, ResNet-50	RMFRD, SMFRD	Deep features of facial areas	VGG-16, alexnet, resnet-50	91.3
Geng et al. (2014)	Facemasknet-21	Custom dataset	Deep metric learning	Facemasknet	88.92
Wang et al. (2003)	Resnet	MFDD, RMFRD	Attention-driven Model		95
Szegedy et al. (2016)	Attention-based	MFDD, RMFRD	Face-eye-based multi-granularity		99
Anwar and Raychowdhury (2008)	MasktheFace	VGGFace2-mini-SM, LFW-SM	MaskTheFace with FaceNet	Facenet	97.25
He et al. (2018)	Wearmask3D	Normalized softmax loss	MFR2, MFW-mini	Resnet-50	95.8
3–37 (2019).	GANs	MFSR, CASIA-WebFace, VGGFace2	IAMGAN with DCR		86.5
Lane (2020)	GANs	Celeb-A, LFW, AR	De-occlusion distillation		95.44
Li et al. (2020)	CBAM	Webface, AR, Yela B, LFW	Face cropping	CBAM	92.61
Deng et al. (2021)	MFCosface	VGGFace2 m, LFW m, CASIAFaceV5 m, MFR2, RMFD	Learning large margin cosine loss	Facenet	98.5
Du et al. (2021)	Siamese networks	Oulu-CASIA NIR-VIS, BUAA-VisNir	Heterogeneous semi-Siamese training	ResNet-50	98.6

### 3.1 Face alignment and preprocessing

Face detection (Karthika and Parameswaran 2016; Kumar et al. 2019) and alignment correction is a vital step in face recognition. Eye-Haar cascade configurations available in the OpenCV library is used to detect eye points, using which the frontal face obtained from the camera is oriented such that the face is perpendicular to the normal of the image. The face alignment algorithm is described in Algorithm. 1 and the corresponding steps are shown in Fig. 4.

#### Algorithm 1 Face alignment and preprocessing

- 1: Detect eyes using an Eye Haar-Cascade Classifier (Figure.4.b).
- 2: Calculate the center of the detected bounding boxes and connect the points (ab) (Figure.4.c).
- 3: Construct a triangle by connecting line ab to the normal of the original image (ac) (Figure.4.d).
- 4: The angle in which the eyes were oriented to the normal of the image was calculated (Compute the angle between ac and bc).
- 5: Orient the image by rotating the image by the calculated angle (Figure.4.e).

### 3.2 Cropping masked region

Masks are usually worn below the eyes and cover a part of the nose region. Therefore, to remove the mask the face image below the eye region is cropped. Although Haar cascades are capable of localizing the eyes correctly the windows which are drawn on each eye are not equal and differ in position. Therefore, the bounding box which is the lower of the two predicted eyes are selected and the region below the bounding box is cropped to remove the mask from the image. The lower of the two bounding boxes are identified by calculating the distance between the bottom line of the bounding box and the bottom center of the image as in Eq. 1.

$$\text{Distance} = \left\| \frac{(P_2 - P_1) * (P_1 - P_3)}{(P_2 - P_1)} \right\| \quad (1)$$

where  $P_1$  and  $P_2$  are the edge points of the bottom line of a bounding box and  $P_3$  is the bottom center point of the image. Fig. 5 shows the mask removal from the input image. One of the major concerns of the cropping-based approach for MFR is: “where to crop the facial image?”

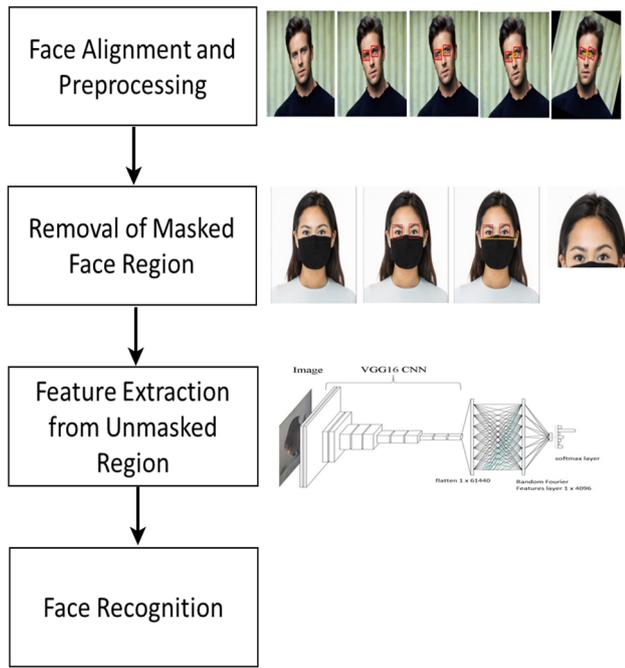


Fig. 3 Architecture of the proposed model

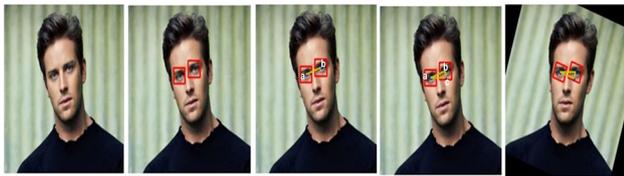


Fig. 4 Face alignment and preprocessing



Fig. 5 Mask removal from the input image

Sample facial images at different cropping proportions are shown in Fig. 6. Different cropping proportions are calculated using three key facial points as shown in Fig. 7. Euclidean distance between  $A(x_1, y_1)$ ,  $B(x_2, y_2)$  (eye key points) in Fig. 7 is calculated as follows:

$$L = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2} \tag{2}$$

The perpendicular bisector of the line AB is calculated (Point C in Fig. 7). Finally, face images are cropped as C as the reference point.

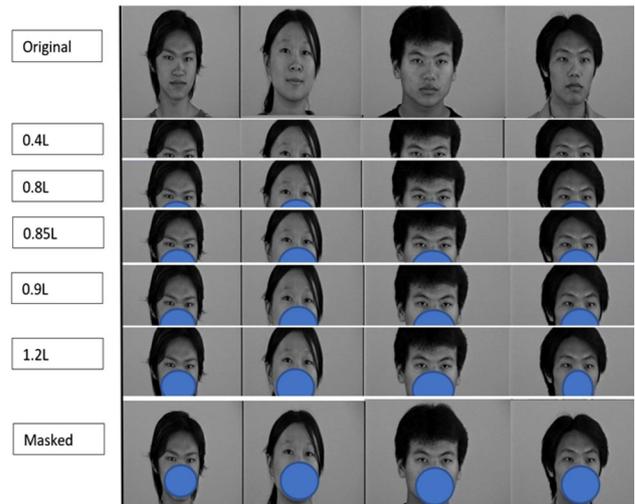


Fig. 6 Cropped facial images at different cropping proportions

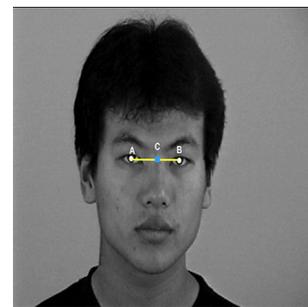


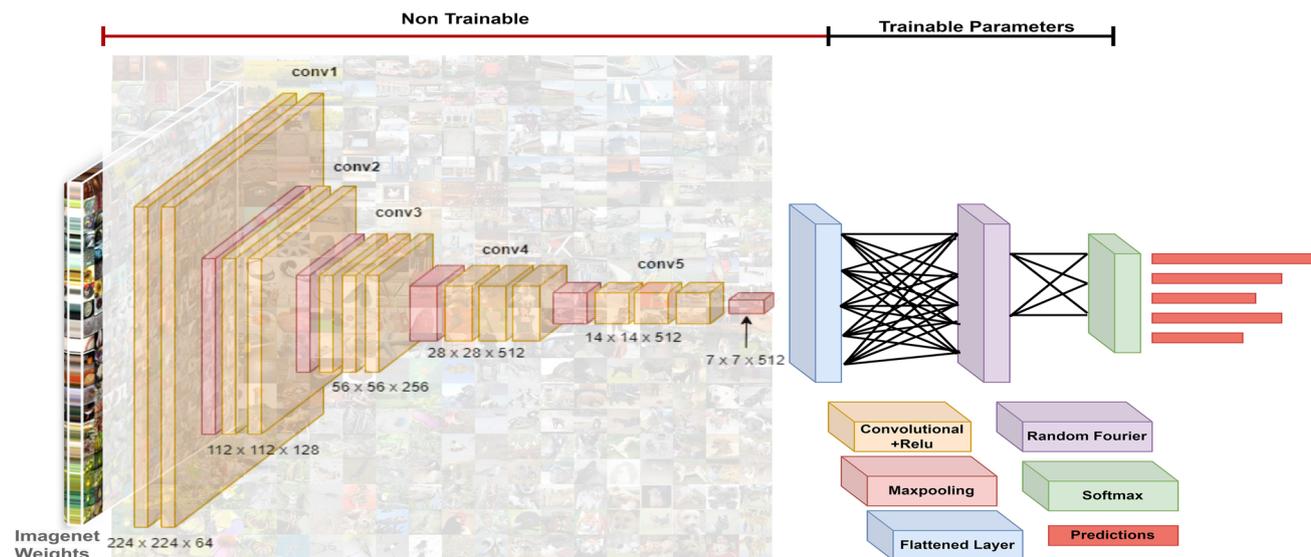
Fig. 7 Calculating cropping proportion (L)

### 3.3 Feature extraction using improved VGG-16 model

VGG16 is a renowned classification model proposed by K. Simonyan and A. Zisserman (Shaheed et al. 2022) which achieved 92.7% accuracy on ImageNet classification with more than 14 million images and 1000 classes. The input convolutional layer accepts RGB image of size 224X224. The image is further passed through a stack of 3X3 convolutional layers. The initial layers of VGG-16 capture low-level features and the deep layers capture high-level features. The original VGG-16 architecture is modified by adding an additional Random Fourier layer after the final fully convolutional layer as shown in Fig. 8. A Random Fourier feature layer is added to extract enhanced features from the partial facial image. The selection of VGG-16 as the backbone network is done based on the comparative analysis of various deep convolutional models.

### 3.4 Random fourier features layer

In general, the lower half of the face contributes more to facial recognition since the features extracted from the



**Fig. 8** Proposed VGG16-Random Fourier hybrid deep learning model for masked face recognition. Feature representation and classification using transfer learned VGG16 model

bottom of the eyes to the chin will be unique for a person. An enhanced feature representation is crucial in masked face recognition where the lower part of the face is completely occluded by the mask. The Random Fourier feature is a well-known, simple, and effective method for scaling-up kernel functions. The underlying principle of the method is a consequence of Bochner’s theorem (Bochner 1932), which states that “any bounded, continuous and shift-invariant kernel is a Fourier transformation to the input features and then training a linear model on top of the transformed features. Depending on the loss function used in the linear model, the incorporation of a random Fourier layer is analogous to kernel SVMs (for hinge loss) and kernel logistic regression (for logistic loss).

The first set of random features consists of random Fourier bases  $\cos(\omega x + b)$  where  $\omega \in \mathbb{R}$  and  $b \in \mathbb{R}$  are random variables. These mappings project data points on a randomly chosen line and then pass the resulting scalar through a sinusoidal function as shown in Fig. 9. Each component of the feature map  $z(x)$  projects  $x$  onto a random direction  $\omega$  drawn from the Fourier transform  $p(\omega)$  of  $k(\Delta)$ , and wraps this line onto the unit circle in  $\mathbb{R}^2$ . After the transformation of two points  $x$  and  $y$  as above, their inner product can be considered as an unbiased estimator of  $k(x, y)$  (Fig. 10). The mapping  $z(x) = \cos(\omega x + b)$  additionally rotates this circle by a random amount  $b$  and projects the points onto the interval  $[0, 1]$  (Fig. 11). In the proposed model a Random Fourier Feature layer of size  $4096 \times 1$  with scale 10 is appended with the final fully connected layer of VGG-16, which is then connected to a multi-layer perceptron for classification with SoftMax layer (Fig. 12). The hybrid VGG-16-Random Fourier model is expected to extract enhanced features from the upper half of the face

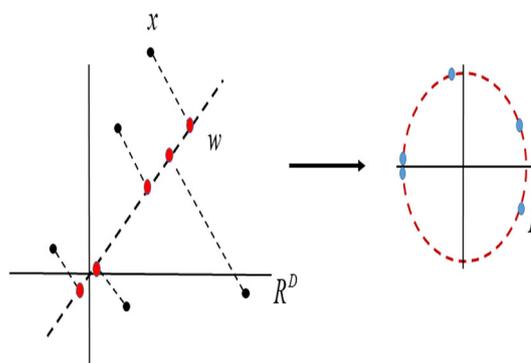
image, which can be further used for facial recognition/authentication.

### 4 Experimental setup

This section introduces the datasets used in the experiment, the preprocessing and exploratory data analysis performed on the dataset, and concludes with feature representation and classification. Three publicly available face datasets are used for evaluating the proposed MFR model: Face detection dataset by Robotics Lab.

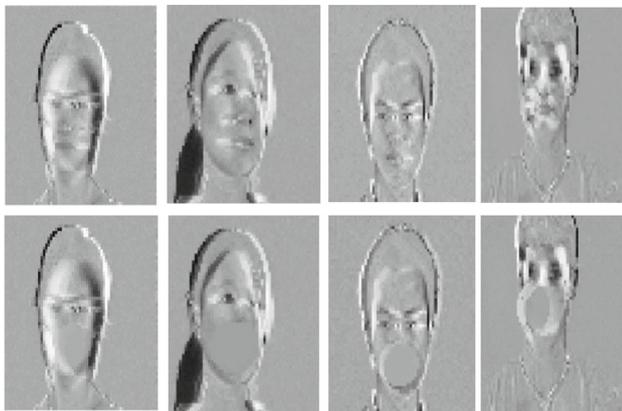
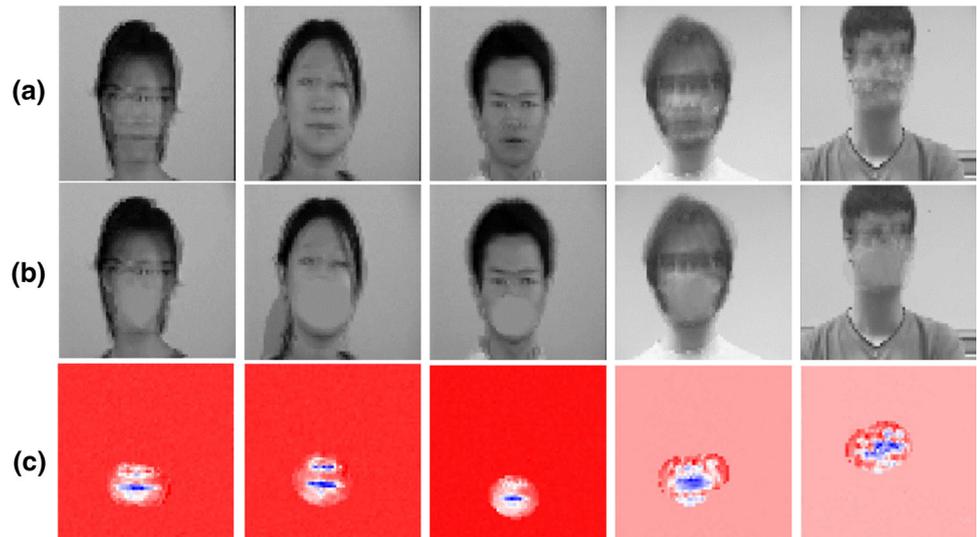
#### 4.1 Datasets used for the experiment and exploratory data analysis

This section describes the datasets used for evaluating the proposed model and exploratory data analysis carried out on the datasets.

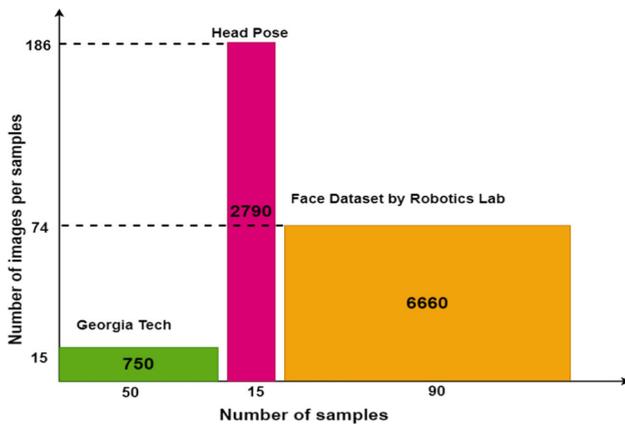


**Fig. 9** Visualization of data projection on to random Fourier bases

**Fig. 10** (a) Average images of normal face image (b) Average images of masked faces (c) Difference between average normal and average masked



**Fig. 11** Eigen images of normal and masked face images



**Fig. 12** Data distribution of three datasets x-axis shows the number of samples and Y axis shows the number of images per sample

### 4.1.1 Face dataset by robotics Lab

This dataset (Chu et al. 2007) includes 6660 images of 90 different subjects. Each subject comprises 74 images, with 37 images captured every 5 degrees in the pan rotation from the right profile (defined as +90°) to the left profile (defined as -90°). The additional 37 images are created (synthesized) by flipping the original 37 photos horizontally using commercial image processing software.

### 4.1.2 Head pose image dataset

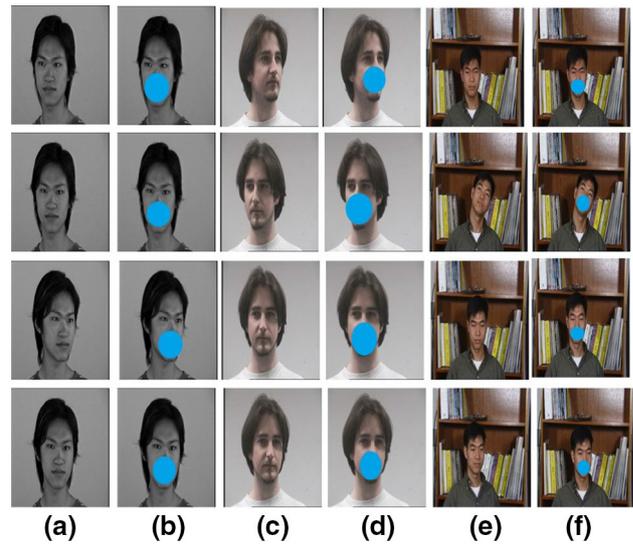
The head pose database is a collection of 2790 monocular face images of 15 people with pan and tilt angles ranging from -90 to +90 degrees (Golwalkar and Mehendale 2022). There are two series 93 images (93 different poses) available for each person. The goal of having two series for each person is to be able to train and test algorithms on both known and unknown faces. In order to focus on face operations, the background is purposefully neutral and uncluttered.

### 4.1.3 Georgia tech face dataset

The Georgia Tech face dataset (Maharani et al. 2020) is a well-known face detection and recognition dataset reported in the literature. It consists of images of 50 people taken in two or three sessions at Georgia Institute of Technology’s Center for Signal and Image Processing between 06/01/99 and 11/15/99. All of the people in the dataset are represented by 15 color JPEG images with cluttered backgrounds taken at 640x480 pixel resolution. The average face size in these images is 150x150 pixels. The images depict frontal and/or tilted faces with various facial expressions, lighting conditions, and scale.

The selection of the aforementioned datasets is done owing to the fact that the Random Fourier layer maps the generated feature vector onto a unit circle. Hence, to generate similar features for individual classes it is better to have images that span from right profile to left profile. Table 2 summarizes the details of the three datasets used in detail. Fig. 13. shows the original and synthesized masked images from the three publically available face datasets. Synthesized masked images are used to train the model since the proposed hybrid deep learning model for masked face recognition uses only the upper half of the face for face recognition hence the presence/absence of mask need not be considered.

Fig. 10 shows the average images generated for normal face images and masked face images of the three datasets described earlier. As shown in Fig. 10b, the average-mask images have more obstructions around the masked areas. In Fig. 10c, which depicts the difference between average-normal and average-mask, it is clear that the upper half of the face averages are consistent across subjects and datasets while the masked regions exhibit high variability. The image locations with the highest variability are highlighted in blue in the difference image (masked region), prompting the proposed model to investigate the upper half of the face image for feature extraction. Fig. 11 shows the Eigen images for four random human samples from the dataset, which are essentially the eigenvectors (components) of PCA of the facial images. For each class, we visualize the principal components that account for 70% of the variability. For masked samples, it is clear that the eigenfaces capture facial key points in the upper half of the face images. Based on these findings, it is possible to conclude that the upper half of the image encompasses features that can be further investigated for masked face recognition. The data distribution (class-wise) of the three datasets under consideration is shown in Fig. 12. All the three datasets are balanced across samples: Georgia Tech dataset contains 50 human samples with 15 images per sample, the head pose dataset contains 15 human samples with 186



**Fig. 13** Sample images used in the experiment. (a, b)Original & Synthesized masked face images from Face Dataset by Robotics Lab. (c, d) Original & Synthesized masked face images from Head Pose Dataset. (e, f)Original & Synthesized masked face images from Georgia Tech Face Dataset

images per sample, and the robotics lab dataset contains 90 human samples with 74 images per sample.

## 4.2 Preprocessing

The input images are resized to 500X500. OpenCV eye Haar cascade configurations are then used to detect the eye locations, which is then used for alignment correction. The masked part of the facial image is then removed based on the eye location.

## 4.3 Feature representation and classification

The proposed model is transfer learned on VGG-16 architecture using ImageNet weights. Stevo Bozinovski and Ante Fulgosi (Bozinovski 1976; Bozinovski and Fulgosi 1976) introduced transfer learning in 1976, and it is widely used to improve the accuracy of Neural Networks

**Table 2** Details of different datasets used in the experiment

Dataset	Total number of persons	Pose, illumination and facial expression variations	Total number of facial images
Face dataset by robotics lab	90	5 degree from right profile(defined as + 90°) to left profile(defined as - 90°) in the pan rotation	6660
Head pose image dataset	15	Variations of pan and tilt angles from -90 to + 90 degrees	2790
Georgia tech face dataset	50	Frontal and/or tilted faces with different facial expressions	750

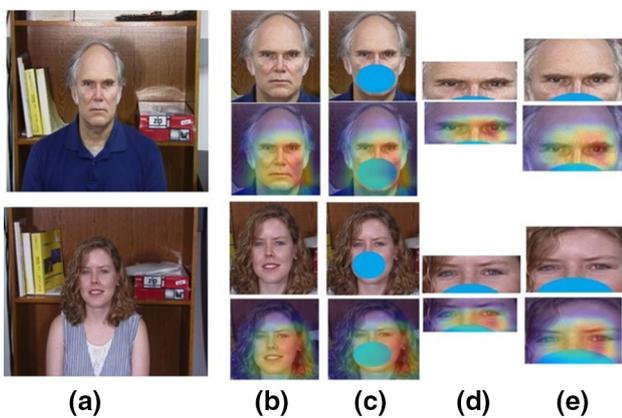
**Table 3** Comparison of three deep CNN models on MFR with/without mask on three benchmark datasets

	Face dataset by robotics lab		Head pose image dataset		Georgia tech face dataset	
	With mask (Acc %)	Without mask (Acc %)	With mask (Acc %)	Without mask (Acc %)	With mask (Acc %)	Without mask (Acc %)
VGG-16	80.623	<b>85.002</b>	82.930	<b>86.141</b>	81.032	<b>86.079</b>
Inception V3	83.607	<b>87.113</b>	84.807	<b>89.32</b>	84.205	<b>88.541</b>
ResNet50	84.250	<b>86.066</b>	85.619	<b>88.995</b>	84.876	<b>87.380</b>

**Table 4** The MFR performance at different cropping proportions without Random Fourier module on Face dataset by Robotics Lab

Cropping proportion	VGG-16 (Acc%)	Inception V3 (Acc%)	ResNet50 (Acc%)
0.4	78.338	76.910	80.408
0.5	78.342	77.285	80.291
0.55	78.94	78.991	81.730
0.6	79.480	79.480	82.776
0.65	81.002	82.632	82.924
0.7	83.619	83.746	83.252
0.75	84.281	85.955	83.925
<b>0.8</b>	84.893	<b>87.113</b>	84.42
<b>0.85</b>	<b>85.002</b>	87.008	85.268
0.9	84.622	86.491	85.702
<b>0.95</b>	82.781	85.087	<b>86.066</b>
1	81.058	84.550	85.82
1.2	80.926	83.9	85.117
No-Crop	80.623	83.607	84.250

Maximum accuracy obtained and the corresponding cropping proportion for each model is shown in bold



**Fig. 14** Visualization of feature maps of two different persons. (a) Original face image (b) Feature map obtained from original face image (c) feature map from synthesized masked images (d) feature map from cropped eye regions (e) feature map from the optimal cropping proportion: locates non- occluded face region precisely and generates unique features for MFR

(Bansal et al. 2021; Liu et al. 2019) over the years. Transfer learning is the process of initializing the weights of a neural

network trained for a task with the weights of previously learned neural networks on a large-scale dataset. The idea is to reuse the intelligence learned by the large-scale neural network to extract high-level features common to both tasks, reducing the need for a massive database to achieve near state-of-the-art predictions. This paper uses a VGG-16 model transfer learned on Imagenet weights for the experiment as depicted in Fig. 8. ImageNet (Karthika and Parameswaran 2016) is a large-scale image database with over 14 million images labeled into approximately 1000 classes. The ImageNet Large-Scale Visual Recognition Challenge was launched in 2010 as a global competition to develop state-of-the-art (SOA) classification, models. Using various SOA neural networks, significant progress has been made in classifying ImageNet over the years. ImageNet weights are widely used for computer vision applications because they are highly reusable for different computer vision tasks and are easily accessible on the web. The layers of VGG-16 model were kept non- trainable during the experiment. The flattened FC layer of size 1X61440 was then mapped into Random Fourier features layer for computing the enhanced lower-dimensional

**Table 5** Comparison of state-of-the-art deep CNN as base models for masked face recognition (With Pose Correction) + (Without Mask)

	Face dataset by robotics lab		Head pose image dataset		Georgia tech face dataset	
	Without random fourier module (Acc %)	With random fourier module (Acc %)	Without random fourier module (Acc %)	With random fourier module (Acc %)	Without random fourier module (Acc %)	With random fourier module (Acc %)
VGG-16	85.002	<b>97.460</b>	86.141	<b>97.634</b>	86.079	<b>97.552</b>
Inception V3	87.113	<b>93.836</b>	89.32	<b>95.209</b>	88.541	<b>94.936</b>
ResNet50	86.066	<b>95.992</b>	88.995	<b>96.47</b>	87.380	<b>96.002</b>

Maximum accuracy obtained for each dataset is shown in bold

feature vector of size 1X4096. The feature vector is then passed on to a softmax layer for classification.

## 5 Results and analysis

This section describes experiments for validating the proposed masked face recognition model. The experiments carried out in this paper are two-fold: (1) Cropping-based approach for MFR using state-of-the-art deep convolutional models (2) Cropping-based approach integrated with Random Fourier layer for MFR.

### 5.1 Cropping-based approach for MFR

The first part of this section discusses the selection of a deep convolutional neural network as the base model for MFR. Secondly, the effect of cropped face images for MFR is experimented with and evaluated. To select the best deep CNN model for masked face recognition, the performance

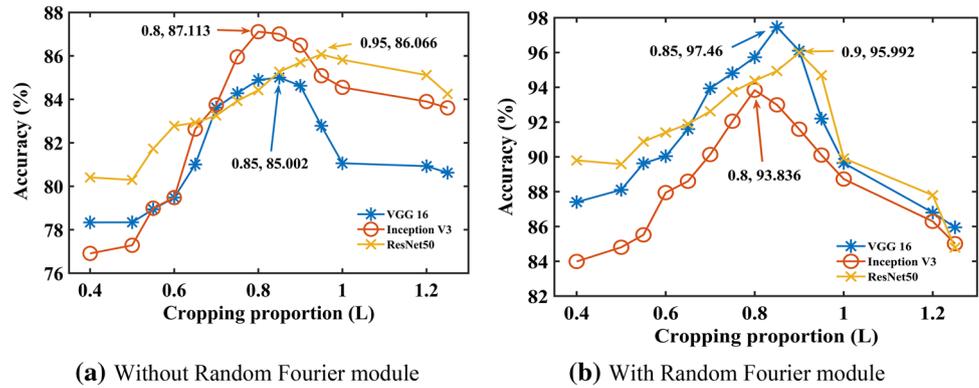
of state-of-the-art deep learning architectures are evaluated in terms of recognition accuracy. VGG- 16(Shaheed et al. 2022), Inception V3, (Soyel and Demirel 2010) and ResNet50 (Hariri 2022) are chosen for the experiment as the base models. Each of the aforementioned models are transfer learned for masked face recognition. Table 3 summarizes the obtained recognition accuracy for two cases: with mask and without mask on three benchmark datasets (The necessary findings are indicated by bold entries). For case 1 the entire face image with mask is considered for the recognition and for case 2 upper half of the face after cropping out the masked region is considered. From the table, it is clear that all the three deep convolutional models produce better accuracy on the cropped face images without masks across the three datasets under study. Inception V3 gave the best accuracy on all the three datasets: On the face dataset by the robotics lab the accuracy is 83.607% for case1 87.113% for case2, with a difference of +3.5% for case 2, on Head pose image dataset 84.807% for case1 and 89.32% for case2, with a difference

**Table 6** The MFR performance at different cropping proportions with Random Fourier on Face dataset by Robotics Lab

Cropping proportion	VGG-16 (Acc%)	Inception V3 (Acc%)	ResNet50 (Acc%)
0.4	87.405	83.991	89.80
0.5	88.11	84.806	89.582
0.55	89.626	85.52	90.881
0.6	90.04	87.947	91.403
0.65	91.597	88.60	91.892
0.7	93.942	90.136	92.620
0.75	94.813	92.052	93.73
<b>0.8</b>	95.729	<b>93.836</b>	94.381
<b>0.85</b>	<b>97.460</b>	92.993	94.94
0.9	96.103	91.591	94.694
<b>0.95</b>	92.2	90.104	<b>95.992</b>
1	89.649	88.728	89.922
1.2	86.80	86.297	87.8
No Crop	85.951	85.014	84.79

Maximum accuracy obtained and the corresponding cropping proportion for each model is shown in bold

**Fig. 15** Cropping proportion comparison in terms of accuracy on Face dataset by Robotics Lab. **(a)** Without Random Fourier module **(b)** With Random Fourier module



**Table 7** Performance comparison of VGG16 without Random Fourier Layer on Masked Face Recognition

	Including masked face region		Excluding masked face region	
	No pose correction	With pose correction	No pose correction	With pose correction
Face dataset by robotics lab				
Accuracy	79.407	80.623	82.519	<b>85.002</b>
True positive	70.529	71.27	74.779	<b>78.635</b>
False positive	24.63	20.129	15.309	<b>9.01</b>
Head pose image dataset				
Accuracy	79.894	82.930	83.612	<b>86.141</b>
True positive	71.4006	72.594	73.087	<b>79.820</b>
False positive	23.918	19.006	14.215	<b>9.929</b>
Georgia tech face dataset				
Accuracy	79.627	81.032	82.981	<b>86.079</b>
True positive	71.295	72.184	74.266	<b>79.015</b>
False positive	24.390	19.860	14.904	<b>9.427</b>

The best results are shown in bold

**Table 8** Performance comparison of VGG16-random fourier hybrid model on masked face recognition

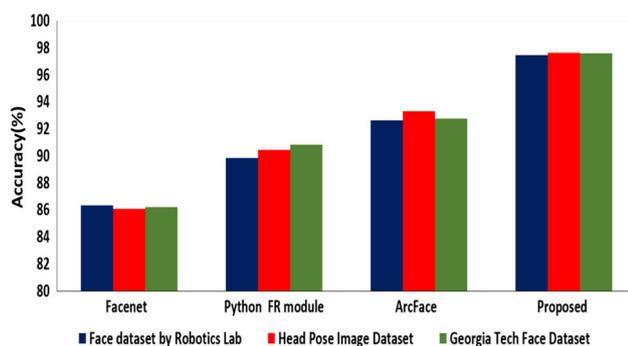
	Including masked face region		Excluding masked face region	
	No pose correction	With pose correction	No pose correction	With pose correction
Face dataset by Robotics Lab				
Accuracy (%)	85.722	85.951	96.381	<b>97.460</b>
True positive (%)	78.860	79.667	91.163	<b>93.001</b>
False positive (%)	14.374	15.396	5.468	<b>2.293</b>
Head Pose Image Dataset				
Accuracy (%)	85.521	87.008	95.986	<b>97.634</b>
True positive (%)	79.8	81.749	91.371	<b>94.601</b>
False positive (%)	13.029	13.962	3.358	<b>1.720</b>
Georgia Tech Face Dataset				
Accuracy (%)	84.92	86.705	95.131	<b>97.552</b>
True positive (%)	79.016	82.140	91.54	<b>93.872</b>
False positive (%)	13.812	14.288	2.830	<b>2.065</b>

The best results are shown in bold

**Table 9** Performance Comparison of state-of-the-art models with the proposed VGG 16-Random Fourier hybrid model on three benchmark datasets

		Accuracy (%)	True positive (%)	False positive (%)
Face dataset by robotics lab	Facenet	86.329	74.192	13.6051
	Python face-recognition 1.3.0 module	89.827	80.251	8.306
	ArcFace	92.610	78.042	12.886
	Proposed	<b>97.460</b>	<b>93.001</b>	<b>2.293</b>
Head pose image dataset	Facenet	86.056	76.721	12.572
	Python face-recognition 1.3.0 module	90.418	79.994	8.76
	ArcFace	93.290	80.672	11.415
	Proposed	<b>97.634</b>	<b>94.601</b>	<b>1.720</b>
Georgia tech face dataset	Facenet	86.193	75.405	12.942
	Python face-recognition 1.3.0 module	90.824	79.218	9.570
	ArcFace	92.73	79.045	11.386
	Proposed	<b>97.552</b>	<b>93.872</b>	<b>2.065</b>

The best results are shown in bold

**Fig. 16** Performance comparison of MFR with different approaches on three benchmark datasets

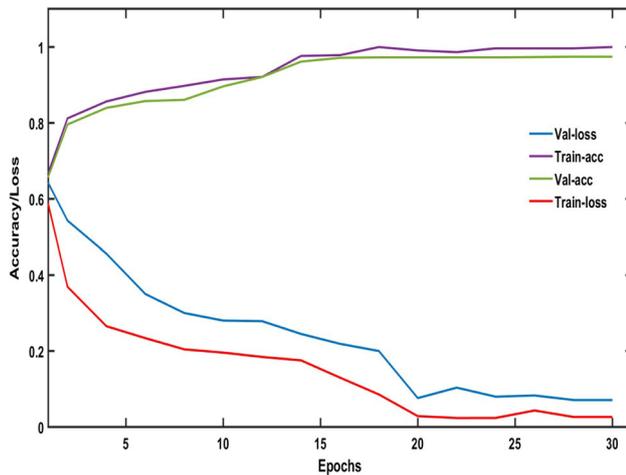
of +4.5% for case2, on Georgia tech face dataset 84.205% for case1 88.541% for case2, with a difference of +4.3% for case2. The cropping-based technique not only reduces the amount of computer resources required for occlusion detection, but also overcomes the problems of occlusion. However, the most crucial factor impacting the method's performance is **where to crop?**. Recognition performance of the three deep convolutional neural networks on MFR at different cropping proportions is tested and the results are tabulated in Table 4. The optimal cropping proportion for VGG-16 is at 0.85 with a recognition accuracy of 85.002% which shows an increase of 4.379% compared to No-Crop case. For Inception V3 and ResNet50 the optimum cropping proportion is at 0.8 and 0.95 with a recognition accuracy of 87.113% and 86.006%, respectively. Compared to No-Crop the recognition accuracy increased by 3.506% and 1.816% for InceptionV3 and ResNet50. From Table 3 and Table 4 it can be concluded that the cropping-based approach significantly improves the recognition accuracy. To further interpret the model's performance on

**Fig. 17** Results of the proposed VGG16-Random Fourier MFR model on identification for masked faces

MFR, class activation maps (CAM) are generated as in Fig. 14. The CAM heat maps are overlaid on the face image for better visualization. The heat map of the original face image concentrates on the eye and the lower face region (Fig. 14b), whereas the heat map of synthesized masked images focuses more on the eye and cheek regions. The cropping-based approach for MFR focuses on the area around eyes, which changes as the cropping proportion changes (Fig. 14.d, Fig. 14.e), so determining the best cropping proportion is critical in MFR. The visualization of CAM (Fig. 14.d, Fig. 14.e) shows that the cropping-based approach for MFR can precisely find the forehead-eye area with the use of prior information about mask location. More importantly, as shown by the proposed approach's

**Table 10** Results of VGG16-Random Fourier hybrid model on Masked Face Recognition

	Accuracy	mAP	Rank 1	Rank 5	Rank 10
Face dataset by robotics lab	97.460	73.87	78.67	84.78	89.56
Head pose image dataset	97.634	75.92	81.73	88.89	92.57
Georgia tech face dataset	97.552	74.67	79.95	86.67	90.83

**Fig. 18** Train-Validation curve of the proposed model on Georgia Tech Face dataset (with pose correction)

CAM maps, the most discriminative areas are not all areas above the mask, but the regions around two eyes.

## 5.2 Integration of random fourier layer with cropping based approach

To evaluate the effectiveness of the Random Fourier layer on feature enhancement for MFR, a cropping-based approach is integrated with a Random Fourier module. The recognition accuracy of different models at the optimal cropping proportion with pose correction, with/without Random Fourier layer is listed in Table 5. It is evident from the table that all the three models gave better accuracy when integrated with Random Fourier layer across all the three benchmark datasets: 85.0% to 97.46% for VGG 16, 87.113% to 93.836% for Inception V3, 86.066% to 95.992% for ResNet50 on Face dataset by Robotics lab. Similarly, 86.141% to 97.634% for VGG 16, 89.32% to 95.209% for Inception V3, 88.995% to 96.47% for ResNet50 on Head Pose Image Dataset and 86.079% to 97.552% for VGG 16, 88.541% to 94.936% for Inception V3, 87.380% to 96.002% for ResNet50 on Georgia Tech Face Dataset. The performance at different cropping proportions with Random Fourier module on Face dataset by Robotics lab is also tested and tabulated in Table 6. The optimal cropping proportion for VGG-16, Inception-V3, and ResNet50 remains the same at 0.85, 0.8, and 0.9, respectively. Compared to No- Crop, the recognition

accuracy increased by 11.515% for VGG-16, 8.822 % for InceptionV3 and 11.20% for ResNet50. Based on these observations we fixed VGG- 16 with Random Fourier layer as the base model for the proposed masked face recognition model because of its superior performance when compared to other models.

Line chart of accuracy at different cropping proportions without Random Fourier module and With Random Fourier Module for three models on Face dataset by Robotics lab is shown in Fig. 15. It can be observed from the figure that for all the three models accuracy increases initially as the cropping proportion increases and then decreases further. This may be because, as the cropping proportion increases masked part in the image also increase which in turn brings inadequate features. Table 7, Table 8 compares the effect of pose correction on the proposed hybrid model and basic VGG-16 model on MFR on three datasets. It is evident from the tables that, both architectures shows improved accuracy with pose correction (Table 8: 85.951%-including mask, 97.460%- excluding mask on Face dataset by Robotics Lab, 87.008%-including mask, 97.634%-excluding mask on Head Pose Image Dataset and 86.705%-including mask, 97.552%-excluding mask on Georgia Tech Face Dataset). To further evaluate the performance of the proposed MFR model, it is compared against state-of-the-art face recognition models on occlusion face recognition at the optimum cropping proportion. Arcface(Deng et al. 2019), Facenet[41]and Python face-recognition 1.3.0 module (Geitgey 2019) are considered for the comparison. The comparison results are tabulated in Table 9 and visualized in Fig. 16. It is clear from the Table 9 that: Our model outperforms Facenet and Arcface, which are the benchmark models for common face recognition by 11.13% (97.46-86.329), 4.85% (97.634-86.056) on Face dataset by Robotics Lab, 11.58% (97.634-86.056),4.344% (97.552- 86.19) on Head Pose Image Dataset and 11.362% (97.552- 86.19),4.82% (97.552-92.73) on Georgia Tech Face Dataset. It can be concluded from the result that models developed for common face recognition are inadequate for Masked Face Recognition since most facial landmarks are obscured by masks, whereas the inputs and feature receptive fields for common face recognition demand the entire face.

It is also evident that the proposed model gave accurate and generalized performance across the three datasets in terms of recognition accuracy. We display the top-7

retrieved images from the VGG16-Random Fourier hybrid model in Figure 17, and the model was able to retrieve a similar identity from the dataset for each of the query images. Table 10 compares the rank scores (rank1, rank5, rank10) and mean Average Precision(mAP) obtained for the proposed model across the three datasets. From the table, it is clear that the proposed model achieves best rank1 score of 81.73% and mAP of 75.92% on Head Pose Image dataset. The proposed model also produced comparable results on Face Dataset by Robotics Lab (Rank1 score of 78.67%, mAP of 73.87%) and on Georgia Tech Face Dataset (Rank1 score of 79.95%, mAP of 74.67%). To further evaluate the behavior of the proposed model, the training-validation graph is plotted in Fig. 18 which includes train-accuracy, validation-accuracy, train-loss, and validation-loss. We ran the model for 30 epochs, and it converged after the 25 epoch, as shown in Fig. 18. The best model with the highest validation accuracy (97.55) was saved and tested.

## 6 Conclusion

The COVID-19 epidemic necessitates people to wear masks; however, the presence of masks raises serious concerns about the accuracy of existing facial recognition systems, since the mask obscures most of the facial features. This paper proposes a cropping-based deep learning architecture to address the issue of masked face recognition. A hybrid VGG16-Random Fourier deep learning model is introduced to extract enhanced features from the upper half of the face, excluding masks for recognition. The selection of VGG-16 as the backbone network was done experimentally after comparing the recognition accuracies with respect to other benchmark CNN models in the literature. The proposed method consists of four major modules: Face alignment and preprocessing is the first module, Eye-Haar cascade configurations available in the OpenCV library are used to detect eye points. The frontal face obtained from the camera is oriented such that the face is perpendicular to the normal of the image with respect to the detected eye points. The second module is mask removal, which involves selecting the lower bounding box of the two predicted eyes and cropping the region below the bounding box to remove the mask from the image. The optimal cropping ratio is investigated to give a better recognition accuracy, and the results show that the optimal cropping proportion for VGG-16 is around 0.85L. The third module is enhanced feature extraction, which is accomplished through the use of the proposed hybrid VGG16-Random Fourier deep learning architecture. Module 4 combines Imagenet's transfer learned VGG16 architecture with a random Fourier layer to provide a better feature

representation in a lower-dimensional plane for masked face recognition. The effectiveness of the proposed model with/without mask is evaluated and verified on three benchmark datasets: Georgia Tech Face Dataset, Head Pose Image Dataset and Face Dataset by Robotics Lab. Experimental re- results show that the proposed approach can increase the recognition accuracy by an average of 10.99% on MFR. Overall, the paper's findings can be summarized as follows:

- Models developed for common face recognition are imprecise for masked face recognition because the mask obscures more than half of the face.
- Cropping-based approaches can significantly improve masked face recognition accuracy using transfer learned CNN models if enhanced feature extraction from the upper half is used.
- Cropping proportion impacts the recognition accuracy of the cropping-based approach for MFR. For VGG-16, the optimal cropping proportion was around 0.85L. The optimal cropping proportion for VGG-16 is around 0.85L for VGG-16. Integration of the optimal cropping and Random-Fourier module achieves the best recognition accuracy for MFR.
- In comparison to the state-of-the-art approach, the proposed VGG16-Random Fourier model delivers superior masked face recognition performance.

This work can be further extended to different application scenarios such as recognition of normal faces with masked faces, and recognition of faces within group/crowd images with multiple faces.

**Authorship contribution** All authors whose names appear on the submission made substantial contributions to the conception or design of the work; or the acquisition, analysis, or interpretation of data; or the creation of new software used in the work.

**Funding** This study was not funded by private/government bodies.

**Data availability** The datasets used during and/or analysed during the current study are publicly available. The corresponding papers are cited accordingly.

## Declarations

**Conflict of interest** The authors declare that they have no conflict of interest.

**Ethical approval** This chapter does not contain any studies with human participants or animals performed by any of the authors.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

## References

- Alzu'bi A, Albalas F, Al-Hadhrami T, Younis LB, Bashayreh A (2021) A masked face recognition using deep learning a review. *Electronics* 10(21):2666
- Andrés AM, Padovani S, Tepper M, Jacobo-Berlles J (2014) Face recognition on partially occluded images using compressed sensing. *Pattern Recognit Lett* 36:235–242
- Anwar A, Raychowdhury A (2020) Masked face recognition for secure authentication. arXiv preprint [arXiv:2008.11104](https://arxiv.org/abs/2008.11104)
- Balaji A, HS V, OK S (2016) Multimodal fingerprint spoof detection using white light. *Proced Comput Sci* 78:330–335
- Bansal M, Kumar M, Sachdeva M, Mittal A (2021) Transfer learning for image classification using vgg19: Caltech-101 image data set. *J Ambient Intell Humaniz Comput* 1–12
- Bochner S (1932) *Lectures on Fourier integrals*, vol. 12. Academic Publishing Company
- Bozinovski S (2020) Reminder of the first paper on transfer learning in neural networks, 1976. *Informatica*, 44(3)
- Bozinovski S, Fulgosi A (1976) The influence of pattern similarity and transfer of learning upon training of a base perceptron b2. (original in croatian: Utjecaj slicnosti likova i transfera ucenja na obucavanje baznog perceptrona b2). In: *Proc. Symp. Informatica*, pp. 3–121
- Cen F, Wang G (2019) Dictionary representation of deep features for occlusion-robust face recognition. *IEEE Access* 7:26595–26605
- Chu WSV, Chen JC, Lien JJJ (2007) Kernel discriminant analysis based on canonical differences for face recognition in image sets. In: *Asian Conference on Computer Vision*, Springer, p 700–711
- Deng H, Feng Z, Qian G, Lv X, Li H, Li G (2021) Mfcos-face: a masked-face recognition algorithm based on large margin cosine loss. *Appl Sci* 11(16):7310
- Deng J, Guo J, Xue N, Zafeiriou S (2019) Arcface: Additive angular margin loss for deep face recognition. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, p. 4690–4699
- Din NU, Javed K, Bae S, Yi J (2020) A novel gan-based network for unmasking of masked face. *IEEE Access* 8:44276–44287
- Dolhansky B, Ferrer CC (2018) Eye in-painting with exemplar generative adversarial networks. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7902–7911
- Du L, Hu H (2019) Nuclear norm based adapted occlusion dictionary learning for face recognition with occlusion and illumination changes. *Neurocomputing* 340:133–144
- Du H, Shi H, Liu Y, Zeng D, Mei T (2021) Towards nir-vis masked face recognition. *IEEE Signal Process Lett* 28:768–772
- Duan Q, Zhang L (2020) Look more into occlusion: realistic face frontalization and recognition with boostgan. *IEEE Transactions on Neural Networks and Learning Systems* 32(1):214–218
- Geitgey A (2019) Face recognition documentation. Release 1.2 3:3–37
- 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
- Golwalkar R, Mehendele N (2022) Masked-face recognition using deep metric learning and facemasknet-21. *Applied Intelligence* pp. 1–12
- Gourier N, Hall D, Crowley JL (2004) Estimating face orientation from robust detection of salient facial features. In: *ICPR international workshop on visual observation of deictic gestures*. citeseer
- Hariri W (2022) Efficient masked face recognition method during the covid-19 pandemic. *SIViP* 16(3):605–612
- Hariri W (2021) Efficient masked face recognition method during the covid-19 pandemic. arXiv preprint [arXiv:2105.03026](https://arxiv.org/abs/2105.03026)
- 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 L, Li H, Zhang Q, Sun Z (2018) Dynamic feature learning for partial face recognition. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7054–7063
- Hong JH, Kim H, Kim M, Nam GP, Cho J, Ko HS, Kim IJ (2021): A 3d model-based approach for fitting masks to faces in the wild. In: *2021 IEEE International Conference on Image Processing (ICIP)*, pp. 235–239. IEEE
- Karthika R, Parameswaran L (2016) Study of gabor wavelet for face recognition invariant to pose and orientation. In: *Proceedings of the international conference on soft computing systems*, Springer, pp. 501–509.
- Kumar A, Kaur A, Kumar M (2019) Face detection techniques A review. *Artif Intell Rev* 52(2):927–948
- Kumar A, Kumar M, Kaur A (2021) Face detection in still images under occlusion and non-uniform illumination. *Multimed Tools Appl* 80(10):14565–14590
- Lane L (2020) Nist finds flaws in facial checks on people with covid masks. *Biometric Technol Today* 2022:2
- Lfw recognition results-<https://vis-www.cs.umass.edu/lfw/results.html>
- Li Y, Guo K, Lu Y, Liu L (2021) Cropping and attention based approach for masked face recognition. *Applied Intelligence* 51(5):3012–3025
- Li C, Ge S, Zhang D, Li J (2020) Look through masks: Towards masked face recognition with de-occlusion distillation. In: *Proceedings of the 28th ACM International Conference on Multimedia*, pp. 3016–3024
- Liu B, Deng W, Zhong Y, Wang M, Hu J, Tao X, Huang Y (2019) Fair loss: Margin-aware reinforcement learning for deep face recognition. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10052–10061
- Lucena O, Junior A, Moia V, Souza R, Valle E, Lotufo R (2017) Transfer learning using convolutional neural networks for face anti-spoofing. In: *International conference image analysis and recognition*, Springer pp. 27–34
- Maharani DA, Machbub C, Rusmin PH, Yulianti L (2020) Improving the capability of real-time face masked recognition using cosine distance. In: *2020 6th International Conference on Interactive Digital Media (ICIDM)*, pp. 1–6. IEEE
- Nefian AV, Hayes MH (2000) Maximum likelihood training of the embedded hmm for face detection and recognition. In: *Proceedings 2000 International Conference on Image Processing (Cat. No. 00CH37101)*, vol. 1, pp. 33–36. IEEE
- Neha R, Nithin S (2018) Comparative analysis of image processing algorithms for face recognition. In: *2018 International Conference on Inventive Research in Computing Applications (ICIRCA)*, pp. 683–688. IEEE
- Nist finds facial recognition-<https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7494276/pdf/main.pdf>
- Priya GN, Banu RW (2014) Occlusion invariant face recognition using mean based weight matrix and support vector machine. *Sadhana* 39(2):303–315
- Schroff F, Kalenichenko D, Philbin J (2015) Facenet: A unified embedding for face recognition and clustering. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 815–823
- Shaheed K, Mao A, Qureshi I, Kumar M, Hussain S, Ullah I, Zhang X (2022) Ds-cnn: A pre-trained xception model based on depth

- wise separable convolutional neural network for finger vein recognition. *Expert Syst Appl* 191:116288
- Simonyan K, Zisserman A (2014) Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556)
- Singh S, Ahuja U, Kumar M, Kumar K, Sachdeva M (2021) Face mask detection using yolov3 and faster r-cnn models: Covid-19 environment. *Multimedia Tools and Applications* 80(13):19753–19768
- Singh T, Mary CJ (2016) Development of accurate face recognition process flow for authentication. In: *International Conference on Emerging Research in Computing, Information, Communication and Applications*, Springer pp. 3–16
- Soyel H, Demirel H (2010) Optimal feature selection for 3d facial expression recognition using coarse-to-fine classification. *Turkish Journal of Electrical Engineering & Computer Sciences* 18(6):1031–1040
- Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826 (2016)
- Wang Z, Wang G, Huang B, Xiong Z, Hong Q, Wu H, Yi P, Jiang K, Wang N, Pei Y et al. (2020) Masked face recognition dataset and application. arXiv preprint [arXiv:2003.09093](https://arxiv.org/abs/2003.09093)
- Wu G (2021) Masked face recognition algorithm for a contactless distribution cabinet. *Math Probl Eng* 2021:1–11
- Yuan, X., Park, I.K.: Face de-occlusion using 3d morphable model and generative adversarial network. In: *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 10062–10071 (2019)

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.