MFI3D: Masked-Face Identification with 3D Face Reconstruction and Deep Learning

Ahmad Alzu'bi[1*], Firas Albalas[1], Tawfik Al-Hadhrami[2], Amjad Albashayreh[1], Lojin Bani Younis[1]

[1*]Department of Computer Science, Jordan University of Science and Technology, Irbid, 22110, Jordan.
[2]School of Science and Technology, Nottingham Trent University, , Nottingham, NG11 8NS, UK.

*Corresponding author(s). E-mail(s): agalzubi@just.edu.jo;
Contributing authors: faalbalas@just.edu.jo;
tawfik.al-hadhrami@ntu.ac.uk; amalbashayreh20@cit.just.edu.jo;
lhbaniyounis19@cit.just.edu.jo;

Abstract

Masked face identification (MFI) aims to identify human faces that are obscured by masks or occluding objects. Today's common practice of wearing masks imposes significant barriers to facial recognition systems that rely on an unobstructed view of the face. To address this challenge, this paper introduces MFI 3D-based deep learning model (MFI3D) to identify occluded faces. Three main architectural components provide a set of discriminant features to decide the face identity, which are synthetic face masking, face unmasking with generator and discriminator, and 3D face reconstruction. The MFI3D pipeline begins by creating a syntactic mask for the face image, simulating real-world situations where the face is partially covered by the mask. Then, effective face detection is applied using a generator that learns to generate unmasked images that are indistinguishable from true unmasked images, and a discriminator that learns to distinguish real images from fake images. As a result, the MFI model can learn to reconstruct facial features from partially masked faces. The use of 3D face reconstruction techniques to generate a detailed model of faces leverages 3D geometry to extract facial features that are not visible in 2D image, providing a superior visual facial representation. Finally, the reconstructed face is matched against a collection of known people to determine their identity. Extensive experiments were conducted on facial datasets reconstructed orderly to build a diverse collection of 3D reconstructed facial images with a benchmarking ground-truth.

The experimental results show the superiority of the proposed MFI3D model in identifying people with occluded faces, achieving a precision of 83.50%.

Keywords: Face Identification; Masked Face Recognition; Deep Learning; 3D Imaging

1 Introduction

Over the past years, face recognition (FR) has undergone notable advancements in both theoretical and practical aspects. It has emerged as a cutting-edge study area in the discipline of pattern recognition. Since Turk et al. [1] presented the historical eigenface approach in the early 1990s, FR has garnered significant attention within the field of computer vision and is deemed one of the most interesting research areas. A face recognition system is used to recognize faces and verify individuals from images [2]. FR systems are frequently utilized to automatically and reliably authenticate a person for a variety of applications, including ATM, passport checking, smart doors, voter verification, criminal investigations, and many more. FR is considered the most widely used biometric technology for such purposes. Moreover, the recognition of occluded face images with objects such as masks, hairstyles, sunglasses, and hat occlusions, frequently causes difficulties in identifying human faces. These occlusions significantly impede the accurate recognition of human faces. A huge occlusion area can substantially damage the details in the original image, causing a recognition failure. The COVID-19 disease, officially declared a worldwide pandemic in 2020, is resulting in unprecedented chaos and disruption worldwide. To prevent the spread of the disease, it is recommended by the centers for disease control and prevention (CDC) [3] that individuals practice social distancing and wear face masks in public places[4].

However, people are occasionally forced, such as in airports and banks, to communicate closely or to remove their masks to disclose their identities. Therefore, it is critical to be able to recognize those wearing masks without requiring them to remove them in order to curb the spread of infection. Recently, many advanced models for masked face recognition (MFR) have been suggested and deployed in a wide range of applications, such as secure authentication at checkpoints [5] masked face tracking for people's safety, and monitoring people wearing face masks [6]. Face unmasking is the process of removing an occlusion from a facial image, such as a mask or a scarf. The purpose of face unmasking is to disclose the individual's true identity by removing the mask or other occlusion, allowing facial recognition technology to identify the person [7]. Face unmasking is necessary in a variety of applications, including security, surveillance, and forensic science. It's worth noting that, in light of the recent COVID-19 pandemic, face unmasking has been studied and improved in order to identify individuals concealing behind masks. Also, 3D face reconstruction and restoration is the process of creating a 3D model of a person's face from 2D images or videos artificially, then improving the appearance of that 3D model. Generating a 3D model typically includes multiple processes such as face detection, alignment, and triangulation. Once the 3D model is built, texture creation, color correction, and detail enhancement can be used to improve the appearance [8].

Due to factors such as the poor quality of the photos used for face unmasking, the results of face unmasking alone may not be sufficient to identify persons. Poorly lit or low-resolution images may not provide enough discriminating details for accurate face identification. The accuracy might also be affected by the quality of the mask being worn. Masks that cover most of the face or are constructed of materials that hide the face might make identifying the individual in 2D images more challenging. The features and form of a person's face can also change due to changing poses and expressions. The unmasking process can be affected by other occlusion such as glasses, hair, and scarves which can make it more difficult to identify the person in 2D images. Therefore, face unmasking in 2D form alone is not always sufficient to identify people. To increase identification precision, it has to be used with additional techniques like facial recognition and biometric analysis. Furthermore, 3D face reconstruction can be used to enhance 2D images and increase the reliability of facial identification. As a result, The accuracy of face identification can be improved by combining face unmasking with 3D face reconstruction since 3D reconstruction can produce an adequate 3D representation that can be used to recreate the face in various orientations, lighting conditions, and expressions [9].

The MFR algorithms, systems and datasets introduced in the literature lack a common development and assessment mainstream. Although there is a wide range of deep learning algorithms dedicated to detecting and recognizing people wearing masks, there is a need to examine the possibilities of such techniques in the MFR context. The existing works focused the attention on a single MFR component. They have discussed the procedures of face detection, face masking, face unmasking, face restoration, or face recognition, but no studies have considered all of the MFR components simultaneously. This motivated us to address the problem of MFR in the 2D and 3D space for face identification. In this paper, a novel pipeline, which we refer to as MFI3D, is introduced for identifying masked faces in a variety of settings and scenarios. The proposed framework integrates synthetic face masking, GAN-based unmasking, and 3D face reconstruction techniques to simulate the overall scenario of masked face identification. The aim of this method is to unmask and reconstruct the 3D contour of the face beneath the mask in order to restore the covered facial features, allowing more accurate identification of people.

The rest of this paper is organized as follows. Section 2 reviews the most related works that addressed the main MFR components; Section 3 introduces the image datasets used for training and evaluation; the proposed methodology is illustrated in Section 4; Section 5 presents the experimental setups and identification results of the proposed MFR model; Section 6 provides a thorough discussion of the model's performance; And section 7 concludes this paper.

2 Literature Review

This section presents the latest MFR works proposed literature focusing on masked face detection, face unmaking, face restoration, and face recognition.

2.1 Masked Face Detection

Face masks, which come in a wide range of shapes, sizes, textures, and colors, have recently become one of the most popular tools for concealing facial features. This points out the significance of accurately training deep learning algorithms to recognize masks [10]. As a face mask detection framework, Zhang et al. [11] proposed a context-attention R-CNN to increase intra-class distance and decrease inter-class distance by acquiring distinguishing features. To detect face masks on the simulated masked face dataset (SMFD), Negi et al. [12] suggested combining intelligent precautionary strategies with CNN and VGG16-based deep learning models. Peng et al. [13] discussed the use of local feature fusion-based deep networks for masked face detection in a nonlinear space. Fan et al [14] introduced RetinaFaceMask as a face mask detection technique evaluated on a new face mask detection dataset, MAFA-FMD, which improved the classification accuracy of mask-wearing states. In addition, the author presented CAM, a new attention module specialized in learning discriminating features related to face mask-wearing states. Alguzo et al [15] presented a multi-graph convolutional architecture for learning and detecting important aspects of occluded faces. This technique surpasses numerous previous state-of-the-art techniques and achieves a high detection accuracy of 97.9%. Jiang et al [16] developed SE-YOLOv3, a quick and accurate mask detection technique with a channel attention mechanism that enhanced the backbone network's feature extraction capabilities, employing 9205 high-quality masked facial images to construct the PWMFD. Furthermore, the authors used GIoU and focal loss, as well as data augmentation, to improve the accuracy of the proposed model. Ge et al [17] suggested a new model for masked face detection called LLECNNs and introduced a new dataset named MAFA. The authors discovered that MAFA is extremely difficult for existing face detectors, whereas the proposed model outperforms them in all scenarios. Face mask assistance (FMA), a face detection technique based on a mobile microscope, is another technique introduced by Chen et al. [18]. They employed the globally and locally consistent image completion (GLCM) to extract texture features from the face mask's microphotos, and they chose contrast, energy, homogeneity, and correlation as the facial features.

2.2 Face Unmasking

Face unmasking is a crucial step in the process of MFI because it can reconstruct the invisible parts of the face and provide a complete facial representation that can be used for face recognition tasks. Shetty et al. [19] suggested A GAN-based model that automatically removes the target object from an input image. Li et al. [20] and Iizuka et al. [21] presented a pair of distinct models to remove the target object and to reconstruct the damaged part while learning global coherency using a GAN setup. To eliminate items from facial images, Khan et al. [22] employed a coarse-to-fine GAN-based technique. Boutros et al. [23] introduced the embedding unmasking model (EUM) which utilizes a feature embedding acquired from a concealed face as its input. It generates a distinct feature embedding that shares similarities with a facial embedding of a similar identity without a mask. Din et al. [24] [7] used a GAN architecture with two discriminators to automatically take off the face mask. Criminisi

et al. [25] suggested a model in which the undesirable portion of an image is removed, a new region is created that fits the empty space, and the remaining portion of the image is then synthetically matched. Wang [26] developed a regularized factor for computing the filling order that adjusts the patch priority function curve. To remove eyeglasses from facial photos, Park et al. [27] employed iterative error correction and principal component analysis (PCA) reconstruction. Hays et al. [28] presented a technique for image completion that looks through a large repository of images for information that is similar and embeds it in the corrupted pixel of an input sample.

2.3 Face Restoration

Before executing the identity matching process to determine whether the identity is recognized or unrecognized, any missing features should first be estimated and recovered after the face has been unmasked. A great deal of work has gone into the sparse representation-based classification (SRC) [29] for robust OFR in image reconstruction. For example, the extended SRC (ESRC) was developed for the job of under-sampled FR [30], and group sparse coding (GSC) [31] was developed to enhance the discriminating capability of face reconstruction. Many alternative methods for reconstructing missing pieces of occluded faces have been proposed. Din et al. [7] [24] employed image-to-image translation to complete images using GAN-based image inpainting. Duan et al. [32] used GANs to manage face frontalization and face completion tasks at the same time. They presented a two-stage generative adversarial network (TSGAN) and an attention model based on occluded masks. Additionally, Luo et al. [33] introduced the EyesGAN framework, which is primarily designed to construct the face around the eyes, using GANs. Learning and preserving face completion network (LPFCN), which Ma et al. [34] introduced, can interpret face images and retrieve face identity-preserving (FIP) attributes simultaneously. This approach relies heavily on CNN, which has been trained to alter FIP characteristics.

2.4 Face Recognition

Many approaches, such as metric learning [35] and sparse representations[36], have been proposed to increase the discrimination level of deep features in order to make the face matching process more precise. Deep learning algorithms for identifying face identities have typically used the softmax loss-based and triplet loss-based models. In order to train a multi-class classifier for one class for each identity in the training dataset, softmax functions are used in softmax-loss models [37] [38]. On the other hand, triplet loss-based models instantly learn the embedding by comparing the outcomes of multiple inputs in order to reduce the intra-class distance and hence maximize the inter-class distance. However, facemask occlusions degrade the performance of softmax loss-based and triplet loss-based models [39] [40]. To address the MFR concerns, many research studies have recently been published in the literature. The use of GAN-based techniques to mask faces before feeding them to the face recognition model [24] [41], features extraction from the top part of the face [40], or training face recognition networks with a mix of masked and unmasked faces [42] [43] are some examples of effective approaches that have demonstrated high FR performance. Anwar et al. [42]

combined the augmented masked faces from the VGG2 dataset [44] with the dataset and trained the model using the original FaceNet pipeline [45], enabling the network to recognize whether a face is wearing a mask or not based on the characteristics of the top half of the face.

However, the MFR literature lacks a comprehensive framework that considers the three architectural components that aid in making accurate decisions about individuals' identities. Our work is distinguished from existing Masked Face Identification (MFI) approaches as we propose a novel 3D-based deep learning model that integrates synthetic face masking, face unmasking with GAN-based generator and discriminator, and 3D face reconstruction techniques to accurately identify occluded faces. This approach provides a more complete representation of the face, enabling accurate identification even in scenarios where the face is partially obscured by a mask or an occluding object.

Table 1 A summary of the related MFR approaches.

Ref.	Model	Method	Requirements	Dataset
[24]	GANs	Map and editing modules	VGG-19	CelebA
[35]	DMHSL	dynamic margin hard sampling loss	Triplet CNNs	LFW, YTF
[36]	SILR	SRC methods	l2-norm	AR, CMU PIE PIECAS-PEAL LFW, CelebA
[37]	CNN-based PDSN	Calculate equivalence between occluded facial blocks and damaged feature elements	MTCNN FCN ResNet	AR, LFW, RMF2
[38]	MSCAN	Spatial Transformer Networks with novel spatial constraints	DCNN	Market1501 CUHK03, MARS VIPeR
[40]	ArcFace SphereFace COTS	CNN,ReseNet-100 MegaMatcher 11.2 SDK	MTCNN	Collected dataset
[41]	GANs	De-occlusion distillation	-	Celeb-A, LFW, AR
[42]	MaskTheFace	MaskTheFace with FaceNet	FaceNet	VGGFace2-mini-SM LFW-SM
[43]	GANs	IAMGAN with DCR	-	MFSR CASIA-WebFace VGGFace2
[45]	FaceNet	a deep CNN	l2-norm Triplet Loss	LFW Youtube Faces DB

3 Datasets

Two benchmarking datasets are used to train and evaluate the 3DMFI model, which are: CelebAMask-HQ and Labeled Faces in the Wild (LFW).

3.1 CelebAMask-HQ

It is a large face image dataset made up of 30,000 high-resolution face images selected from the CelebA dataset [46] by CelebA-HQ [47]. Each image contains a segmentation mask of CelebA's facial features. CelebAMask-HQ masks have 512 x 512 resolution and Nineteen classes, which include all face details such as nose, mouth, eyes, ears, eyebrows, skin, lip, hat, hair, eyeglass, necklace, earring, cloth, and neck [48].

3.2 Labeled Faces in the Wild (LFW)

It is a dataset of 13,233 low-resolution images of people's faces gathered from the internet and annotated with their names. There are 5749 IDs in this dataset, and 1680 persons have two or more images. The verification accuracy on 6000 face pairs is measured in the conventional LFW evaluation technique. Face recognition algorithms are frequently trained and evaluated using this benchmarking dataset. The photos were obtained on the internet and depict a wide range of poses, expressions, and lighting conditions. [49].

Using the face masking technique, the 9000 photos from the CelebAMask-HQ dataset and all of the images from the LFW dataset were masked synthetically. The CelebAMask-HQ is primarily used to train the unmasking model, by which the complete model is trained using the masked photos. And the masked LFW images are used to test the entire framework and identify the faces. The statistics of both datasets are summarized in Table 2.

Table 2 Statistics of the used datasets.

Dataset	CelebAMask-HQ	LFW
Number of Images	30,000	13,233
Number of Identities	307	5749
Number of Masked Images	9000	13,233

4 Methodology

This section describes the development phases of the Masked Face Recognition system. The general methodology is based primarily on deep learning models, which are widely used to learn distinguishing features of masked faces. As illustrated in the following subsections, several critical steps are implemented toward creating the final Identification system.

4.1 The Procedure of Face Masking

Due to the lack of a dataset that has sufficient masked and unmasked facial image pairs, the (Mask the Face) tool [42] was used to generate synthetic masked datasets to train and evaluate the proposed unmasking model. Mask the Face is based on several computer vision and machine learning methods to generate masks and fit them into

facial images. The general structure of the masking tool, as shown in Figure 1, is as follows:

- The first step is to detect the face landmarks by focusing on the face tilt and six important key features such as the eyes, nose, lips, and face edges using a face landmark detection algorithm called dlib [50].
- The mask key positions in the face are then estimated.
- The corresponding mask pattern from the mask library is chosen based on face tilt.
- The mask is finally converted to fit properly on the face based on the six key features.

In our work, five mask types are used: KN95 mask, cloth mask, gas mask, surgical mask, and blue surgical mask. These masks are applied to create synthetic masked facial images for training and testing the proposed unmasking model. The tool was automated to apply these types of masks randomly and adjust the brightness of the masks based on the facial landmarks. As a result, a new synthetic masked CelebA_HQ collection is created for training, and a new synthetic masked LFW dataset is created for testing purposes.

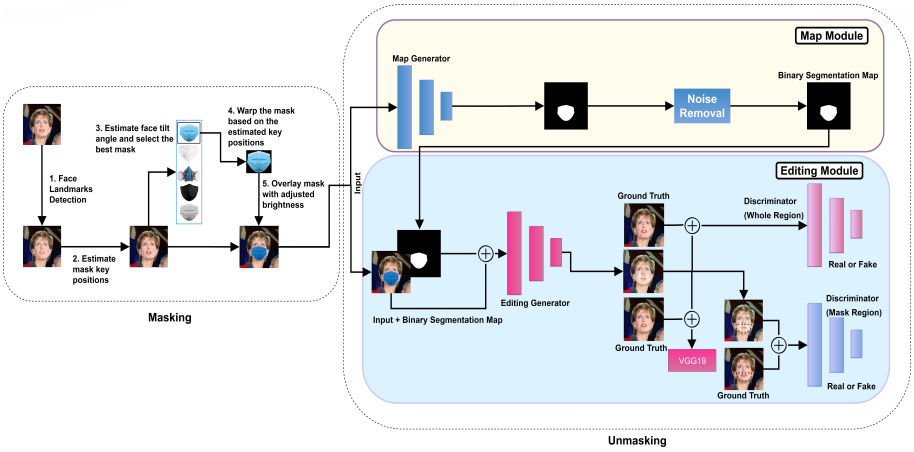

Fig. 1 The general MFI3D approach of face masking with two GAN-based discriminators.

4.2 The Procedure of Face Unmasking

In the proposed framework, the GAN-based unmasking approach [24] is adopted to unmask and reconstruct the facial features of partially masked faces. The overall structure of the 3DMFI pipeline is divided into two main modules: map module and editing module, as shown in Figure 1.

4.2.1 Map Module

The map module returns a binary segmentation map, I_mask map, with value 1 denoting the mask object and value 0 indicating the remaining image pixels. G_mask is a

map generator that uses a CNN-based encoder and decoder of a modified U-Net version [51]. The generator's encoder is composed of five blocks of convolutional layers. Except for the first encoder layer, each block represents a convolution layer followed by a L*relu* activation function and an instant_norm layer. The decoder architecture is identical to the encoder architecture, with the exception of a deconvolution layer replacing the convolution layer. The decoder's last layer employs the tanh activation function without a normalization layer. In addition, local and global information is incorporated by concatenating the deconvolution layer results with the encoder feature maps at the same level. The map generator network takes the image I_input as input and uses the encoder network to down-sample it to the bottleneck layer. The decoder network is then up-sampled for projecting a binary map.

4.2.2 Editing Module

This module's goal is to remove the mask and reconstruct the left-behind portion so that it is visually and structurally consistent with the ground-truth image. Using the input image I_{input} and the object map I_{mask_map}, it creates a complete image without the mask. The editing generator, discriminators, and perceptual network make up the primary components of this module.

Editing Generator: The editing generator G_{edit} shares the same design as the map generator G_{mask}. Unlike G_{mask}, the squeeze and excitation (SE) block [52] is used at the output of the encoder's first three blocks. Furthermore, four layers of atrous convolution, rate: 2,4,8,16, [53] are used between the encoder and decoder to make the missing part generation cohesive with the rest of the face image by collecting vast fields of view. The generator concatenates the input image, I_{input}, with the map module output, I_{mask_map}, to produce a new generated image, I_{edit}.

Visual Discriminator: This network contains two discriminators labeled as D_{whole_region} and D_{mask_region}. Both discriminators have the same architecture as the discriminator in pix2pix [54]. They penalize different structures at 70x70 patch scale. Both discriminators' roles are to force the editing generator to provide visually plausible and semantically compatible images. Instead of training both discriminators with the editing generator at the same time, the editing generator is trained alongside D_{whole_region} for the first 2/5 of the total training iterations. This helps to ensure that the generator's output is structurally compatible with the original input face image.

Perceptual Network: A perceptual network is the third component of the editing module. It is a pre-trained VGG19 network [55]. The goal of this network is to enable the generator output, I_edit, to have feature representations that are comparable to the ground truth, I_{gt}. By constructing a feature level distance measure between the intermediate feature maps of I_{edit} and I_{gt} based on a pre-trained network, a perceptual loss L_{perc} [56] is utilized to penalize the outputs that are not acceptable perceptually. Figure 2 shows the architecture of the generator and discriminators of the editing module. Table 3 demonstrates the hyperparameters of the weights used in the modules.

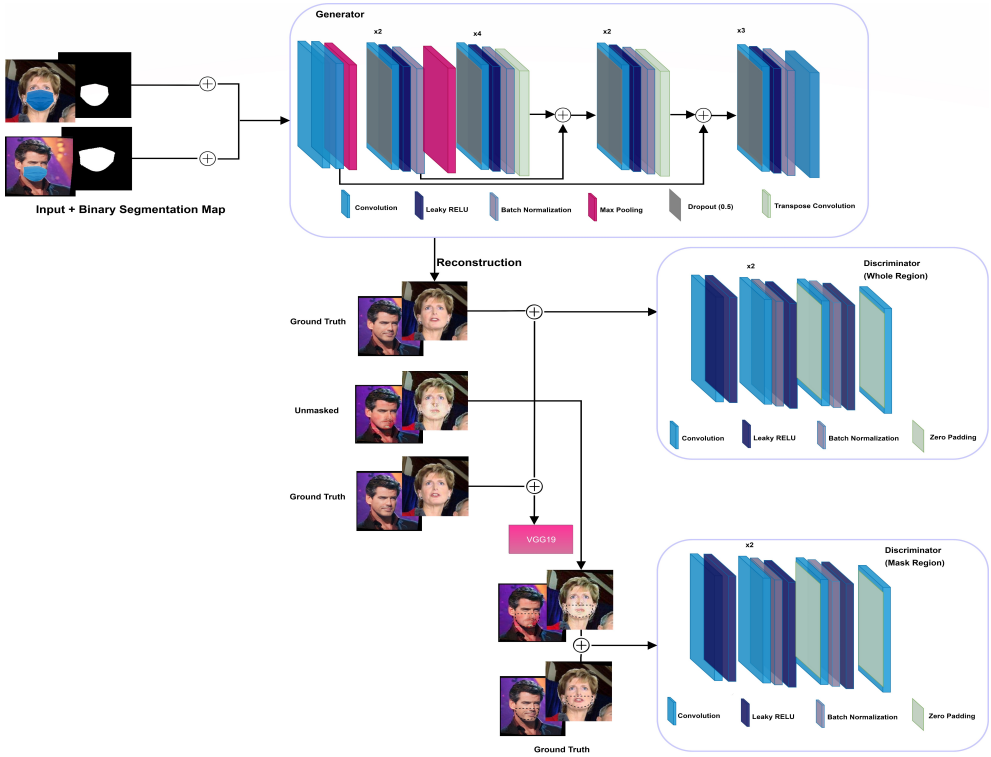

Fig. 2 The architecture of the Editing Module generator and two discriminators.

Table 3 Hyperparameters of the Unmasking Module.

Hyperparameter	Value
λ_{rc}	100
$\lambda_{D_{whole\_region}}$	0.3
$\lambda_{D_{mask\_region}}$	0.7
$\lambda_{adv_{whole\_region}}$	0.3
$\lambda_{adv_{mask\_region}}$	0.7

4.3 3D Face Reconstruction

3D face reconstruction is used to produce a digital representation of a person's face from 2D images or videos. It is crucial because it allows for more precise and detailed renderings of a person's face than conventional 2D techniques. Biometrics, virtual reality, animation, and forensic analysis are just a few of the many uses for 3D reconstruction. Face identification uses 3D face reconstruction to build a detailed and accurate facial image, which is then compared to a database of known faces. This has the potential to be utilized for the recognition of individuals within security systems, including building access control systems and mobile devices. Traditional 2D facial

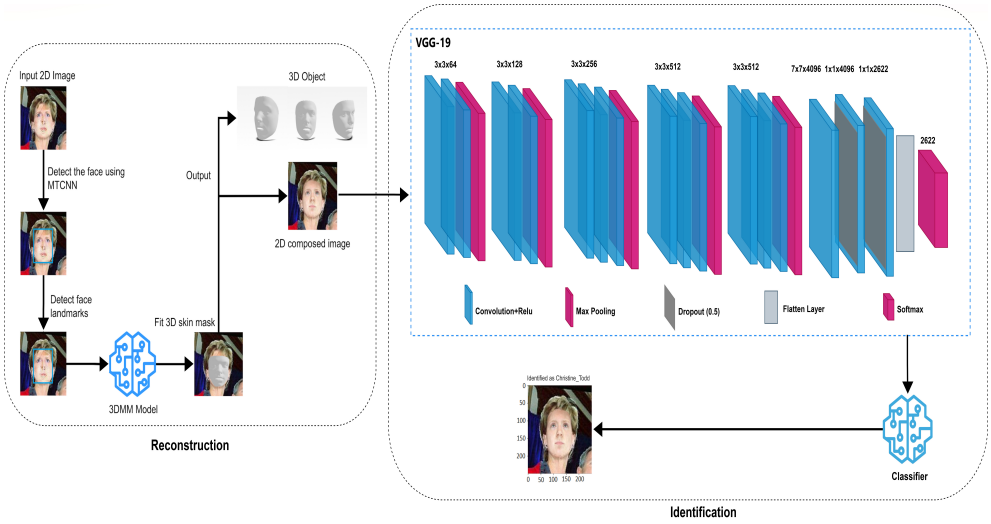

Fig. 3 The MFI3D pipeline of masked face reconstruction and identification.

recognition systems, which can be impacted by factors such as illumination and face angle, can also benefit from 3D face reconstruction. Furthermore, due to its consideration of facial depth and structure in addition to texture, 3D face recognition has the ability to exhibit greater resilience against attempts to deceive the system, such as the use of static images or videos depicting a person's face.

The 3D Morphable Model (3DMM) [57] [58] is a statistical model for generating 3D face shapes and textures. It is often generated by collecting a large set of 3D scans or images of real human faces and using statistical techniques to recreate the observed variations in shape and texture in the data. By changing a set of parameters that indicate variations in shape and texture, the model can be utilized to create new 3D faces after it has been generated. This permits more effective and accurate 3D face reconstruction when compared to current techniques that rely on manual intervention. The 3DMM is generally composed of two main parts: the shape model and the texture model. The shape model is a statistical model of 3D face shape variations. It is commonly represented as a linear combination of a collection of basis shapes intended to capture the most common shape variations found in the data. The linear combination coefficients are the shape model's parameters, and by varying the relative weights of the basic shapes, they may be used to create different face shapes. The texture model is a statistical model of the variations in the texture of the face, such as the skin color, wrinkles, and other details. Both shape and texture models are usually constructed from a large dataset of 3D scans or images of real human faces, using statistical techniques such as Principal Component Analysis (PCA) or Orthogonal Procrustes Analysis (OPA).

The 3DMM fitting approach was employed in this step of our approach, as shown in Figure 3, to create new 2D faces with 3D structures to reconstruct the faces and restore their features, making the process of facial recognition simpler. Using a training RGB

image, the R-Net is employed to calculate a coefficient vector x. This vector is then used in simple and differentiable mathematical derivations to generate an analytically reconstructed image. After detecting the facial landmarks, the 3D Morphable Model (3DMM) is employed to fit a 3D mask onto the face. This process involves utilizing the Base Face Model (BFM). The model's output is a 2D image of the input face with a 3D mask fitted onto it, resulting in a reconstructed face. The 3D mask is generated as a mesh output from the model, which is applied to both the ground truth and testing datasets.

4.4 Masked Face Identification

Face recognition is a method for determining a person's identity by examining and comparing patterns in their facial features. This process frequently involves the use of computer vision and machine learning algorithms that can automatically identify and evaluate a person's face in an image. The system then compares the extracted facial traits to a database of recognized faces in order to identify the person. Face matching in face recognition systems can be done by face verification (one-to-one matching) or face identification (one-to-many matching). In the case of our research, deep features are extracted from the most informative facial regions using a pre-trained variant of the VGG-19 model [55], and logistic regression is employed as a classifier. This section describes the procedure adopted for face identification including the deep learning model, feature extraction process, and the used classification method.

4.4.1 Feature Extraction

The VGG-19 network architecture is used to extract a set of image features used as key facial discriminators. The VGG-19 network incorporated in Figure 3 has a deep architecture consisting of 3×3 convolution layers, 1×1 padding layer, 2×2 pooling layers, and three fully connected layers. After preprocessing the images, they are fed sequentially into CNN as a multidimensional array of pixel densities to extract features from them. The input for RGB images is a 224×224×3 array. Each 2D convolutional layer filters the preceding layer, resulting in an activation output shape that represents the input of the subsequent layer. Also, to decrease the number of nodes by downsampling the activation maps, several max-pooling layers are added to the network architecture. The network's fully connected layers are used to learn the classification function. The feature descriptors are extracted from the first fully connected layer's output and L2-normalized by separating the components using the feature vector's L2-norm. We used a feature vector of size 1×4096. The generated normalized features are then used in the training, validation, and testing phases.

4.4.2 Classification Method

In the testing phase, the identification decision is made by employing logistic regression as the classifier to categorize unseen samples presented to the system. Additionally, PCA is utilized to reduce the dimensional subspace and enhance the performance of

the logistic regression classifier. This is achieved by generating an embedding vector for each image in the dataset using a pre-trained VGG-19 model, normalizing pixel values, scaling RGB values to the [0, 1] interval, encoding the labels, and standardizing the feature values. PCA is a useful technique for face identification as it has the potential to minimize the high-dimensional image descriptors, concentrate on the most relevant facial features, increase the algorithm's performance, and avoid overfitting.

5 Experimental Results

5.1 Experiments Setup

In the process of accomplishing the masked face identification task, various stages were implemented. However, our primary focus in this work revolved around two crucial phases: face unmasking and identification. These phases held significant importance in achieving the final result. The experimental configuration used in each step is detailed in the following subsections.

5.1.1 Face Unmasking Configuration

In the unmasking phase, we fed an input image I_input into the deep CNN-based network to build a binary map I_mask map that is similar to the target binary map I_tm for map module training. The resulting binary map I_mask is then fed into the editing network, alongside the input image I_input, to generate the final output I_edit. The model was trained on a set of 9,000 training synthetic images, each measuring 128×128, using a batch size of 10 and Adam optimizer for 20 epochs in the map module and 100 epochs in the editing module. The training of the editing module consists of several stages, starting by training the G_edit alone and then retraining it along with the D_whole_region for nearly half of the training iterations to establish an acceptable general shape of the face. Once a shape is created, D_mask_region is added to focus more on the missing area for the rest of the training cycles in order to construct the deep missing semantics more realistically.

5.1.2 Face Identification Configuration

The objective of the second experiment was to evaluate the performance of the enhanced VGG-19 model in the task of face identification specifically for reconstructed faces. In this sense, the system was initially fed a set of images of identified samples during the training and validation phases. In the testing phase, logistic regression was employed as a classifier to identify new unknown samples supplied to the system in order to make an identification decision. Moreover, an embedding vector was generated for each image in the dataset using the pre-trained VGG-19 model. Many post-processing steps were also applied to the extracted features, such as normalizing pixel values, scaling RGB values to the [0,1] interval, encoding the labels, standardizing the feature values, and utilizing PCA to reduce dimensional subspace and to improve the performance of the logistic regression classifier.

5.2 The Results of MFI3D Model

Under the same experimental conditions, the proposed model was evaluated in a number of distinct scenarios with various data amounts. The effectiveness of a face identification algorithm and its capacity for generalization might be impacted by the amount of the data. A balanced training and testing setup has been implemented in the first scenario. The model successfully identified masked faces with an accuracy score of 86.60%. Additionally, it showed excellent recall, precision, and F1-score. Figure 4 demonstrates a sample result from each phase of the MFI3D model, starting with the masking phase and ending with the identification phase.

Scenarios 2 and 3 are similar to Scenario 1, but with smaller test sets of 400 and 300 images, mimicking slightly more difficult real-world conditions. Despite the decreased test size, the model performed well, demonstrating its robustness under more restrictive conditions. The model produced consistently good outcomes. Regarding scenario 5, the smallest test set, which has only 100 images, has been used to simulate a scenario with restricted data availability. Despite the small sample size, the model performed well, demonstrating its ability to recognize masked faces in difficult settings. In scenario 6, a larger training set of 799 images was used. The model's accuracy decreased to 76.80%, indicating overfitting as a result of the larger dataset. Other metrics were similarly lower, highlighting a need for caution when utilizing a large training set. Scenario 7 is similar to Scenario 6 but with 400 images. The model showed better performance, but complaints about overfitting remained. Scenario 8 displays a

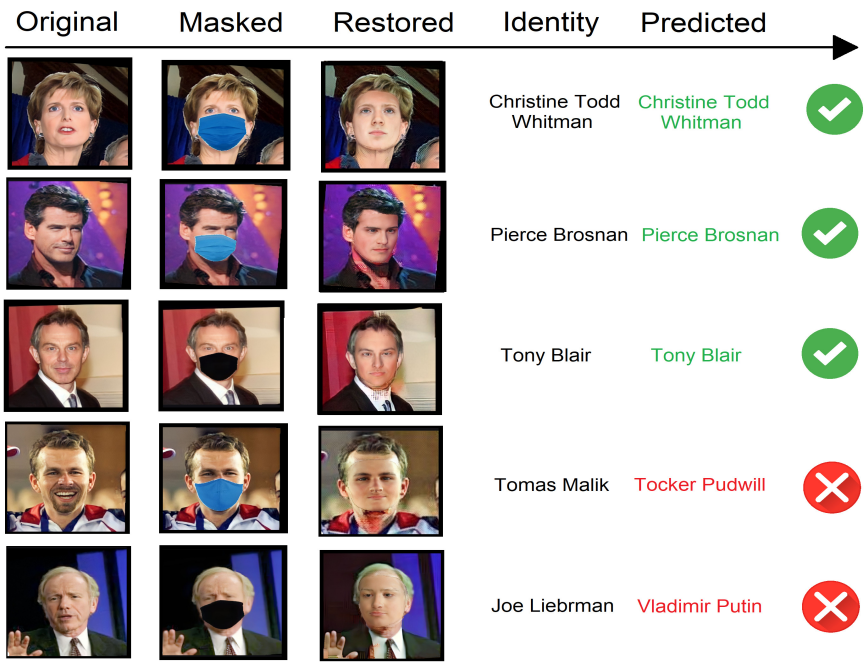

Fig. 4 Sample masked faces predicted by the proposed MFI3D model.

test of the model with 300 images to see if it can generalize more effectively. The model improved in accuracy and the other measures, demonstrating that generalization was improved by reducing the test size. Regarding scenarios 9 and 10, the test set was further reduced to 200 and 100 images. Even with a small test dataset, the model maintains consistent performance. Its high precision, recall, and F1-score demonstrate its dependability.

As shown in Table 4, the first five scenarios were trained with 500 images, while the remaining five scenarios were trained with 799 images. Among the ten scenarios evaluated, the initial five scenarios exhibited superior results compared to the latter five scenarios, primarily attributable to the disparity in training data size. Notably, when the model was trained with a larger dataset comprising 799 images, it acquired a higher degree of generalization capability. However, this increased generalization resulted in a reduction in the model's performance when dealing with unseen data. The inclusion of a larger dataset introduces a broader spectrum of examples and variations, which can potentially create confusion for the model and pose challenges in effectively handling diverse data variations. Additionally, when the model is trained on a smaller dataset but tested on a larger one, it may exhibit suboptimal performance due to overfitting. Overfitting implies that the model becomes excessively tailored to the training data and struggles to accurately identify unknown data instances. Hence, ten different scenarios were executed, employing varying data sizes, to assess the model's ability for identifying identities within the testing data. The highest performance was observed in Scenario 1, wherein 500 samples were used for both training and testing. It is worth noting that a larger testing dataset can contribute to minimizing result variance and enhancing performance robustness by offering a more comprehensive set

Table 4 A summary of the face identification results.

Scenario	Train size	Val size	Test size	Accuracy	Precision	Recall	F1-score
1	500	100	500	**86.60**	81.13	**86.60**	82.79
2	500	100	400	85.25	81.04	85.66	82.24
3	500	100	300	85.66	81.88	85.66	83.11
4	500	100	200	86.00	82.91	86.00	83.91
5	500	100	100	85.00	**83.50**	85.00	**84.00**
6	799	100	500	76.80	73.58	76.80	74.50
7	799	100	400	77.25	75.25	77.25	75.79
8	799	100	300	78.33	76.50	78.33	77.03
9	799	100	200	75.50	74.41	75.50	74.70
10	799	100	100	76.00	75.16	76.00	75.46

of representative samples of masked faces. However, it is important to acknowledge that a large testing dataset is not always necessary. Scenario five exemplifies this, as it achieved higher precision and F1-score with a smaller testing dataset, indicating that satisfactory performance can be achieved even with limited testing samples.

Leveraging a sizable testing dataset has the potential to enhance the generalizability of a model and reduce result variance. However, it is not universally indispensable or feasible in all situations. It is crucial to consider multiple factors, including the dataset size, the complexity of the problem at hand, and the variability across different scenarios, when determining the optimal testing dataset size. Due to the identical training and validation data amounts, the validation accuracy for the first five situations was the same as it was for the last five. As can also be observed in Table 4, the recall score is higher than the precision score in all scenarios. Recall measures how well the model detects all relevant occurrences, whereas precision measures how well the model avoids false positives. A high recall score indicates that the model can identify the majority of relevant MFI cases and has a low incidence of false negatives. However, the cost of false positives and false negatives is not the same in our case of MFI3D, and the decision is based on the cost of each type of error. As a result, it is critical to assess the cost of each type of error and to employ a measure that takes into account the system goal. In the context of face identification, a true positive occurs when the system accurately recognizes a facial image of an individual as a match to a reference image that has been stored.

On the other hand, a false positive happens when the system incorrectly identifies that two different facial images belong to the same person. Simply defined, a true positive indicates that the system correctly identified a person, whereas a false positive indicates that the system mistakenly identified a person. This work aims to improve the system's ability to detect people's identities by increasing true positives and decreasing false positives. For example, if a face identification system is used to identify someone entering a building, a true positive happens if the system successfully identifies the person as the one who has already registered. A false positive arises when the system erroneously identifies an individual as the registered person and grants them access to the building. In such instances, the cost associated with false positives is considerable, and it is desirable to minimize it to the greatest extent possible. Precision holds significant importance in face identification as it evaluates the system's ability to correctly identify positive matches out of all positive identifications, encompassing both true and false positives. It essentially quantifies the accuracy of the system in recognizing individuals. A low precision implies that the system produces a considerable number of false positive identifications, which can introduce security concerns and inconvenience for users.

False positive identifications can be time and resource-consuming. As an illustration, when a face identification system erroneously identifies an individual as a match to a stored reference image, security personnel may need to intervene and verify the person's identity. This can lead to delays and additional expenses incurred for authentication purposes. When the system inaccurately identifies an individual (false positive identification), it can give rise to privacy concerns, including unintended disclosure of personal information or data breaches. Consequently, in our study, we considered

precision as a significant metric to evaluate the accuracy and reliability of the MFI3D model. By minimizing costs and safeguarding privacy, we aimed to ensure the model's overall effectiveness. In this study, weighted average metrics were employed to assess the performance of MFI3D. This approach takes into account the varying impact of each metric on the final performance evaluation, considering factors such as the class distribution in the dataset and the desired outcome. By utilizing weighted average metrics, a broader range of results is obtained, allowing for a more comprehensive evaluation of MFI3D's performance. The best performance outcomes were achieved in the first and fifth scenarios, showcasing a precision of 83.50%, a recall of 86.60%, and an F1-score of 84.0%. Figure 5 depicts the validation and testing accuracy for each scenario. Our initial goal was to evaluate the model's performance under various conditions by proposing diverse scenarios. However, in Scenario 1, the model was trained and tested in a balanced setup, which was critical in achieving extremely close results in terms of validation and testing accuracy. The model achieved 89.0% validation accuracy and 86.60% testing accuracy, suggesting strong performance. Scenario 8 involved training the model on a larger set of 799 samples while testing it on a smaller dataset of 300 identities. It obtained 80.0% validation accuracy and 78.33% testing accuracy, which is lower than Scenario 1 but still represents a good level of performance. Therefore, scenarios 1 and 8 are considered the best scenarios in terms of accuracy due to their close accuracy rates.

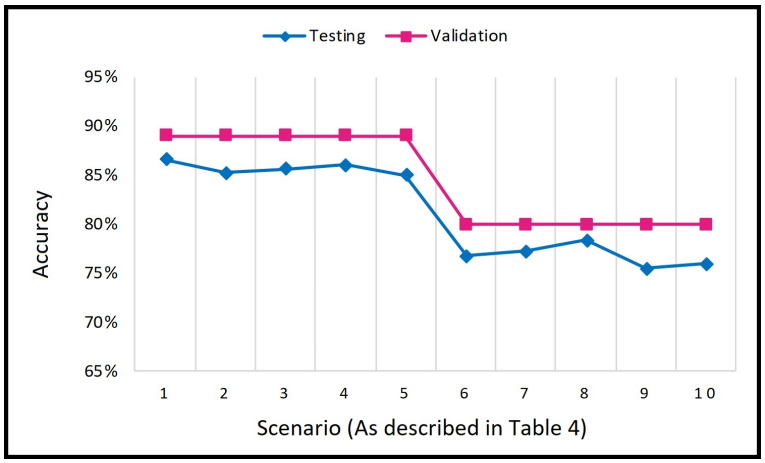

Fig. 5 The MFI3D validation and testing accuracy using various scenarios.

6 Discussion

6.1 Semantic Identification of Masked Faces

To the best of our knowledge, no prior studies have specifically tackled the issue of masked face identification by incorporating the three fundamental architectural components that offer a collection of distinctive features: masking, unmasking, and

subsequent 3D face restoration. The process of face masking is an essential component of our proposed MFI3D framework. The purpose of the synthetic face masking is to mimic several scenarios in which the face may be partially covered by a mask or any occluding items, giving a broad and thorough benchmarking ground-truth for the MFI3D evaluation. To accomplish this, a synthetic face masking technique was employed, which involved superimposing different types of masks onto the face. Various synthetic masks, characterized by distinct shapes and sizes, were created and seamlessly applied to the face to achieve a realistic appearance. The masks were subsequently eliminated from these masked face images using a GAN-based generator and discriminator that had been trained to learn the mapping between masked and unmasked faces. Then the 3DMM was applied to reconstruct the faces after unmasking them. As a result, we were able to create a broad and comprehensive dataset of occluded and non-occluded facial images, providing a credible and representative benchmark for testing and assessing the MFI3D model. Overall, the synthetic face masking technique utilized in this work offers a flexible and adaptable approach, enabling the modeling of diverse masking scenarios and the evaluation of MFI models under various conditions. This technique provides more semantic facial traits to reliably identify occluded faces, even in cases where facial features are partially concealed by masks or other occluding objects. As a result, it enhances the performance of facial recognition systems in real-world applications, contributing to improved accuracy and effectiveness.

Furthermore, previous studies that implemented 3D reconstruction on faces have demonstrated the effectiveness of this approach in enhancing the accuracy of face identification. For instance, Zhang et al. [59] described a technique for identifying faces from unconstrained images using a combination of face frontalization, data representation, and support vector-guided dictionary learning. The approach utilized in this study entailed the automatic generation of frontal views of faces from unconstrained images, achieved through the utilization of a single 3D face model. Subsequently, the face regions were segmented, and the coding vectors' discrimination quality was enhanced using an adaptive weighting scheme. This methodology aimed to improve the accuracy and effectiveness of face recognition by effectively handling unconstrained images and optimizing the encoding process. The authors evaluated the system on the LFW database and achieved an identification accuracy of 97.17% using 7 images per individual for training and 3 images per individual for testing. They also suggested that the use of a 3D face model for generating frontal views contributes to the effectiveness of the proposed method because the 3D model can be used to estimate the pose and illumination of the face in the original image, and then generate a frontal view that is more consistent and easier to recognize. Moreover, the utilization of 3D modeling can effectively address several challenges encountered in unconstrained face recognition, including occlusion, partial views, and image noise. By leveraging a 3D model, it becomes possible to infer missing information from the original image, leading to the generation of more accurate representations of the face [60]. This enables improved robustness and accuracy in recognizing faces that would otherwise be challenging with traditional 2D approaches.

Yu et al. [61] a 2D-aided deep 3D face identification framework that employed a deep learning-based 3D face reconstruction method. This approach involved reconstructing numerous 3D face scans from a vast 2D face database. Subsequently, normal component images (NCI) derived from the reconstructed 3D face scans were utilized to train a deep convolutional neural network (DCNN). By combining 2D and 3D techniques, this framework aimed to enhance the performance of face identification by exploiting the benefits of both modalities. The proposed approach achieved state-of-the-art rank-1 scores on the FRGC v2.0 (97.6%), Bosphorus (98.4%), and BU-3DFE (98.8%) databases, demonstrating the usefulness of reconstructed 3D facial surfaces and the effectiveness of 3D modeling in improving the accuracy of face identification. Additionally, Papadopoulos et al. [62] proposed an approach that represented each dynamic sequence of facial expressions as a spatio-temporal graph, which is constructed using 3D facial landmarks. They emphasized that the use of 3D facial landmarks in constructing the spatio-temporal graph and the subsequent application of the ST-GCN highlights the effectiveness of 3D modeling in capturing the dynamic nature of facial expressions and improving the accuracy of face identification. Liu et al. [63] also proposed a framework for regressing dense 3D face shapes from single 2D images while explicitly addressing the identity and residual components of the 3D face shape. This is accomplished using a composite 3D face-shape model with latent representations. The training process for the proposed network involves a joint loss function measuring both face identification error and 3D face shape reconstruction error. To construct the training data, the authors developed a method for fitting a 3DMM to multiple 2D images of a subject.

As a result, these studies provide compelling evidence that the integration of 3D modeling can play a pivotal role in face identification, enhancing recognition accuracy and effectively addressing challenges like occlusion, partial views, and image noise. In alignment with this notion, our study leverages 3D modeling with 3D Morphable Models (3DMM) to further enhance the accuracy of identifying unmasked faces. By reconstructing the face and generating a discriminating representation, the application of 3D modeling augments the precision and reliability of the identification process.

6.2 The Quality of Facial Deep Features

This work used the deep VGG-19 architecture as a baseline due to its effectiveness in serving many computer vision tasks. Therefore, we have modified the VGG-19 architecture to make better learnable feature maps from the input images. The baseline VGG-19 has a huge number of layers and parameters, which allow it to capture complex features and patterns in images, which aid the MFI3D framework in making highly accurate decisions. Also, VGG-19 can be easily adjusted for face identification purposes. For example, the last few layers of the network can be adjusted to generate feature vectors that represent individual faces, which can subsequently be used to determine people's identities. The convolutional layers of VGG-19 are all 2D convolutional layers, which means they only work on 2D images. The filters in a neural network's convolutional layer are the weights that are learned during training to recognize particular features or patterns in the input data. Each convolutional layer in VGG-19 has a collection of learnable filters used to extract various features from the

input image. The filters in each convolutional layer are 3x3 matrices of weights. As we progress deeper into the architecture, the number of filters in each convolutional layer rises, from 64 in the first layer to 512 in the last. Each filter is applied to a different region of the input image, and the output is then passed through an activation function to generate the associated feature map. The filters are randomly initialized and then trained to minimize the loss function using back-propagation. They are tuned during training to become progressively more specialized in detecting specific properties of the input data. The filters in the first convolutional layers learn to detect simple features such as edges or corners, but the filters in the deeper layers learn to detect more complex elements like face landmarks, textures, and shapes.

The filters are designed to reduce the distance between embeddings of the same identity's face while increasing the distance between embeddings of other identities' faces. The input image is subjected to a series of filters, creating a collection of feature maps. The locations and intensities of the learned features in the input image that the network has identified through the convolution operation are displayed on a feature map, which is a representation of the output of a convolutional layer. It can be challenging to interpret deep neural networks like VGG-19 due to their complexity and nonlinearity. The representation of the input data by the network and the features it uses to create predictions can both be seen visually in feature maps. We can learn more about how the network processes the input data and what kinds of features it is learning by looking at the feature maps at various layers of the network. Therefore, we constructed a modified model that is a subset of the convolutional layers by removing the pooling, activation, dropout, and flatten layers of the original VGG-19 model, which they used to visualize the feature maps. There are five basic blocks in this scheme, which culminate in a pooling layer.

The size of the filters varies by block. Using this new model, we predicted a list of feature maps, with one feature map output for each of the last convolutional layers in each block. The resultant feature map of the first, middle, and final layers is obtained to depict the learned features in each layer, as shown in Figure 6. We limited

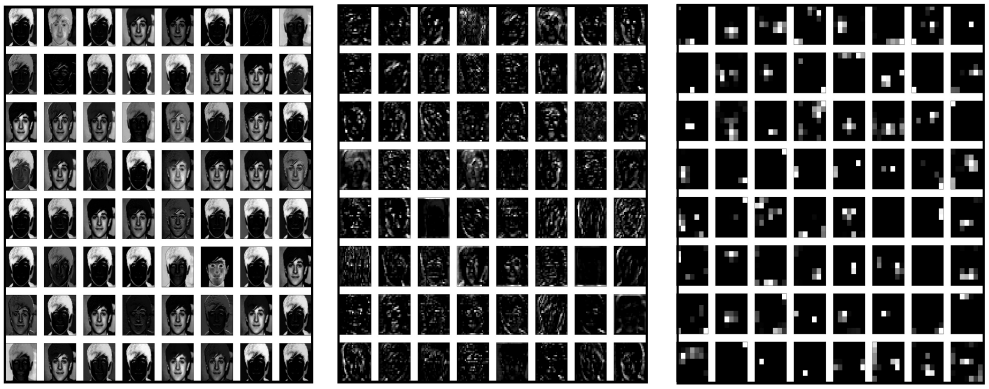

Fig. 6 The feature maps output for multiple convolutional layers.

the number of feature maps presented to 64 for consistency, although the first layer actually contains 64 feature maps that increase in depth to 256 and 512. In the visual representation, the dark hues correspond to small weights, while the light colors indicate large weights. It is noticeable that the feature maps located closer to the model's input effectively capture finer details present in the image. However, as we move further into the model, the displayed feature maps gradually lose some of these intricate details. This pattern is commonly observed in the VGG-19 model, where the deeper feature maps in the final layers tend to exhibit a higher level of abstraction and complexity. Consequently, interpreting and analyzing these feature maps becomes more challenging when compared to the early feature maps.

6.3 Limitations of the Study

Over the past two decades, we've witnessed significant technological transformations driven by disruptive advancements in both software and hardware. A key feature of these changes is the integration of the digital realm with the physical world through IoT technologies. Notably, the latest shift is from a focus on "connected things" to a new era of "connected intelligence" [64]. This shift is particularly significant in the context of MFI, as AI systems now benefit from this connected intelligence ecosystem, allowing them to utilize data from IoT devices, collaborate with other intelligent entities, and make more context-aware and informed identification decisions. This evolution enhances the overall effectiveness and relevance of MFI in our increasingly interconnected world.

While the proposed framework demonstrates effectiveness in accurately identifying masked faces, it is important to acknowledge certain limitations encountered in the study. For instance, the current version of the proposed MFI3D is limited to identifying individuals based on their facial features alone. Other factors, such as body shape, clothing, and context, can also contribute to the identification process in real-world scenarios. Another limitation is that the synthetic face masking technique used to create the benchmarking ground-truth may not fully capture the complexity and variability of real-world masking scenarios. Although we made efforts to create a diverse and representative dataset of occluded and non-occluded facial images, it is necessary to investigate a wider range of scenarios that may be not fully accounted for in the synthetic collection. Additionally, our proposed model relies on the assumption that the facial features of partially masked faces can be accurately reconstructed using 3D face reconstruction and GAN-based unmasking techniques. However, there are instances where the obscured facial features can be excessively complex or ambiguous, making it difficult to reconstruct them accurately. As a result, this can lead to inaccuracies in the identification process. Future work could involve the integration of additional modalities and the collection and annotation of real-world datasets comprising occluded facial images. By doing so, the aim is to enhance the generalization and applicability of the MFI3D in real-world scenarios.

The 3D face reconstruction algorithms typically require large amounts of computing power and memory to process and store the 3D models. This can limit the scalability and practicality of the approach, particularly in real-time applications or scenarios with limited computing resources. Furthermore, the quality and type of input

images can significantly impact the sensitivity of 3D face reconstruction. The accuracy and robustness of the 3D reconstruction can be affected by various factors, such as lighting conditions, camera angles, occlusion levels, and facial expressions. There are instances where the input images may contain excessive noise or incomplete information, resulting in challenges when attempting to generate accurate 3D models. Also, the 3DMM suffers from sensitivity to facial expressions and poses. 3DMMs are usually trained on neutral faces, which means that they may not accurately represent the variation in facial shape and appearance under different expressions or poses.

Future research endeavors could concentrate on investigating the utilization of 3DMMs specifically designed to effectively handle diverse facial expressions and poses. This could involve training 3DMMs on a wider range of facial expressions and poses to capture the variation in facial shape and appearance that occurs under these conditions. A potential avenue is to mutually integrate 3DMMs with complementary techniques, such as deep learning, to develop resilient models that possess the capability to handle an extensive array of variations. This amalgamation has the potential to empower the models in effectively accommodating a broader range of facial expressions, poses, and other diverse factors.

7 Conclusion

The significance of masked face identification arises from the prevalent adoption of masks as a preventive measure against the transmission of COVID-19. The presence of masks poses challenges in identifying individuals solely based on their facial features, thereby diminishing the effectiveness of conventional face recognition technologies. This paper introduced a framework for masked face identification employing synthetic masking, face unmasking, and 3D face reconstruction techniques. The integration of masking, unmasking, and 3D face reconstruction holds significance as it capitalizes on the unique strengths of each technique to effectively address the challenges presented by facial masks or occluding object. By leveraging the capabilities of these techniques in combination, it becomes possible to overcome the obstacles associated with masked face identification. The strength of the synthetic masking phase can be evaluated by how well it occludes the face; the strength of the unmasking phase can be evaluated by how well it reveals the occluded face; and the strength of the reconstruction phase can be evaluated by how well it recreates the full face from the unmasked information. The performance of the proposed MFI3D model was assessed across various scenarios involving diverse data sizes, and the outcomes were promising and encouraging. Exhibiting an accuracy of 86.60 and a precision of 83.50, the model demonstrated its capability to accurately identify individuals, even in instances where their faces are partially concealed by a mask. This underscores its reliability in masked face identification tasks. The achieved precision score indicates that the system successfully identifies people with high accuracy while limiting false positive identifications. Across all scenarios, the model consistently showed higher recall than precision results, indicating its proficiency in identifying a significant portion of relevant cases while maintaining a low occurrence of false negatives. There exist numerous intriguing avenues for future research, including expanding the presented

approach to encompass traditional occlusion objects like haircuts, sunglasses, and hats. Additionally, exploring methods to preserve facial sentiment features in order to generate more realistic reconstructed images could also be a promising direction for further investigation.

Funding: This research was funded by the Jordan University of Science and Technology, grant number 20210166.

Data Availability Statement: The images in this study were taken and analyzed from the publicly archived dataset with no ethics implications. CelebA datasets are available in the MMLAB repository, https://mmlab.ie.cuhk.edu.hk/projects/CelebA.html, and LWF dataset is available in the UMass repository, http://vis-www.cs.umass.edu/lfw/. The new collection of images generated synthetically and processed during the current study is available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

References

[1] Turk, M.A., Pentland, A.P.: Face recognition using eigenfaces. In: Proceedings. 1991 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 586–587 (1991). IEEE Computer Society

[2] Chaudhari, J.: Face recognition using pca algorithm. Inventi Rapid: Image & Video Processing (2012)

[3] CDC: Centers for Disease Control and Prevention. U.S. Department of Health & Human Services (2022). https://www.cdc.gov/

[4] CDC: Coronavirus Disease 2019 (COVID-19) (2020). https://www.cdc.gov/coronavirus/2019-ncov/prevent-getting-sick/

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

[6] Prasad, S., Li, Y., Lin, D., Sheng, D.: maskedfacenet: a progressive semi-supervised masked face detector. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pp. 3389–3398 (2021)

[7] Din, N.U., Javed, K., Bae, S., Yi, J.: Effective removal of user-selected foreground object from facial images using a novel gan-based network. IEEE Access **8**, 109648–109661 (2020)

[8] Lin, J., Yuan, Y., Shao, T., Zhou, K.: Towards high-fidelity 3d face reconstruction from in-the-wild images using graph convolutional networks. In: Proceedings of

the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5891–5900 (2020)

[9] Alzu'bi, A., Albalas, F., Al-Hadhrami, T., Younis, L.B., Bashayreh, A.: Masked face recognition using deep learning: A review. Electronics **10**(21), 2666 (2021)

[10] Albalas, F., Alzu'bi, A., Alguzo, A., Al-Hadhrami, T., Othman, A.: Learning discriminant spatial features with deep graph-based convolutions for occluded face detection. IEEE Access **10**, 35162–35171 (2022)

[11] Zhang, J., Han, F., Chun, Y., Chen, W.: A novel detection framework about conditions of wearing face mask for helping control the spread of covid-19. Ieee Access **9**, 42975–42984 (2021)

[12] Negi, A., Kumar, K., Chauhan, P., Rajput, R.: Deep neural architecture for face mask detection on simulated masked face dataset against covid-19 pandemic. In: 2021 International Conference on Computing, Communication, and Intelligent Systems (ICCCIS), pp. 595–600 (2021). IEEE

[13] Peng, X.-Y., Cao, J., Zhang, F.-Y.: Masked face detection based on locally non-linear feature fusion. In: Proceedings of the 2020 9th International Conference on Software and Computer Applications, pp. 114–118 (2020)

[14] Jiang, M., Fan, X., Yan, H.: Retinamask: A face mask detector. arXiv preprint arXiv:2005.03950 (2020)

[15] Alguzo, A., Alzu'bi, A., Albalas, F.: Masked face detection using multi-graph convolutional networks. In: 2021 12th International Conference on Information and Communication Systems (ICICS), pp. 385–391 (2021). IEEE

[16] Jiang, X., Gao, T., Zhu, Z., Zhao, Y.: Real-time face mask detection method based on yolov3. Electronics **10**(7), 837 (2021)

[17] Ge, S., Li, J., Ye, Q., Luo, Z.: Detecting masked faces in the wild with lle-cnns. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2682–2690 (2017)

[18] Chen, Y., Hu, M., Hua, C., Zhai, G., Zhang, J., Li, Q., Yang, S.X.: Face mask assistant: Detection of face mask service stage based on mobile phone. IEEE Sensors Journal **21**(9), 11084–11093 (2021)

[19] Shetty, R.R., Fritz, M., Schiele, B.: Adversarial scene editing: Automatic object removal from weak supervision. Advances in Neural Information Processing Systems **31** (2018)

[20] Li, Y., Liu, S., Yang, J., Yang, M.-H.: Generative face completion. In: Proceedings

of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3911–3919 (2017)

[21] Iizuka, S., Simo-Serra, E., Ishikawa, H.: Globally and locally consistent image completion. ACM Transactions on Graphics (ToG) **36**(4), 1–14 (2017)

[22] Khan, M.K.J., Ud Din, N., Bae, S., Yi, J.: Interactive removal of microphone object in facial images. Electronics **8**(10), 1115 (2019)

[23] Boutros, F., Damer, N., Kirchbuchner, F., Kuijper, A.: Unmasking face embeddings by self-restrained triplet loss for accurate masked face recognition. arxiv 2021. arXiv preprint arXiv:2103.01716

[24] Din, N.U., Javed, K., Bae, S., Yi, J.: A novel gan-based network for unmasking of masked face. IEEE Access **8**, 44276–44287 (2020)

[25] Criminisi, A., Pérez, P., Toyama, K.: Region filling and object removal by exemplar-based image inpainting. IEEE Transactions on image processing **13**(9), 1200–1212 (2004)

[26] Wang, J., Lu, K., Pan, D., He, N., Bao, B.-k.: Robust object removal with an exemplar-based image inpainting approach. Neurocomputing **123**, 150–155 (2014)

[27] Park, J.-S., Oh, Y.H., Ahn, S.C., Lee, S.-W.: Glasses removal from facial image using recursive error compensation. IEEE transactions on pattern analysis and machine intelligence **27**(5), 805–811 (2005)

[28] Hays, J., Efros, A.A.: Scene completion using millions of photographs. ACM Transactions on Graphics (ToG) **26**(3), 4 (2007)

[29] Wright, J., Yang, A.Y., Ganesh, A., Sastry, S.S., Ma, Y.: Robust face recognition via sparse representation. IEEE transactions on pattern analysis and machine intelligence **31**(2), 210–227 (2008)

[30] Deng, W., Hu, J., Guo, J.: Extended src: Undersampled face recognition via intraclass variant dictionary. IEEE Transactions on Pattern Analysis and Machine Intelligence **34**(9), 1864–1870 (2012)

[31] Huang, J., Nie, F., Huang, H., Ding, C.: Supervised and projected sparse coding for image classification. In: Twenty-Seventh AAAI Conference on Artificial Intelligence (2013)

[32] Duan, Q., Zhang, L.: Look more into occlusion: Realistic face frontalization and recognition with boostgan. IEEE transactions on neural networks and learning systems **32**(1), 214–228 (2020)

[33] Luo, X., He, X., Qing, L., Chen, X., Liu, L., Xu, Y.: Eyesgan: Synthesize human

face from human eyes. Neurocomputing **404**, 213–226 (2020)

[34] Ma, R., Hu, H., Wang, W., Xu, J., Li, Z.: Photorealistic face completion with semantic parsing and face identity-preserving features. ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM) **15**(1), 1–18 (2019)

[35] Yu, J., Hu, C.-H., Jing, X.-Y., Feng, Y.-J.: Deep metric learning with dynamic margin hard sampling loss for face verification. Signal, Image and Video Processing **14**, 791–798 (2020)

[36] Yang, S., Wen, Y., He, L., Zhou, M., Abusorrah, A.: Sparse individual low-rank component representation for face recognition in the iot-based system. IEEE Internet of Things Journal **8**(24), 17320–17332 (2021)

[37] Song, L., Gong, D., Li, Z., Liu, C., Liu, W.: 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 (2019)

[38] Li, D., Chen, X., Zhang, Z., Huang, K.: Learning deep context-aware features over body and latent parts for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 384–393 (2017)

[39] Lane, L.: Nist finds flaws in facial checks on people with covid masks. Biom. Technol. Today **8**(2), 30101–6 (2020)

[40] Damer, N., Grebe, J.H., Chen, C., Boutros, F., Kirchbuchner, F., Kuijper, A.: The effect of wearing a mask on face recognition performance: an exploratory study. In: 2020 International Conference of the Biometrics Special Interest Group (BIOSIG), pp. 1–6 (2020). IEEE

[41] Li, C., Ge, S., Zhang, D., Li, J.: Look through masks: Towards masked face recognition with de-occlusion distillation. In: Proceedings of the 28th ACM International Conference on Multimedia, pp. 3016–3024 (2020)

[42] Anwar, A., Raychowdhury, A.: Masked face recognition for secure authentication. arXiv preprint arXiv:2008.11104 (2020)

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

[44] Cao, Q., Shen, L., Xie, W., Parkhi, O.M., Zisserman, A.: 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 (2018). IEEE

[45] Schroff, F., Kalenichenko, D., Philbin, J.: Facenet: A unified embedding for face recognition and clustering. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 815–823 (2015)

[46] Liu, Z., Luo, P., Wang, X., Tang, X.: Deep learning face attributes in the wild. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3730–3738 (2015)

[47] Karras, T., Aila, T., Laine, S., Lehtinen, J.: Progressive growing of gans for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196 (2017)

[48] Lee, C.-H., Liu, Z., Wu, L., Luo, P.: Maskgan: Towards diverse and interactive facial image manipulation. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 5549–5558 (2020)

[49] Huang, G.B., Mattar, M., Berg, T., Learned-Miller, E.: Labeled faces in the wild: A database forstudying face recognition in unconstrained environments. In: Workshop on Faces in'Real-Life'Images: Detection, Alignment, and Recognition (2008)

[50] http://dlib.net/

[51] Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18, pp. 234–241 (2015). Springer

[52] Hu, J., Shen, L., Sun, G.: Squeeze-and-excitation networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7132–7141 (2018). IEEE

[53] Chen, L.-C., Papandreou, G., Schroff, F., Adam, H.: Rethinking atrous convolution for semantic image segmentation. arXiv preprint arXiv:1706.05587 (2017)

[54] Isola, P., Zhu, J.-Y., Zhou, T., Efros, A.A.: Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5967–5976 (2017)

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

[56] Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: European Conference on Computer Vision, pp. 694–711 (2016). Springer

[57] Blanz, V., Vetter, T.: A morphable model for the synthesis of 3d faces. In: Proceedings of the 26th Annual Conference on Computer Graphics and Interactive Techniques, pp. 187–194 (1999)

[58] Blanz, V., Vetter, T.: Face recognition based on fitting a 3d morphable model. IEEE Transactions on pattern analysis and machine intelligence **25**(9), 1063–1074 (2003)

[59] Zhang, Z., Xu, X., Liang, J., Sun, B.: Unconstrained face identification based on 3d face frontalization and support vector guided dictionary learning. Mathematical Problems in Engineering **2020**, 1–16 (2020)

[60] Alzu'bi, A., Younis, L.B., Madain, A.: An interactive attribute-preserving fashion recommendation with 3d image-based virtual try-on. International Journal of Multimedia Information Retrieval **12**(2), 24 (2023)

[61] Yu, C., Zhang, Z., Li, H.: Reconstructing a large scale 3d face dataset for deep 3d face identification. arXiv preprint arXiv:2010.08391 (2020)

[62] Papadopoulos, K., Kacem, A., Shabayek, A., Aouada, D.: Face-gcn: A graph convolutional network for 3d dynamic face identification/recognition. arXiv preprint arXiv:2104.09145 (2021)

[63] Liu, F., Zhu, R., Zeng, D., Zhao, Q., Liu, X.: Disentangling features in 3d face shapes for joint face reconstruction and recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5216–5225 (2018)

[64] Jiang, Y., Li, X., Luo, H., Yin, S., Kaynak, O.: Quo vadis artificial intelligence? Discover Artificial Intelligence **2**(1), 4 (2022)