Archives of Computational Methods in Engineering (2022) 29:3475–3507
https://doi.org/10.1007/s11831-021-09705-4

REVIEW ARTICLE

3D Face Reconstruction in Deep Learning Era: A Survey

Sahil Sharma[1] · **Vijay Kumar**[2]

Received: 10 October 2021 / Accepted: 21 December 2021 / Published online: 10 January 2022

Abstract
3D face reconstruction is the most captivating topic in biometrics with the advent of deep learning and readily available graphical processing units. This paper explores the various aspects of 3D face reconstruction techniques. Five techniques have been discussed, namely, deep learning, epipolar geometry, one-shot learning, 3D morphable model, and shape from shading methods. This paper provides an in-depth analysis of 3D face reconstruction using deep learning techniques. The performance analysis of different face reconstruction techniques has been discussed in terms of software, hardware, pros and cons. The challenges and future scope of 3d face reconstruction techniques have also been discussed.

1 Introduction

3D face reconstruction is a problem in biometrics, which has been expedited due to deep learning models. Several 3D face recognition research contributors have improved in the last five years (see Fig. 1). Various applications such as reenactment and speech-driven animation, facial puppetry, video dubbing, virtual makeup, projection mapping, face aging, and face replacement are developed [1]. 3D face reconstruction faces various challenges such as occlusion removal, makeup removal, expression transfer, and age prediction. Occlusion can be internal or external. Some of the well-known internal occlusions are hair, beard, moustache, and side pose. External occlusion occurs when some other object/person is hiding the portion of the face, viz. glasses, hand, bottle, paper, and face mask [2].The primary reason behind the growth of research in 3D face reconstruction is the availability of multicore central processing units (CPUs), smartphones, graphical processing unit (GPU) and cloud applications such as Google Cloud Platform (GCP), Amazon Web Services (AWS), and Microsoft Azure [3–5]. 3D data is represented in voxels, point cloud, or a 3D mesh that GPUs can process (see Fig. 2). Recently, researchers have started working on 4D face recognition [6, 7]. Figure 3 depicts the taxonomy of 3D face reconstruction.

1.1 General Framework of 3D-Face Reconstruction Problem

3D reconstruction-based face recognition framework involves pre-processing, deep learning, and prediction. Figure 4 shows the phases involved in the 3D face restoration technique. There are various forms of 3D images that can be acquired. All of them have different pre-processing steps based on the need. Face alignment may or may not be done for sending it to the reconstruction phase. Sharma and Kumar [2, 8, 9] have not used face alignment for their reconstruction techniques.

The face reconstruction can be done using a variety of techniques, viz. 3D morphable model-based reconstruction, epipolar geometry-based reconstruction, one-shot learning-based reconstruction, deep learning-based reconstruction, and shape from the shading-based reconstruction. Further, the prediction phase is required as the outcome of the reconstruction of the face. The prediction may be based on applications of face recognition, emotion recognition, gender recognition, or age estimate.

1.2 Word Cloud

The word cloud represents the top 100 keywords of 3D face reconstruction (see Fig. 5). From this word cloud, the keywords related to face reconstruction algorithm such as "3D

✉ Sahil Sharma
sahil301290@gmail.com

Vijay Kumar
vijaykumarchahar@gmail.com

[1] Computer Science and Engineering Department, Thapar Institute of Engineering and Technology, Patiala, India

[2] Computer Science and Engineering Department, National Institute of Technology, Hamirpur, India

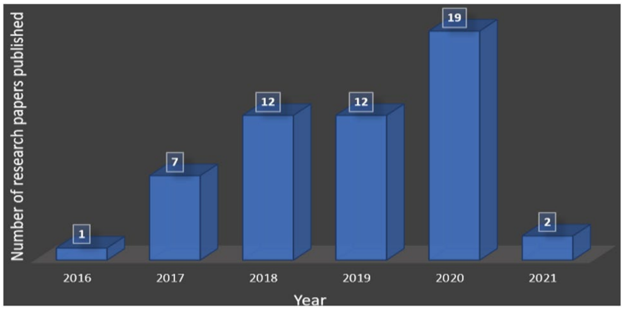


Fig. 1 Number of research papers published in 3D face reconstruction from 2016–2021

face", "pixel", "image", and "reconstruction" are widely used. The keyword "3D face reconstruction" has fascinated the researchers as a problem domain of face recognition techniques.

Face reconstruction involves completing the occluded face image. Most 3D face reconstruction techniques use 2D images during the reconstruction process [10–12]. Recently, researchers have started working on mesh and voxel images [2, 8]. Generative adversarial networks (GANs) are used for face swap and facial features modification [13] in 2D faces. These are yet to be explored using deep learning techniques.

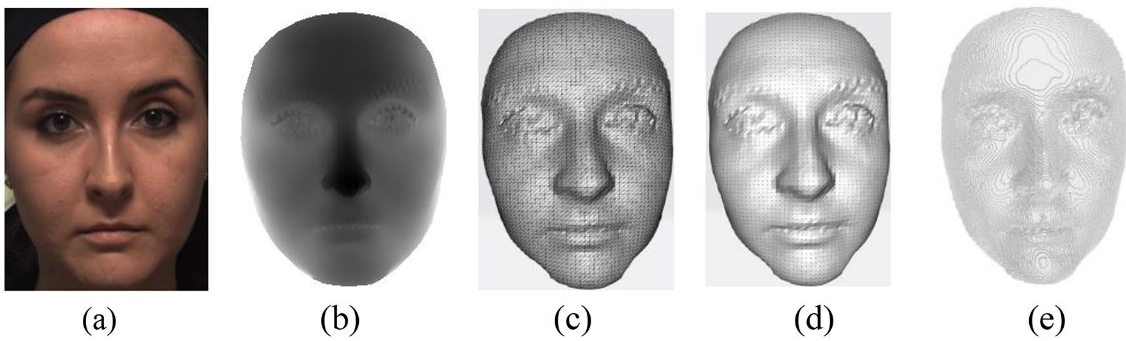


Fig. 2 3D face images: **a** RGB Image, **b** Depth Image, **c** Mesh Image, **d** Point Cloud Image, **e** voxel Image

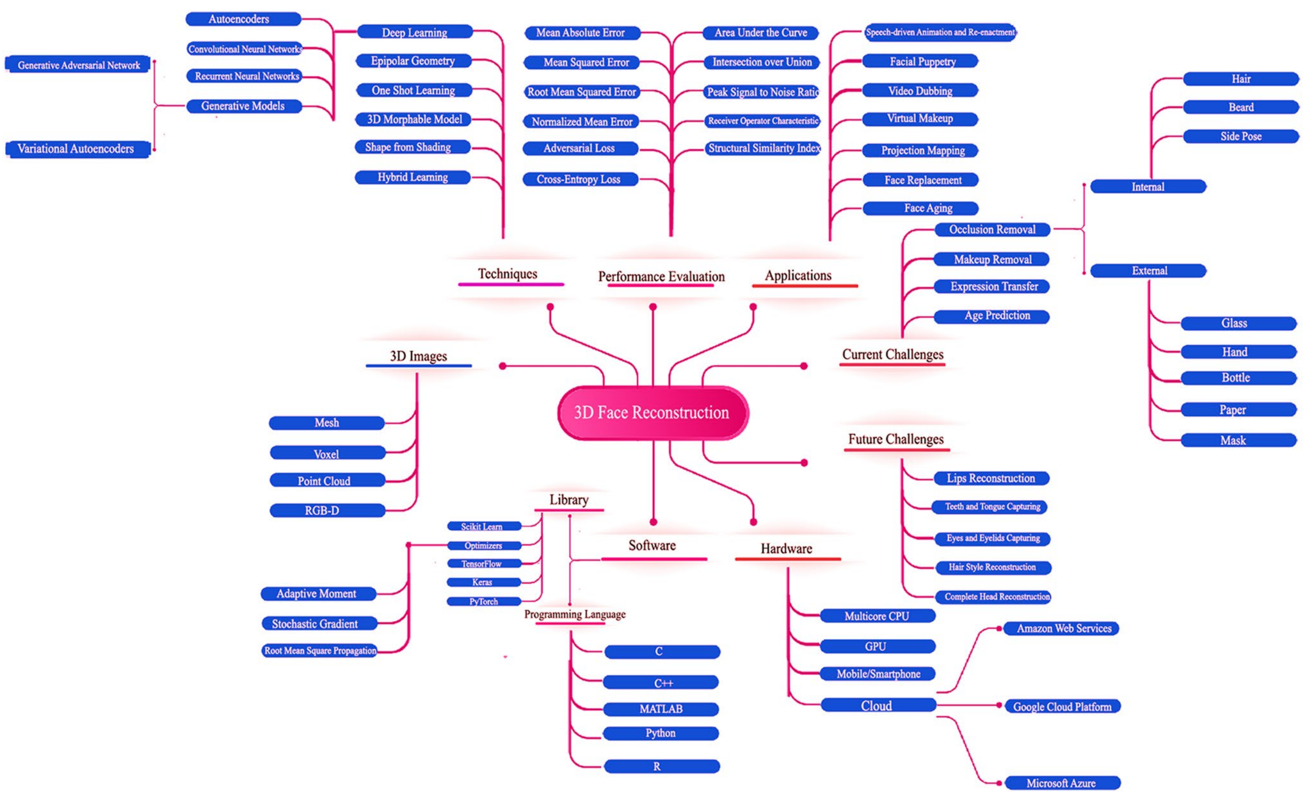


Fig. 3 Taxonomy of 3D face reconstruction

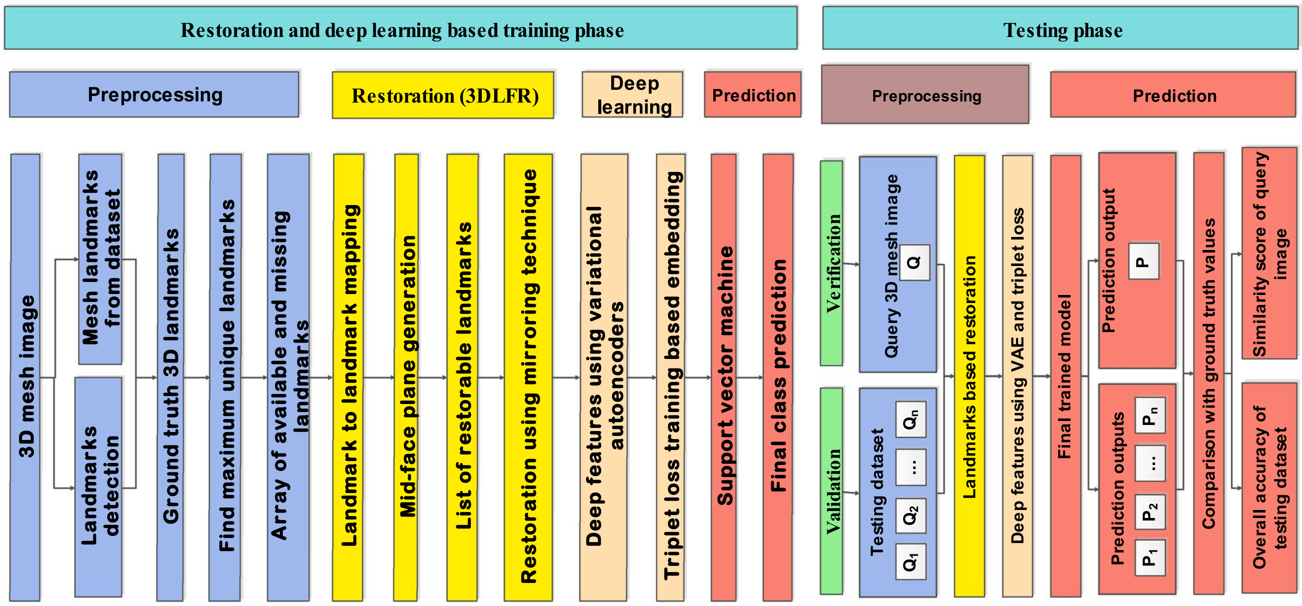


Fig. 4 General Framework of 3D Face Reconstruction Problem [9]

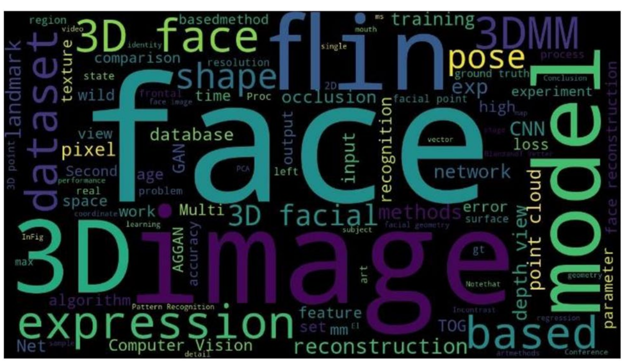


Fig. 5 Word cloud of 3D face reconstruction literature

The presented work's motivation lies in the detailed research surveys with deep learning of 3d point clouds [14] and person re-identification [15]. As seen in Fig. 1, in the last five years, 3d face research has grown with every passing year. Most of the reconstruction research has preferred using GAN-based deep learning techniques. This paper aims to study 3D face reconstruction using deep learning techniques and their applications in a real-life scenario. The contribution of this paper is four-fold.

1. Various 3D face reconstruction techniques are discussed with pros and cons.
2. The hardware and software requirements of 3D face reconstruction techniques are presented.
3. The datasets, performance measures, and applicability of 3D face reconstruction are investigated.
4. The current and future challenges of 3D face reconstruction techniques have also been explored.

The remainder of this paper is organised as follows: Sect. 2 covers variants of the 3D face reconstruction technique. Section 3 discusses the performance evaluation measures followed by datasets used in reconstruction techniques in Sect. 4. Section 5 discusses the reconstruction process' tools and techniques. Section 6 discusses 3D face reconstruction's potential applications. Section 7 summarises current research challenges and future research directions. Section 8 holds the concluding remarks.

2 3D Face Reconstruction Techniques

3D face reconstruction techniques are broadly categorised into five main classes such as 3D morphable model (3DMM), deep learning (DL), epipolar geometry (EG), one-shot learning (OSL), and shape from shading. Figure 6 shows the 3D face reconstruction techniques. Most of the researchers are working on hybrid face reconstruction techniques and are considered as sixth class.

2.1 3D Morphable Model-based Reconstruction

A 3D Morphable Model (3DMM) is a generative model for facial appearance and shape [16]. All the faces to be generated are in a dense point-to-point correspondence, which can be achieved through the face registration process. The morphological faces (*morphs*) are generated through dense

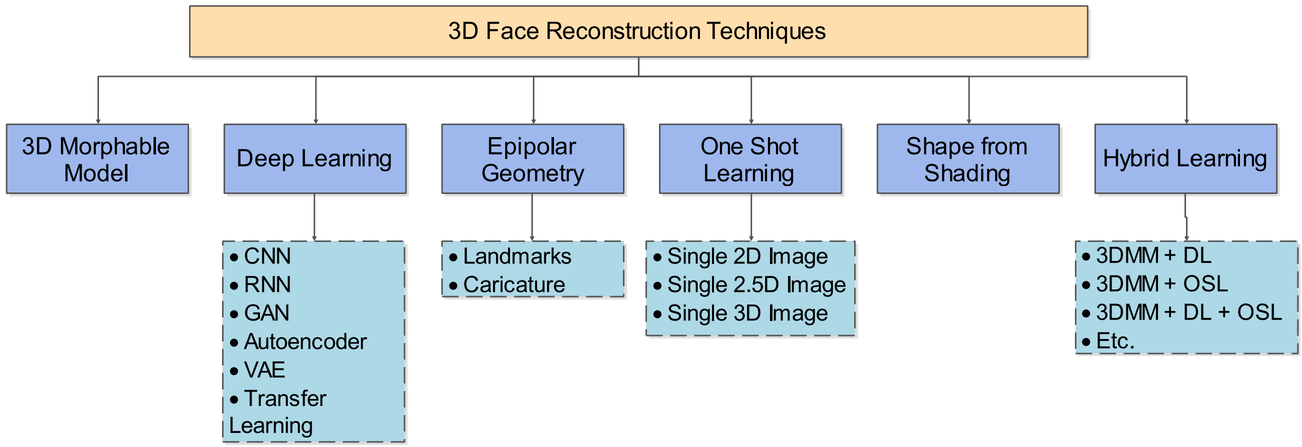


Fig. 6 3D Face Reconstruction Techniques

correspondence. This technique focuses on disentangling the facial colour and shape from the other factors, such as illumination, brightness, contrast, etc. [17]. 3DMM was introduced by Blanz and Vetter [18]. Variants of 3DMM are available in the literature [19–23]. These models use low-dimensional representations for facial expressions, texture, and identity. Basel Face Model (BFM) is one of the publicly available 3DMM models. The model is constructed by registering the template mesh corresponding to the scanned face obtained from Iterative Closest Point (ICP) and Principal Component Analysis (PCA) [24].

Figure 7 shows the progressive improvement in 3DMM during the last twenty years [18, 25–28]. The results from the original paper of Blanz and Vetter 1999 [18], the first publicly available Morphable Model in 2009 [25], and state-of-the-art facial re-enactment results [28] and GAN-based models [27] have been presented in the figure.

Maninchedda et al. [29] proposed an automatic reconstruction of the human face and 3D epipolar geometry for eye-glass-based occlusion. A variational segmentation model was proposed, which can represent a wide variety of glasses. Zhang et al. [30] proposed reconstructing a dense 3D face point cloud from a single data frame captured from an RGB-D sensor. The face region's initial point cloud was captured using the K-Mean Clustering algorithm. An artificial neural network (ANN) estimates the neighbourhood of the point cloud.

In addition, Radial Basis Function (RBF) interpolation is used to achieve the final approximation of the 3D face centred on the point cloud. Jiang et al. [31] proposed a 3D face restoration algorithm (PIFR) based on 3DMM. The input image was normalised to get more information about the visibility of facial landmarks. The pros of the work are pose-invariant face reconstruction. However, the reconstruction needs improvement over large poses. Wu et al. [32]

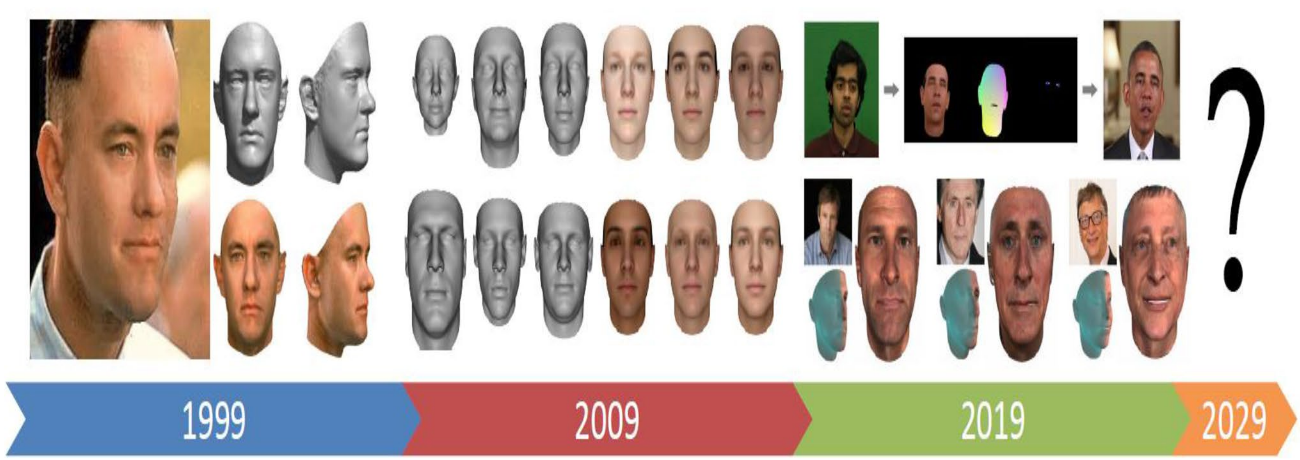


Fig. 7 Progressive improvement in 3DMM over last twenty years [17]

presented a 3D face expression reconstruction technique using a single image. A cascaded regression framework was used to calculate the parameters for 3DMMs. The histogram of oriented gradients (HOG) and landmark displacement was used for the feature extraction phase. Kollias et al. [33] proposed a novel technique for synthesising facial expressions and the degree of positive/ negative emotion. Based on the valence-arousal (VA) technique, 600 K frames were annotated from the 4DFAB dataset [34]. This technique works for in-the-wild face datasets. However, 4DFAB is not publicly available. Lyu et al. [35] proposed a Pixel-Face dataset consisting of high-resolution images by using 2D images. Pixel-3DM was proposed for 3D facial reconstruction. However, the external occlusions are not considered in this study.

2.2 Deep Learning-based Reconstruction

3D generative adversarial networks (3DGANs) and 3D convolutional neural networks (3DCNN) are the deep learning techniques of 3D face reconstruction [27]. The main advantages of these methods are high fidelity and better performance in terms of accuracy and mean absolute error (MAE). However, it takes a lot of time to train GANs. The face reconstruction in canonical view can be done through the face-identity preserving (FIP) method [36]. Tang et al. [37] introduced a multi-layer generative deep learning model for image generation under new lighting situations. In face recognition, the training corpus was responsible for providing the labels to the multi-view perceptron-based approach. Synthetic data were augmented from a single image using facial geometry [38]. Richardson et al. [39] proposed the unsupervised version of the reconstruction mentioned above. Supervised CNNs were used to implement facial animation tasks [40]. 3D texture and shape were restored using deep convolutional neural networks (DCNNs). In [41], facial texture restoration provided better fine details than the 3DMM [42]. Figure 8 shows the different phases of 3D face recognition using the restoration of the occluded region.

Kim et al. [26] proposed a deep convolutional neural network-based 3D face recognition algorithm. 3D face augmentation technique synthesised a variety of facial expressions using a single scan of the 3D face. Transfer learning-based model is faster to train. However, 3D data is lost when the 3D point cloud image is converted to a 2.5D image. Gilani et al. [43] proposed a technique for developing a huge corpus for labelled 3D faces. They trained a face recognition 3D convolutional neural network (FR3DNet) to recognise 3D faces over 3.1 million faces of 100 K people. The testing was done on 31,860 images of 1853 people. Thies et al. [44] presented a neural voice puppetry technique for generating photo-realistic output video from the source input audio. This was based on DeepSpeech recurrent neural networks using the latent 3D model space. Audio2ExpressionNet was responsible for converting the input audio to a particular facial expression.

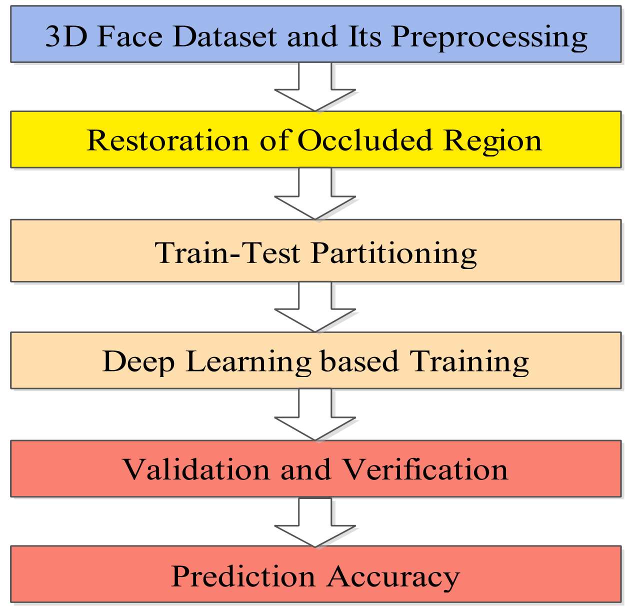


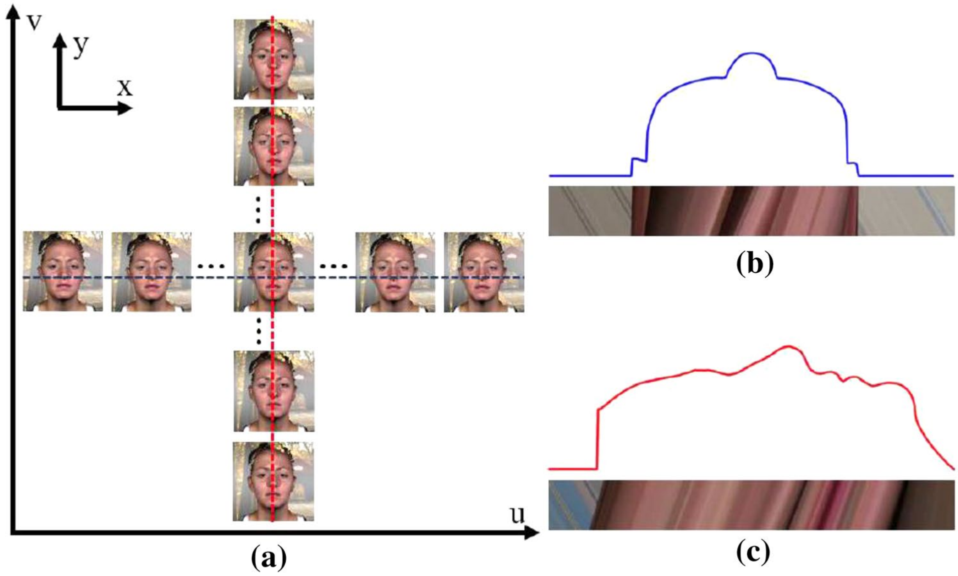
Fig. 8 Phases of 3D face recognition using restoration [9]

Li et al. [45] proposed SymmFCNet, a symmetry consistent convolutional neural network for reconstructing missing pixels on the one-half face using the other half. SymmFCNet consisted of illumination-reweighted warping and generative reconstruction subnet. The dependency on multiple networks is a significant drawback. Han et al. [46] proposed a sketching system that creates 3D caricature photos by modifying the facial features. An unconventional deep learning method was designed to get the vertex wise exaggeration map. They used the FaceWarehouse dataset [20] for training and testing. The advantage was the conversion of a 2D image into a 3D face caricature model. However, the caricatured quality is affected in the presence of eyeglasses. Besides this, the reconstruction is affected by the varying light conditions. Moschoglou et al. [47] implemented an autoencoder such as 3DFaceGAN for modelling 3D facial surface distribution. Reconstruction loss and adversarial loss were used for generator and discriminator. On the downside, GANs are hard to train and cannot be applied to real-time 3D face solutions.

2.3 Epipolar Geometry based reconstruction

The epipolar geometry-based facial reconstruction approach uses various non-synthesising perspective images of one subject to generate a single 3D image [48]. Good geometric fidelity is the main advantage of these techniques. The calibrated camera and the orthogonal images are two main challenges associated with these techniques. Figure 9 shows the horizontal and vertical

Fig. 9 **a** Epipolar Plane Images corresponding to 3D face curves, **b** horizontal EPI, and **c** vertical EPI [48]



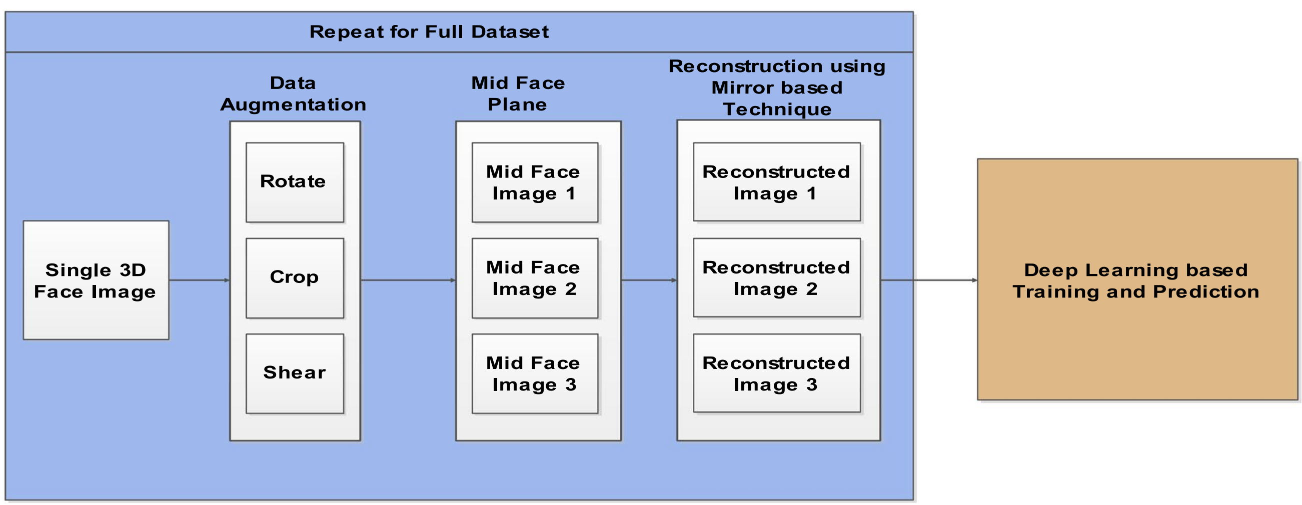


Fig. 10 General framework of one-shot learning-based 3D face reconstruction

Epipolar Plane Images (EPIs) obtained from the central view and sub-aperture images [48].

Anbarjafari et al. [49] proposed a novel technique for generating 3D faces captured by phone cameras. A total of 68 facial landmarks were used to divide the face into four regions. Different phases were used during texture creation, weighted region creation, model morphing, and composing. The main advantage of this technique is the good generalisation obtained from the feature points. However, it is dependent on the dataset having good head shapes, which affects the overall quality.

2.4 One-Shot Learning-based Reconstruction

The one-shot learning-based reconstruction method uses a single image of an individual to recreate a 3D recognition model [50]. This technique utilises a single image per subject to train the model. Therefore, these techniques are quicker to train and also generates promising results [51]. However, this approach cannot be generalised to videos. Nowadays, One-shot learning-based 3D reconstruction is an active research area.

The ground truth 3D models are needed to train the model for mapping from the 2D-to-3D image. Some researchers used depth prediction for reconstructing 3D structures [52, 53]. While other techniques directly predict 3D shapes [54, 55]. Few works have been done on 3D face reconstruction by utilising one 2D image [38, 39]. The optimum parameter values for the 3D face are obtained by using deep neural networks and model parameter vectors. The major enhancement has been achieved over [56, 57]. However, this approach fails to handle pose variation adequately. The major drawbacks of this technique are the creation of multi-view 3D faces and reconstruction degradation. Figure 10 shows the general framework of the one shot-based face reconstruction technique.

Xing et al. [58] presented a 3D face reconstruction technique using a single image without considering the ground-truth 3D shape. The face model rendering was used in the reconstruction process. The fine-tuning bootstrap method was used to send feedback for further improvement in the quality of rendering. This technique provides the reconstruction of 3D shape from 2D image. However, the con is that rigid-body transformation is used for pre-processing.

2.5 Shape from shading based reconstruction

Shape from shading (SFS) method is based on the recovery of 3D shape from shading and lighting cues [59, 60]. It uses an image that produces a good shape model. However, the occlusion cannot be dealt with when shape estimates have interfered with a target's shadow. It operates well under lightening from the non-frontal face view (see Fig. 11). Jiang et al. [61] method was inspired by the face animation using RGB-D and monocular video. The computation of coarse estimation was done for the target 3D face by using a parametric model fitting to the input image. The reconstruction of a 3D image from a single 2D image is the main drawback of this technique. On the contrary, the SFS technique depends upon pre-defined knowledge about facial geometry, such as facial symmetry.

2.6 Hybrid Learning-based Reconstruction

Richardson et al. [38] proposed a technique for generating the database with photo-realistic face images using geometries. ResNet model [62] was used for building the proposed network. This technique was unable to restore the images having different facial attributes. It failed to generalise the training process for new face generations. Liu et al. [63] proposed a 3D face reconstruction technique using the mixture of 3DMM and shape-from-shading method. Mean absolute error (MAE) was plotted for the convergence of reconstruction error. Richardson et al. [39] proposed a single shot learning model for extracting the coarse-to-fine facial shape. CoarseNet and FineNet were used for the recovery of coarse facial features. The high detail face reconstruction includes wrinkles from a single image. However, it fails to generalise the facial features that are available in the training data. The dependence on synthetic data is another drawback. Jackson et al. [51] proposed a CNN-based model for reconstructing 3D face geometry using a single 2D facial image. This method did not require any kind of facial alignment. It works on various types of expressions and poses.

Tewari et al. [64] proposed a generative model based on a convolutional autoencoder network for face reconstruction. They used AlexNet [65] and VGGFace [66] models. However, it fails under occlusion such as beard or external object. Dou et al. [67] proposed a deep neural network (DNN) based technique for end-to-end 3D face reconstruction using a single 2D image. Multitask loss function and fusion CNN were hybridised for face recognition. The main advantage of this method is the simplified framework with the end-to-end model. However, the proposed approach suffers from the dependency of synthetic data. Han et al. [68] proposed a sketching system for 3D faces and caricatured

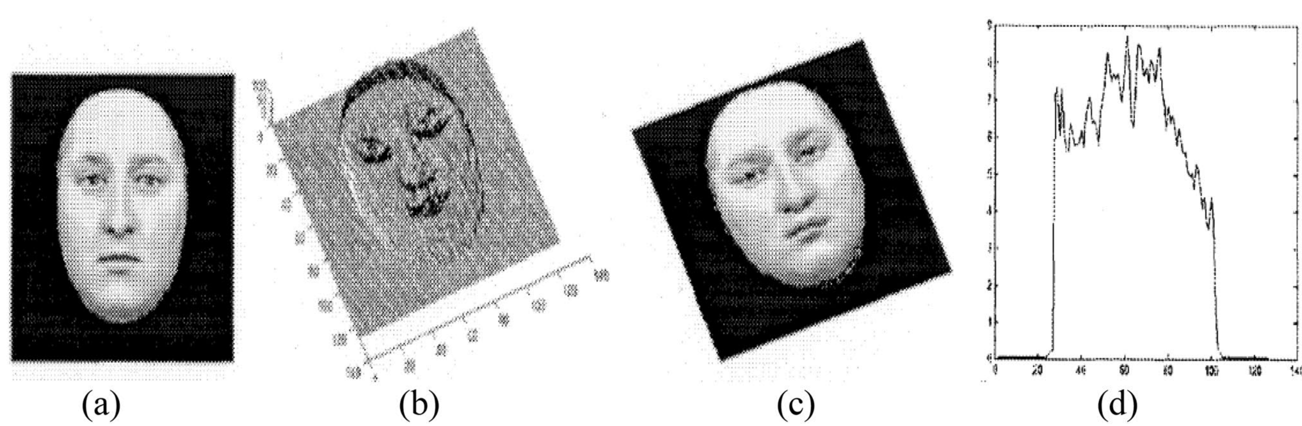


Fig. 11 3D face shape recovery **a** 2D image, **b** 3D depth image, **c** Texture projection, and **d** Albedo histogram[59]

modelling using CNN-based deep learning. Generally, the rich facial expressions were generated through MAYA and ZBrush. However, it includes the gesture-based interaction with the user. The shape level input was appended with a fully connected layer's output to generate the bilinear output.

Hsu et al. [69] proposed two different approaches for cross-pose face recognition. One technique is based on 3D reconstruction, and the other method is built using deep CNN. The components of the face were built out of the gallery of 2D faces. The 3D surface was reconstructed using 2D face components. CNN based model can easily handle in-the-wild characteristics. 3D component-based approach does not generalise well. Feng et al. [48] developed a FaceLFnet to restore 3D face using Epipolar Plane Images (EPI). They used CNN for the recovery of vertical and horizontal 3D face curves. The photo-realistic light field images were synthesised using 3D faces. A total of 14 K facial scans of 80 different people were used during the training process, making up to 11 million facial curves/EPIs. The model is a superior choice for medical applications. However, this technique requires a huge amount of epipolar plane image curves.

Zhang et al. [70] proposed a 3D face reconstruction technique using the combination of morphable faces and sparse photometric stereo. An optimisation technique was used for per-pixel lighting direction along with the illumination at high precision. The semantic segmentation was performed on input images and geometry proxy to reconstruct details such as wrinkles, eyebrows, whelks, and pores. The average geometric error was used for verifying the quality of reconstruction. This technique is dependent on light falling on the face. Tran et al. [71] proposed a bump map based 3D face reconstruction technique. The convolutional encoder-decoder method was used for estimating the bump maps. Max-pooling and rectified linear unit (ReLU) were used along with the convolutional layers. The main disadvantage of the technique is that the unoptimised soft symmetry implementation is slow. Feng et al. [72] presented the benchmark dataset consisting of 2 K faces for 135 people. Five different 3D face reconstruction approaches were evaluated on the proposed dataset.

Feng et al. [73] proposed a 3D face reconstruction technique called a Position map Regression Network (PRN) based on the texture coordinates UV position maps. CNN regressed 3D shape from a one-shot 2D image. The weighted loss function used different weights in the form of a weight mask during the convolution process. The UV position map can generalise as well. However, it is difficult to be applied in real-world scenarios. Liu et al. [74] proposed an encoder-decoder based network for regressing 3D face shape from 2D images. The joint loss was computed based on 3D face reconstruction as well as identification error. However, the joint loss function affects the quality of face shapes. Chinaev et al. [75] developed a CNN based model for 3D face reconstruction using mobile devices. MobileFace CNN was used for the testing phase. This method was fast training on mobile devices and real-time application. However, the annotation of 3D faces using a morphable model is costly at the pre-processing stage. Gecer et al. [27] proposed a 3D face reconstruction based on DCNNs and GANs. In UV space, GAN was used to train the generator for facial textures. An unconventional 3DMM fitting strategy was formulated on differentiable renderer and GAN. Deng et al. [76] presented a CNN based single-shot face reconstruction method for weakly supervised learning. The perception level and image-level losses were combined. The pros of this technique are large pose and occlusion invariant. However, the confidence of the model is low on occlusion during the prediction phase.

Yuan et al. [77] proposed a 3D face restoration technique for occluded faces using 3DMM and GAN. Local discriminator and global discriminator were used for verifying the quality of 3D face. The semantic mapping of facial landmarks led to the generation of synthetic faces under occlusion. Contrasting to that, multiple discriminators increase the time complexity. Luo et al. [78] implemented a Siamese CNN method for 3D face restoration. They utilised the weighted parameter distance cost (WPDC) and contrastive cost function to validate the quality of the reconstruction method. However, the face recognition has not been tested in the wild, and the number of training images are low. Gecer et al. [79] proposed a GAN based method for synthesising the high-quality 3D faces. Conditional GAN was used for expression augmentation. 10 K new individual identities were randomly synthesised from 300 W-LP dataset. This technique generates high-quality 3D faces with fine details. However, GANs are hard to train and yet cannot be applied in real-time solutions. Chen et al. [80] proposed a 3D face reconstruction technique using a self-supervised 3DMM trainable VGG encoder. A two-stage framework was used to regress 3DMM parameters for reconstructing the facial details. Faces are generated with good quality under normal occlusion. The details of the face are captured using UV space. However, the model fails on extreme occlusion, expression, and large pose. CelebA [81] dataset was used for training, and LFW [82] dataset was used along with CelebA for the testing process. Ren et al. [83] developed an encoder-decoder framework for video deblurring of 3D face points. The identity knowledge and facial structure were predicted by the rendering branch and 3D face reconstruction. Face deblurring is done over the video handling challenge of pose variation. High computational cost is the major drawback of this technique.

Tu et al. [10] developed a 2D assisted self-supervised learning (2DASL) technique for 2D face images. The noisy information of landmarks was used to improve the quality of 3D face models. Self-critic learning was developed for improving the 3D face model. The two datasets, namely AFLW-LFPA [84] and AFLW2000-3D [85], were used for 3D face restoration and face alignment. This method works for in-the-wild 2D faces along with noisy landmarks. However, it has a dependency on 2D-to-3D landmarks annotation. Liu et al. [86] proposed an automatic method for generating pose-and-expression-normalised (PEN) 3D faces. The advantages of this technique are the reconstruction from a single 2D image and the 3D face recognition invariant of pose and expression. However, it is not occlusion invariant. Lin et al. [24] implemented a 3D face reconstruction technique based on single-shot image in-the-wild. Graph convolutional networks were used for generating the high-density facial texture. FaceWarehouse [20] along with CelebA [81] database were used for training purposes. Ye et al. [87] presented a big dataset of 3D caricatures. They generated a PCA based linear 3D morphable model for caricature shapes. 6.1 K portrait caricature images were collected from pinterest.com as well as WebCaricature dataset [88]. High-quality 3D caricatures have been synthesised. However, the quality of the caricature is not good for occluded input face images. Lattas et al. [89] proposed a technique for producing high-quality 3D face reconstruction using arbitrary images. The large scale database was collected using 200 different subjects based on their geometry and reflectance. The image translation networks were trained to estimate specular and diffuse albedo. This technique generated high-resolution avatars using GANs. However, it fails to generate avatars of dark skin subjects.

Zhang et al. [90] proposed an automatic landmark detection and 3D face restoration for caricatures. 2D image of caricature was used for regressing the orientation and shape of 3D caricature. ResNet model was used for encoding the input image to a latent space. The decoder was used along with the fully connected layer to generate 3D landmarks on the caricature. Deng et al. [91] presented DISentangled precisely-COntrollable (DiscoFaceGAN) latent embedding for representing fake people with various poses, expressions, and illumination. Contrastive learning was employed to promote disentanglement by comparing rendered faces with the real ones. The face generation is precise over expressions, poses, and illumination. The low quality of the model is generated under low lighting and extreme poses. Li et al. [92] proposed a 3D face reconstruction technique to estimate the pose of a 3D face using coarse-to-fine estimation. They used an adaptive reweighting method to generate the 3D model. The pro of this technique was the robustness to partial occlusions and extreme poses. However, the model fails when 2D and 3D landmarks are wrongly estimated for occlusion. Chaudhuri et al. [93] proposed a deep learning method to train personalised dynamic albedo maps and the expression blendshapes. 3D face restoration was generated in a photo-realistic manner. The face parsing loss and blendshape gradient loss captured the semantic meaning of reconstructed blend shapes. This technique was trained in-the-wild videos, and it generated high-quality 3D face and facial motion transfer from one person to other. It did not work well under external occlusion. Shang et al. [94] proposed a self-supervised learning technique for occlusion aware view synthesis. Three different loss functions, namely, depth consistency loss, pixel consistency loss, and landmark-based epipolar loss, were used for multi-dimensional consistency. The reconstruction is done through the occlusion-aware method. It does not work well under external occlusion such as hands, glasses, etc.

Cai et al. [95] proposed an Attention Guided GAN (AGGAN), which is capable of 3D facial reconstruction using 2.5D images. AGGAN generated a 3D voxel image from the depth image using the autoencoder technique. 2.5D to 3D face mapping was done using attention-based GAN. This technique handles a wide range of head poses and expressions. However, it is unable to fully reconstruct the facial expression in case of a big open mouth. Xu et al. [96] proposed training the head geometry model without using 3D ground-truth data. The deep synthetic image with head geometry was trained using CNN without optimisation. The head pose manipulation was done using GANs and 3D warping. Table 1 presents the comparative analysis of 3D facial reconstruction techniques. Table 2 summarises the pros and cons of 3D face reconstruction techniques.

3 Performance Evaluation Measures

Evaluation measures are important to know the quality of the trained model. There are various evaluation metrics, namely, mean absolute error (MAE), mean squared error (MSE), normalised mean error (NME), root mean squared error (RMSE), cross-entropy loss (CE), area under the curve (AUC), intersection over union (IoU), peak signal to noise ratio (PSNR), receiver operator characteristic (ROC), and structural similarity index (SSIM). Table 3 presents the evaluation of 3D face reconstruction techniques in terms of performance measures. During the face reconstruction, the most important performance measures are MAE, MSE, NME, RMSE, and adversarial loss. These are the five widely used performance measures. Adversarial loss is being used since 2019 with the advent of GANs in 3D images.

4 Datasets Used for Face Recognition

Table 4 depicts the detailed description of datasets used in 3D face reconstruction techniques. The analysis of different datasets highlights the fact that most of the 3D face datasets are publicly available datasets. They do not have a high

Table 1 Comparative analysis of 3D facial reconstruction techniques

Reference	Year	Approaches/Models used	Is Face Alignment Done?	Convergence factor	Is Deep Learning Done?	Synthetic Data used?
[38]	2016	SFS with landmarks and deep learning	No	MSE	Yes	Yes
[63]	2017	3DMM and SFS, learning cascaded regression	Yes	Mean Absolute Error (MAE)	No	Yes
[39]	2017	CNN and Coarse-to-fine details	No	MSE	Yes	Yes
[51]	2017	Volumetric regression networks (Multi-task and Guided)	Yes	Normalized Mean Error (NME)	Yes	No
[64]	2017	Auto encoder-based CNN	Yes	Geometric, Photometric, and Landmark error	Yes	No
[67]	2017	DNN architecture	No	Root Mean Square Error (RMSE)	Yes	No
[68]	2017	CNN based deep regression network	No	Mean error	Yes	Yes
[29]	2017	Automatic reconstruction of a human face and 3D epipolar geometry	Yes	Mean and Standard Deviation	No	No
[26]	2018	3D deep feature vector and 3D augmentation of faces	Yes	Cumulative Matching Characteristic (CMC) and Receiver Operator Characteristic (ROC) curve	Yes	Yes
[61]	2018	Coarse face modeling, Medium face modeling, and Fine face modeling	Yes	RMSE	No	No
[30]	2018	Clustering and interpolation-based reconstruction	No	Error distribution	No	No
[69]	2018	Faster Region-based CNN (RCNN) and reduced tree structure model	No	Moving Least Squares (MLS)	Yes	Yes
[43]	2018	FR3DNet, CNN, Data augmentation. Main objective of the paper is to close the gap between size of 2D and 3D datasets for effective 3DFR	Yes	ROC curve	Yes	Yes
[48]	2018	FaceLFNet. 3DMM with facial geometry. EPIs and CNN	No	RMSE	Yes	No
[70]	2018	Leverages sparse photometric stereo (PS)	No	Average geometric error for reconstruction	No	Yes
[71]	2018	Deep convolution encoder-decoder	No	ROC curve	No	No
[72]	2018	3D dense face reconstruction algorithm. 3DMM-CNN	No	RMSE	Yes	No
[73]	2018	UV position maps	Yes	NME	Yes	No
[74]	2018	Encoder-decoder network	No	RMSE	Yes	No
[31]	2018	PIFR based 3DMM	Yes	Mean euclidean metric (MEM)	No	No
[32]	2019	3DMM. Cascaded regression	No	RMSE and MAE	No	No
[49]	2019	Blended model	Yes	Structural Similarity Index Metric (SSIM)	No	No
[75]	2019	MobileNet CNN	No	Area Under the Curve (AUC)	Yes	No
[58]	2019	Deep learning model	Yes	NME	Yes	Yes
[27]	2019	3DMM fitting based on GANs and a differentiable renderer	No	Mean and Standard Deviation	Yes	No
[76]	2019	CNNs	Yes	RMSE	Yes	No
[77]	2019	Inverse 3DMM GAN model	Yes	Peak Signal to Noise Ratio (PSNR) and SSIM	Yes	Yes
[78]	2019	Siamese network-based CNN model	Yes	ROC curve	Yes	No
[79]	2019	GAN	No	Wasserstein GAN (W-GAN) adversarial loss	Yes	Yes
[80]	2019	Self-supervised 3DMM encoder	Yes	RMSE	No	No
[83]	2019	Encoder-decoder framework	Yes	PSNR and SSIM	Yes	Yes
[44]	2019	Neural rendering	No	RMSE	Yes	Yes
[33]	2020	Blended model	No	Pearson correlation and MSE	Yes	Yes

Table 1 (continued)

Reference	Year	Approaches/Models used	Is Face Alignment Done?	Convergence factor	Is Deep Learning Done?	Synthetic Data used?
[45]	2020	SymmFCNet	Yes	PSNR, SSIM, Identity Distance, and Perceptual Similarity	Yes	No
[46]	2020	Deep learning-based technique	Yes	MSE	Yes	Yes
[10]	2020	2D-Assisted Self Supervised Learning (2DASL)	Yes	NME	Yes	Yes
[47]	2020	VAE-GAN	Yes	Normalized dense vector error	Yes	No
[86]	2020	Summation model and cascaded regression	Yes	MAE and NME	No	No
[24]	2020	Graph convolutional networks	No	W-GAN adversarial loss	Yes	No
[87]	2020	PCA model	No	Adversarial loss, Bidirectional cycle-consistency loss, Cross-domain character loss, and User control loss	Yes	Yes
[89]	2020	Generation of per-pixel diffuse and specular components	No	PSNR	Yes	Yes
[90]	2020	Parametric model based on vertex deformation space	Yes	Cumulative error distribution	Yes	Yes
[91]	2020	DiscoFaceGAN	Yes	Adversarial loss, Imitative loss, and Contrastive loss	Yes	Yes
[92]	2020	Joint 2D and 3D metaheuristic method	Yes	3D Root Mean Square Error (3DRMSE)	No	No
[93]	2020	End-to-end deep learning framework	Yes	NME and AUC	Yes	No
[94]	2020	MGCNet	Yes	RMSE	No	No
[35]	2020	3D morphable model-based Pixel-3DM	Yes	NME and RMSE	No	Yes
[95]	2020	Attention Guided Generative Adversarial Networks (AGGAN)	Yes	Intersection-over-Union (IoU) and Cross-Entropy Loss (CE)	Yes	No
[96]	2020	GAN	Yes	Adversarial loss	Yes	Yes
[2]	2020	Variational autoencoder, bi-LSTM, and triplet loss training	No	MAE	Yes	No
[8]	2020	Deep learning process, game-theory based generator and discriminator	No	MAE	Yes	No
[9]	2021	Variational autoencoders and triplet loss training	No	MAE	Yes	No
[97]	2021	Single shot learning based weakly supervised multi-face reconstruction technique	Yes	NME	Yes	No

number of images to train the model compared to 2D face publicly datasets. This makes the research in 3D faces more interesting because the scalability factor has not been tested and has become an active area of research. It is worth mentioning that only three datasets, namely, Bosphorus, KinectFaceDB, and UMBDB datasets, have occluded images for occlusion removal.

5 Tool and Techniques in 3D Face Reconstruction Techniques

Table 5 presents the techniques along with hardware used in terms of a graphical processing unit (GPU), size of random-access memory (RAM), central processing unit (CPU), and brief applications. The comparison highlights the importance of deep learning in 3D face reconstruction. GPUs play a vital role in the deep learning-based model. With the advent of Google Collaboratory, GPUs are freely accessible.

6 Applications

Based on the artificial intelligence-based AI + X technique [128], where X is domain expertise in face recognition, a plethora of applications are affected by 3D face reconstruction. Facial puppetry, speech-driven animation and enactment, video dubbing, virtual makeup, projection mapping, face replacement, face aging, and 3D printing in medicine

Table 2 Comparison of 3D face reconstruction techniques in terms of pros and cons

References	Year	Pros	Cons
[38]	2016	Efficient on various illumination conditions and expressions Generalises well on the synthetic data	Fail to reconstruct new faces in the test set Fails on facial attributes not present in synthetic data
[63]	2017	No annotation needed for invisible landmarks	The facial texture quality is low Unable to apply in real-world scenarios
[39]	2017	Fine details such as wrinkles can be reconstructed	Fails to generalise on facial features not in training data Dependency on synthetic data
[51]	2017	Reconstruction is done using a single 2D image	Lack of facial alignment leads to the generation of almost identical faces after reconstruction
[64]	2017	End-to-end deep learning	Fail under occlusion such as beard or external object Shrinked reconstruction
[67]	2017	Simplified framework with end-to-end model instead of iterative model	Dependency on synthetic data
[68]	2017	Free hand sketch of the face gets converted to 3D caricature model	Generates unnatural results when exaggeration in expression and shape are inconsistent
[29]	2017	Model training can be done on mobile A general variational segmentation model is proposed to generalise the glasses	Occlusion-invariant only on glasses
[26]	2018	Transfer Learning-based model is faster to train	Loses 3D information while converting 3D point cloud image to 2.5D
[61]	2018	3D face reconstructed from single 2D image	SFS technique depends on pre-defined knowledge about facial geometry, such as facial symmetry
[30]	2018	100 K point clouds are generated with a single shot of RGB-D sensor Low-cost method	External hardware is required
[69]	2018	3D component-based approach requires only a few 3D models CNN based models can easily handle in-the-wild characteristics	3D component-based approach does not generalise well CNN based approach requires a large dataset for training
[43]	2018	Deep CNN model trained on 3.1 M 3D faces of 100 K subjects	Training a CNN from scratch is time-consuming
[48]	2018	Method generalises well This model-free approach is a superior choice in medical applications	A huge amount of epipolar plane image curves are required
[70]	2018	High-quality 3D faces are generated with fine details	Dependent on light falling on the face
[71]	2018	Bump map regression takes 0.03 s/image Face segmentation requires 0.02 s/image	Unoptimised soft symmetry implementation takes 50 s/image
[72]	2018	3D face reconstruction from single 2D image	Texture based reconstruction technique takes high computational time
[73]	2018	Size of the model is 160 MB in contrast to 1.5 GB of Volumetric Regression Network The UV position map can generalise well	Unable to apply in real-world scenarios
[74]	2018	It is a lightweight network	The joint loss function affects the quality of face shapes
[31]	2018	Good reconstruction invariant of poses	The reconstruction needs improvement for large poses
[32]	2019	The reconstruction technique is robust to light variation	Less landmarks are available. Hence landmark displacement features can be improved
[49]	2019	Good generalisation through UV map based on feature points	Dependent on database with good head shapes Due to this the overall head shape lacks good quality
[75]	2019	Fast training on mobile devices Real-time application	Annotation of 3D faces using morphable model is costly at a pre-processing stage of proposed MobileNet
[58]	2019	Rendering of input image to multiple view angles 2D image to 3D shape is reconstructed	Rigid-body transformation is used for pre-processing

Table 2 (continued)

References	Year	Pros	Cons
[27]	2019	High texture 3D images are generated using UV maps with GANs	GANs are hard to train It cannot be applied in real-time solutions
[76]	2019	Large pose and occlusion invariant Weakly supervised learning	Confidence of model is low on occlusion during prediction
[77]	2019	Synthetic faces under occlusion images generated with semantic mapping to facial landmarks	Multiple discriminators add on to the model complexity GANs are hard to train
[78]	2019	Siamese CNN based reconstruction has achieved at-par normalised mean error when compared to 3D dense face alignment method	Face recognition has not been tested in-the-wild The number of images in training set are low
[79]	2019	High-quality 3D faces are generated with fine details	GANs are hard to train Cannot be applied in real-time solutions
[80]	2019	Faces are generated with good quality under normal occlusion Details of face are captured using UV space	Model fails on extreme occlusion, expression, and large pose
[83]	2019	Face deblurring is done over video handling challenge of pose variation	High computational cost
[44]	2019	Technique is fast to train as it depends on transfer learning and not training from scratch	Video quality needs to be improved for real-world applications
[33]	2020	Works for in-the-wild face datasets	4DFAB dataset is not publicly available
[45]	2020	Missing pixels are regenerated using a generative model	Dependency on multiple networks
[46]	2020	2D image as input can be converted to caricature of 3D face model	Caricature quality is affected when occlusion such as eyeglasses exist Not invariant to lighting conditions
[10]	2020	Works for in-the-wild 2D faces along with the noisy landmarks. Self-supervised learning	Dependency on 2D-to-3D landmarks annotation
[47]	2020	3D GAN method for generation and translation of 3D face High-frequency details are preserved	GANs are hard to train Cannot be applied in real-time 3D face solutions
[86]	2020	3D face reconstruction from single 2D image 3D face recognition invariant of pose and expression	Does not include occlusion invariance
[24]	2020	No large-scale dataset is required Detailed and coloured 3D mesh image	Does not work for occluded face regions
[87]	2020	End-to-end deep learning method 6.1 K 3D meshes of caricature are synthesised Generates high-quality caricatures	Does not work well if the input image is occluded
[89]	2020	Generates high-resolution avatars using GANs	Fails to generate avatars of dark skin subjects Minor alignment errors lead to blur of pore details
[90]	2020	3D caricature shape is directly built from 2D caricature image	Does not work well if the input image is occluded
[91]	2020	Face generation is precise over expressions, poses, and illumination	Low quality of model is generated under low lighting and extreme poses
[92]	2020	Robust to partial occlusions and extreme poses	Fails when 2D and 3D landmarks are wrongly estimated for occlusion
[93]	2020	Trained through in-the-wild videos Generates high-quality 3D face and facial motion transfer from one person to other	Does not work well under external occlusion
[94]	2020	Reconstruction is done through an occlusion-aware method	Does not work well under external occlusion such as glasses, hands, etc
[35]	2020	Proposed technique can do 3D face analysis and generation effectively	External occlusions are not considered

Table 2 (continued)

References	Year	Pros	Cons
[95]	2020	2.5D to 3D face mapping is done with attention-based GAN Handles a wide range of head poses and expressions	Unable to fully reconstruct the facial expression in case of big open mouth
[96]	2020	The CNN-based model learns head-geometries even without ground-truth data	Unable to handle large pose variations
[2]	2020	The mirroring technique is faster for reconstruction as compared to deep learning-based methods The pre-processing time, reconstruction time, and verification time are faster in computation as compared to the state-of-the-art methods	Preprocessing does not include facial alignment The technique has not been tested on a big dataset of 3D faces
[8]	2020	End-to-end deep learning technique Occlusion invariant reconstruction is done	Not tested for being scalable
[9]	2021	One-shot learning-based 3D face restoration technique Landmarks based reconstruction is faster as compared to mesh-based reconstruction using the mirroring technique	Facial alignment is missing The maximum size of the dataset is 4666 3D images. Deep learning model can be trained better if the dataset is huge
[97]	2021	A single network model for multi-face reconstruction Faster pre-processing for feature extraction	Low facial texture Multiple GPUs required

are some of the well-known applications. These are discussed in succeeding subsections.

6.1 Facial Puppetry

The games and movie industry use facial cloning or puppetry in video-based facial animation. The expressions and emotions are transferred from user to target character through video streaming. When artists dub for animated characters for a movie, 3D face reconstruction can help the expressions transfer from the artist to the character. Figure 12 illustrates the puppetry demonstration in real-time by digital avatars [129, 130].

6.2 Speech-driven Animation and Reenactment

Zollhofer et al. [1] discussed various video-based face reenactment works. Most of the methods depend on the reconstruction of source and target face using a parametric facial model. Figure 13 presents the pipeline architecture of neural voice puppetry [44]. The audio input is passed through deep speech based on a recurrent neural network for feature extraction. Furthermore, the autoencoder-based expression features with the 3D model are transferred to the neural renderer to receive the speech-driven animation.

6.3 Video Dubbing

Dubbing is an important part of filmmaking where an audio track is added or replaced in the original scene. The original actor's voice is to be replaced with the dubbed actor. This process requires ample training for dubbed actors to match their audio with the original actor's lip-sync [131]. To minimise the discrepancies in visual dubbing, the reconstruction of mouth in motion complements the dialogues spoken by the dubbed actor. It involves the mapping of dubber's mouth movements with the actor's mouth [132]. Hence, the technique of image swapping or transferring parameters is used. Figure 14 presents the visual dubbing by VDub [131] and Face2Face with live enabled dubbing [132]. Figure 14 shows DeepFake example in 6.S191 [133], showing the course instructor dubbing his voice to famous personalities using deep learning.

6.4 Virtual Makeup

Virtual makeup is excessively used in online platforms for meetings and video chats where a presentable appearance is indispensable. It includes digital image changes such as applying suitable colour lipstick, face masks, etc. It can be useful for beauty product companies as they can advertise digitally, and consumers can experience the real-time effect

Table 3 Evaluation of 3D face restoration techniques in terms of performance measures

References	Year	Adversarial Loss	AUC	CE	IoU	MAE	MEM	MSE	NME	PSNR	RMSE	ROC	SSIM	Other
[38]	2016	×	×	×	×	×	×	✓	×	×	×	×	×	×
[63]	2017	×	×	×	×	✓	×	×	×	×	×	×	×	×
[39]	2017	×	×	×	×	×	×	✓	×	×	×	×	×	×
[51]	2017	×	×	×	×	×	×	×	✓	×	×	×	×	×
[64]	2017	×	×	×	×	×	×	×	×	×	×	×	×	✓
[67]	2017	×	×	×	×	×	×	×	×	×	✓	×	×	×
[68]	2017	×	×	×	×	×	×	×	×	×	×	×	×	✓
[29]	2017	×	×	×	×	×	×	×	×	×	×	×	×	✓
[26]	2018	×	×	×	×	×	×	×	×	×	×	✓	×	✓
[61]	2018	×	×	×	×	×	×	×	×	×	✓	×	×	×
[30]	2018	×	×	×	×	×	×	×	×	×	×	×	×	✓
[69]	2018	×	×	×	×	×	×	×	×	×	×	×	×	✓
[43]	2018	×	×	×	×	×	×	×	×	×	×	✓	×	×
[48]	2018	×	×	×	×	×	×	×	×	×	✓	×	×	×
[70]	2018	×	×	×	×	×	×	×	×	×	×	×	×	✓
[71]	2018	×	×	×	×	×	×	×	×	×	×	✓	×	×
[72]	2018	×	×	×	×	×	×	×	×	×	✓	×	×	×
[73]	2018	×	×	×	×	×	×	×	✓	×	×	×	×	×
[74]	2018	×	×	×	×	×	×	×	×	×	✓	×	×	×
[31]	2018	×	×	×	×	×	✓	×	×	×	×	×	×	×
[32]	2019	×	×	×	×	✓	×	×	×	×	✓	×	×	×
[49]	2019	×	×	×	×	×	×	×	×	×	×	×	✓	×
[75]	2019	×	✓	×	×	×	×	×	×	×	×	×	×	×
[58]	2019	×	×	×	×	×	×	×	✓	×	×	×	×	×
[27]	2019	×	×	×	×	×	×	×	×	×	×	×	×	✓
[76]	2019	×	×	×	×	×	×	×	×	×	✓	×	×	×
[77]	2019	×	×	×	×	×	×	×	×	✓	×	×	✓	×
[78]	2019	×	×	×	×	×	×	×	×	×	×	✓	×	×
[79]	2019	✓	×	×	×	×	×	×	×	×	×	×	×	×
[80]	2019	×	×	×	×	×	×	×	×	×	✓	×	×	×
[83]	2019	×	×	×	×	×	×	×	×	✓	×	×	✓	×
[44]	2019	×	×	×	×	✓	×	×	×	×	×	×	×	✓
[33]	2020	×	×	×	×	×	×	✓	×	×	×	×	×	×
[45]	2020	×	×	×	×	×	×	×	×	✓	×	×	✓	×
[46]	2020	×	×	×	×	×	×	✓	×	×	×	×	×	×
[10]	2020	×	×	×	×	×	×	×	✓	×	×	×	×	×
[47]	2020	×	✓	×	×	×	×	×	×	×	×	×	×	✓
[86]	2020	×	×	×	×	✓	×	×	✓	×	×	×	×	×
[24]	2020	✓	×	×	×	×	×	×	×	×	×	×	×	×
[87]	2020	✓	×	×	×	×	×	×	×	×	×	×	×	×
[89]	2020	×	×	×	×	×	×	×	×	✓	×	×	×	×
[90]	2020	×	×	×	×	×	×	×	✓	×	×	×	×	×
[91]	2020	✓	×	×	×	×	×	×	×	×	×	×	×	×
[92]	2020	×	×	×	×	×	×	×	×	×	✓	×	×	×
[93]	2020	×	✓	×	×	×	×	×	✓	×	×	×	×	×
[94]	2020	×	×	×	×	×	×	×	×	×	✓	×	×	×
[35]	2020	×	×	×	×	×	×	×	✓	×	✓	×	×	×
[95]	2020	×	×	✓	✓	×	×	×	×	×	×	×	×	×
[96]	2020	✓	×	×	×	×	×	×	×	×	×	×	×	×

Table 3 (continued)

References	Year	Adversarial Loss	AUC	CE	IoU	MAE	MEM	MSE	NME	PSNR	RMSE	ROC	SSIM	Other
[2]	2020	×	×	×	×	✓	×	×	×	×	×	×	×	✓
[8]	2020	✓	×	×	×	✓	×	×	×	×	×	×	×	×
[9]	2021	×	×	×	×	✓	×	×	×	×	×	×	×	✓
[97]	2021	×	×	×	×	×	×	×	✓	×	×	×	×	×

of their products on their images. It is implemented by using different reconstruction algorithms.

The synthesised virtual tattoos have been shown adjusting to the facial expression [134] (see Fig. 15a). Viswanathan et al. [135] gave a system in which two face images are given as input, one with eyes-open and the other with eyes-closed. An augmented reality-based face is proposed to add one or more makeup shapes, layers, colours, and textures to the face. Nam et al. [136] proposed an augmented reality-based lip makeup method, which used pixel-unit makeup compared to polygon unit makeup on lips, as seen in Fig. 15b.

6.5 Projection-Mapping

Projection mapping uses projectors to amend the features or expressions of real-world images. This technique is used to bring life to static images and give them a visual display. Different methods are used for projection mapping in 2D and 3D images to alter the person's appearance. Figure 16 presents the live projection mapping system called FaceForge [137].

Lin et al. [24] presented a technique of 3D face projection to the input image by passing the input image through CNN and combining the information with 3DMM to get the fine texture of the face (see Fig. 17).

6.6 Face Replacement

Face Replacement is commonly used in the entertainment industry, where the source face is replaced with the target face. This technique is based on parameters such as the track of identity, facial properties, and expressions of both faces (source and target). The source face is to be rendered so that it matches the conditions of the target face. Adobe After Effects is a famous tool used in the movie and animation industry and can help face replacement [138] (see Fig. 18).

6.7 Face Aging

Face ageing is an effective technique to convert 3D face images into 4D. If a single 3D image can be synthesised using aging GAN, it would be useful to create 4D datasets. Face aging is also called age progression or age synthesis as it revives the face by changing the facial features. Various techniques are used to enhance the features of the face so that the original image is preserved. Figure 19 shows the transformation of the face using age-conditional GAN (ACGAN) [139].

Shi et al. [140] used GANs for face aging because different face parts have different ageing speeds over time. Hence, they used attention based conditional GAN using normalisation for handling the segmented face aging. Fang et al. [141] proposed a progressive face aging using the triple loss function at the generator level of GAN. The complex translation loss helped them in handling face age effectively. Huang et al. [142] worked on face aging using the progressive GAN for handling three aspects such as identity preservation, high fidelity, and aging accuracy. Liu et al. [143] proposed a controllable GAN for manipulating the latent space of the input face image to control the face aging. Yadav et al. [144] proposed face recognition over various age gap using two different images of the same person. Sharma et al. [145] worked on fusion-based GAN using a pipeline of CycleGAN for aging progression and enhanced super-resolution GAN for high fidelity. Liu et al. [146] proposed a face aging method for young faces modeling the transformation over appearance and geometry of the face.

As shown in Table 6, facial reconstruction can be used in three different types of settings. Facial puppetry, speech-driven animation, and face enactment are all examples of animation-based face reconstruction. Face replacement and video dubbing are two examples of video-based applications. Face ageing, virtual makeup, and projection mapping are some of the most common 3D face applications.

7 Challenges And Future Research Directions

This section discusses the main challenges faced during 3D face reconstruction, followed by future research directions.

7.1 Current Challenges

The current challenges in 3D face reconstruction are occlusion removal, makeup removal, expression transfer, and

Table 4 Detail description of datasets used

Dataset Name	Modalities	Total Images	Total Subjects	Emotion Label Availability	Occlusion Label Availability	Publicly Available	Techniques using the dataset
Annotated faces-in-the-wild (AFW) [98]	2D + Landmarks	468	–	No	No	Yes	[63, 31]
Annotated facial landmarks in the wild (AFLW) [99]	2D + Landmarks	25993	–	No	No	Yes	[31]
AFLW2000-3D [85]	2D + Landmarks	2000	–	No	No	Yes	[73, 58, 77, 100, 86, 92, 94]
AFLW-LFPA [84]	2D + Landmarks	1299	–	No	No	Yes	[100]
Age Database (AgeDB) [101]	2D + Age	16488	568	No	No	Yes	[79]
Basel Face Model (BFM2009) [102]	3D	200	–	No	No	Yes	[63, 61]
Bosphorus [103]	2D, 3D	4666	105	Yes	Yes	Yes	[38, 39, 26, 43, 32]
Binghamton University 3D Facial Expression (BU3DFE) [104]	3D	2500	100	Yes	No	Yes	[63, 43, 48, 74, 86]
Binghamton University 4D Facial Expression (BU4DFE) [104]	3D + Time	606	101	Yes	No	Yes	[43, 48, 75]
Chinese Academy of Sciences Institute of Automation (CASIA–3D) [105]	3D	4624	123	Yes	No	Yes	[26, 43]
CASIA-WebFace [106]	2D	494414	10575	No	No	Yes	[45]
CelebFaces Attributes Dataset (CelebA) [81]	2D	202599	10177	Yes	Yes	Yes	[24, 64],[76, 77, 80, 96]
Celebrities in Frontal-Profile in the Wild (CFP) [107]	2D	7000	500	No	No	Yes	[79]
FaceScape [108]	3D	18760	938	Yes	No	Yes	–
Facewarehouse [109]	3D	–	150	Yes	No	Yes	[64, 61, 76, 46, 24]
Face Recognition Grand Challenge (FRGC-v2.0) [110]	3D	4950	466	Yes	No	Yes	[38, 67, 43, 86, 94]
GavabDB [111]	3D	549	61	Yes	No	Yes	[43]
Helen Facial Feature Dataset (Helen) [112]	2D + Landmarks	2330	–	No	No	Yes	[95]
Hi-Lo [47]	3D	6000	–	No	No	Yes	[47]
IARPA Janus Benchmark A (IJB–A) [113]	2D + Landmarks	5712	500	No	No	Yes	[76]
KinectFaceDB [114]	2D, 2.5D, 3D	936	52	Yes	Yes	Yes	[2, 8]
Labeled Faces in the Wild (LFW) [82]	2D + Landmarks	13233	5749	No	No	Yes	[64, 74, 76, 80, 96]

Table 4 (continued)

Dataset Name	Modalities	Total Images	Total Subjects	Emotion Label Availability	Occlusion Label Availability	Publicly Available	Techniques using the dataset
Labeled Face Parts in the Wild (LFPW) [115]	2D + Landmarks	3000	–	No	No	Yes	[31]
Large Scale 3D Faces in the Wild (LS3D-W) [11]	2D + Landmarks	230000	–	No	No	Yes	[58, 76]
Media Integration and Communication Center Florence (MICC-Florence)[116]	2D, 3D	53	53	No	No	Yes	[74, 58, 76, 92, 94, 97]
Notre Dame (ND-2006) [117]	3D	13450	888	Yes	No	Yes	[43]
Texas 3D Face Recognition Database (TexasFRD) [118]	2D + Landmarks, 2.5D	1149	105	No	No	Yes	[43]
University of Houston Database (UHDB31) [119]	3D	25872	77	No	No	Yes	[67]
University of Milano Bicocca 3D Face Database (UMBDB) [120]	2D, 3D	1473	143	Yes	Yes	Yes	[43]
Visual Geometric Group Face Dataset (VGG-Face) [121]	2D + Time	2.6M	2622	No	No	Yes	[64]
VGGFace2 [66]	2D	3.31M	9131	No	No	Yes	[45]
VidTIMIT [122]	Video + Audio	–	43	No	No	Yes	[83]
VoxCeleb2 [123]	2D + Audio	1.12M Audios	6112	No	No	Yes	[93]
WebCaricature [88]	2D	6042 Caricatures + 5974 Images	252	No	No	Yes	[87]
YouTube Faces Database (YTF) [124]	2D + Time	3425 Videos	1595	No	No	Yes	[74]
300 Videos in the Wild (300 VW) [125]	2D+Time	218595	300	No	No	Yes	[64, 83]
300 Faces in-the-wild Challenge (300W-3D) [126]	2D + Landmarks	600	300	No	No	Yes	[75, 58, 77]
300 Faces in-the-wild with Large Poses (300W-LP) [85]	2D, 3D	61225	–	No	No	Yes	[79, 76, 96, 95, 51, 78]
3D Twins Expression Challenge (3D-TEC) [127]	3D	428	214	Yes	No	Yes	[26]
4D Facial Behaviour Analysis for Security (4DFAB) [34]	3D+Time	1.8M	180	Yes	No	No	[33, 47]

Table 5 Comparative analysis of 3D face reconstruction in terms of technique, hardware, and applications

References	Year	Technique	Hardware	RAM (GB)	Applications
[38]	2016	CNN on synthetic data	GPU	8	Real Images
[63]	2017	Cascaded regression	Intel Core i7	8	Real Images
[39]	2017	CNN on synthetic data	Intel Core i7	8	Facial Reenactment
[51]	2017	Direct volumetric CNN regression	GPU	8	Real Images
[64]	2017	Unsupervised deep convolutional autoencoder	GPU	8	Real Images
[67]	2017	End-to-end Deep Neural Network	GPU	8	Real Images
[68]	2017	Deep learning-based sketching	GPU	8	Avatar, Cartoon characters
[29]	2017	Glass-based explicit modelling	Mobile Phone	4	Real Images
[26]	2018	Deep CNN and 3D augmentation	GPU	8	Captured 3D face reconstruction
[61]	2018	Coarse to fine optimization strategy	Intel Core i7	8	Real Images
[30]	2018	RBF interpolation	Intel Core i7	8	3D films and virtual reality
[69]	2018	Landmark localization and deep CNN	Intel Core i7	8	Real Images
[43]	2018	Deep CNN	GPU	8	Real Images
[48]	2018	Epipolar plane images and CNN	GPU	8	Real Images
[70]	2018	Sparse photometric stereo	Intel Core i7	8	Semantic labeling of face into fine region
[71]	2018	Bump map estimation with deep convolution autoencoder	GPU	12	Real Images
[72]	2018	Morphable model, Basel face model, and Cascaded regression	Intel Core i7	8	Real Images
[73]	2018	Position map regression network	GPU	8	Real Images
[74]	2018	Encoder decoder based	GPU	32	Real Images
[31]	2018	3DMM	Intel Core i7	8	Real Images
[32]	2019	Cascaded regression	Intel Core i7	8	Estimation of high-quality 3D face shape from single 2D image
[49]	2019	Best fit blending	Intel Core i7	16	Virtual reality
[75]	2019	CNN regression	GPU	8	Real time application
[58]	2019	Unguided volumetric regression network	Intel Core i7	8	Real Images
[27]	2019	GANs and Deep CNNs	GPU	11	Image augmentation
[76]	2019	Weakly supervised CNN	GPU	8	Real Images
[77]	2019	GANs	GPU	11	De-occluded face generation
[78]	2019	Siamese CNN	Intel Core i7	8	Real Images
[79]	2019	GANs	GPU	4	Image augmentation
[80]	2019	3DMM	Intel Core i7	8	Easy combination of multiple face views
[83]	2019	Encoder decoder based	GPU	8	Video quality enhancement
[44]	2019	RNN and autoencoder	GPU	11	Video avatars, facial reenactment, video dubbing
[33]	2020	Deep neural networks	Intel Core i7	8	Facial affect synthesis of basic expressions
[45]	2020	Symmetry consistent CNN	GPU	11	Natural Images
[46]	2020	Deep CNN	GPU	11	Expression modeling on caricature
[10]	2020	2D assisted self-supervised learning	Intel Core i7	8	Real Images
[47]	2020	GANs	GPU	32	3D face augmentation
[86]	2020	Cascaded coupled regression	GPU	8	Real Images
[24]	2020	Graph convolutional networks	GPU	11	Generate high fidelity 3D face texture
[87]	2020	GANs	GPU	11	Animation, 3D printing, virtual reality
[89]	2020	3DMM and GAN algorithm	Intel Core i7	16	Generation of 4Kx6K 3D face from single 2D face image
[90]	2020	ANN	GPU	16	Expression modeling on caricature
[91]	2020	GANs and VAEs	GPU	8	Vision and graphics
[92]	2020	Adaptive reweighing based optimization	Intel Core i7	8	Real Images
[93]	2020	3DMM and blendshapes	GPU	8	Personalized reconstruction
[94]	2020	Multiview geometry consistency	Intel Core i7	8	Real Images

Table 5 (continued)

References	Year	Technique	Hardware	RAM (GB)	Applications
[35]	2020	3DMM	Intel Core i7	8	Expression modeling
[95]	2020	Attention guided GAN	Intel Core i7	8	2.5D to 3D face generation
[96]	2020	GANs	GPU	12	Face animation and reenactment
[2]	2020	Clustering, VAE, BiLSTM, SVM	GPU	32	Real Images
[8]	2020	End-to-end deep learning	GPU	32	Real Images
[9]	2021	VAE and Triplet Loss	GPU	32	Real Images
[97]	2021	Encoder decoder	GPU	32	Multiface reconstruction

Fig. 12 Face puppetry in real-time [129]

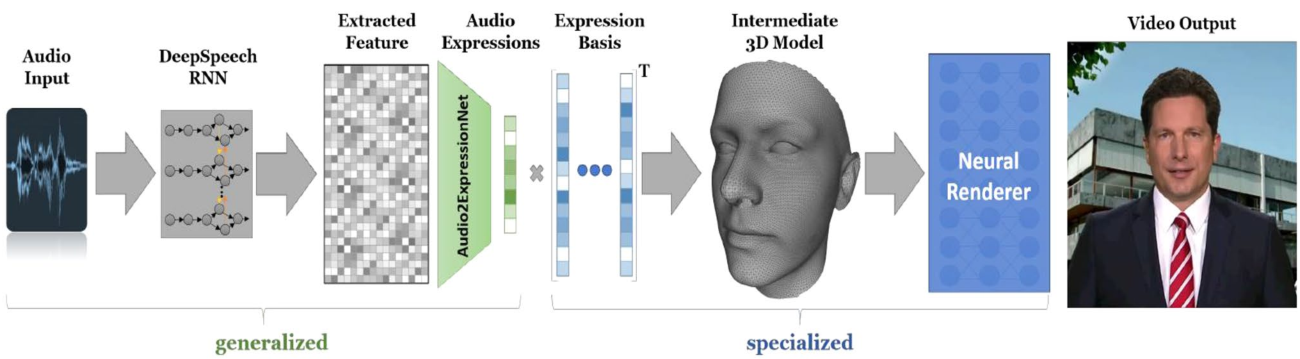


Fig. 13 Neural Voice Puppetry [44]

age prediction. These are discussed in the succeeding subsections.

7.1.1 Occlusion Removal

The occlusion removal is a challenging task for 3D face reconstruction. Researchers are working to handle 3D face occlusion using voxels and 3D landmarks [2, 8, 9]. Sharma and Kumar [2] developed a voxel-based face reconstruction technique. After the reconstruction process, they used a pipeline of variational autoencoders, bidirectional LSTM, and triplet loss training to implement 3D face recognition.

Sharma and Kumar [20] proposed voxel-based face reconstruction and recognition method. They used a generator and discriminator based on game theory for the generation of triplets. The occlusion was removed after the missing information was reconstructed. Sharma and Kumar [22] used 3D face landmarks to build a one-shot learning 3D face reconstruction technique (see Fig. 20).

7.1.2 Applying Make-up and its Removal

Applying the facial makeup and its removal is challenging in virtual meetings during the COVID-19 pandemic [154–156]. makeup bag [154] presented an automatic

Fig. 14 DeepFake example in 6.S191 [133]

(a)

(b)

Fig. 15 **a** Synthesised virtual tattoos [134] and **b** Augmented reality-based pixel-unit makeup on lips [136]

makeup style transfer technique by solving the makeup disentanglement and facial makeup application. The main advantage of MakeupBag is that it considers the skin tone and colour while doing the makeup transfer (Fig. 21).

Li et al. [155] proposed a makeup-invariant face verification system. They employed a semantic aware makeup cleaner (SAMC) to remove face makeup under various expressions and poses. The technique worked

Fig. 16 FaceForge based live projection mapping [137]

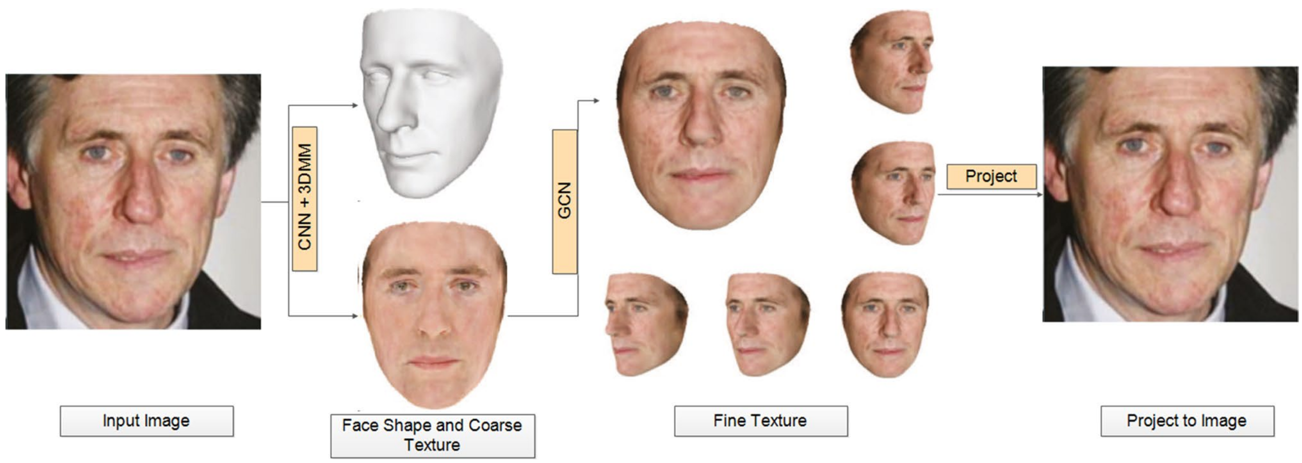


Fig. 17 Projection mapping of a 2D face combined with 3DMM model [24]

Fig. 18 Expression invariant face replacement system [138]

unsupervised while locating the makeup region in the face and used an attention map ranging from 0 to 1, denoting the degree of makeup. Horita and Aizawa [156] proposed a style and latent-guided generative adversarial networks

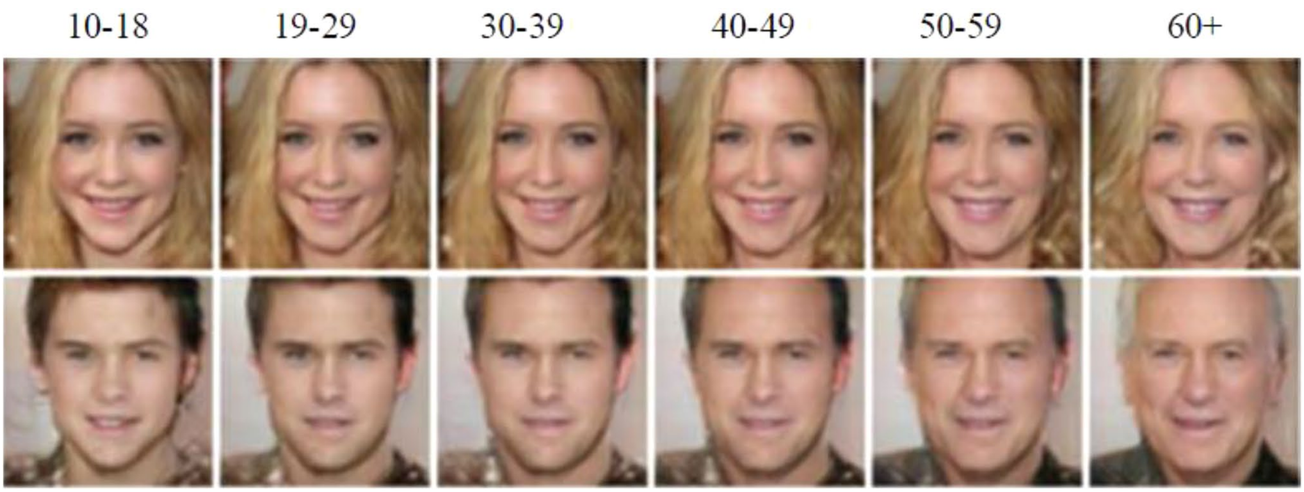


Fig. 19 Transformation of the face using ACGAN [139]

Table 6 Applications of 3D face reconstruction

Broad Area	Target Problems	Techniques / Tools	References
Animation	Facial Puppetry	Displaced dynamic expression (DDE) model and dynamic expression model (DEM)	[129, 130]
	Speech-driven Animation	RNN and Autoencoders	[44]
	Face Enactment	RNN, GAN, Attention-based CNN	[132, 147–151]
Video	Video Dubbing	DeepFake, GANs	[131–133]
	Face Replacement	CNN-based transfer learning, GANs, Adobe Premiere Elements, Apple Final Cut Pro, Filmora	[138, 152, 153]
3D Face	Face Aging	GANs	[139–146]
	Virtual Makeup	GANs, Autoencoders, Augmented Reality	[134–136]
	Projection Mapping	CNN	[24, 137]

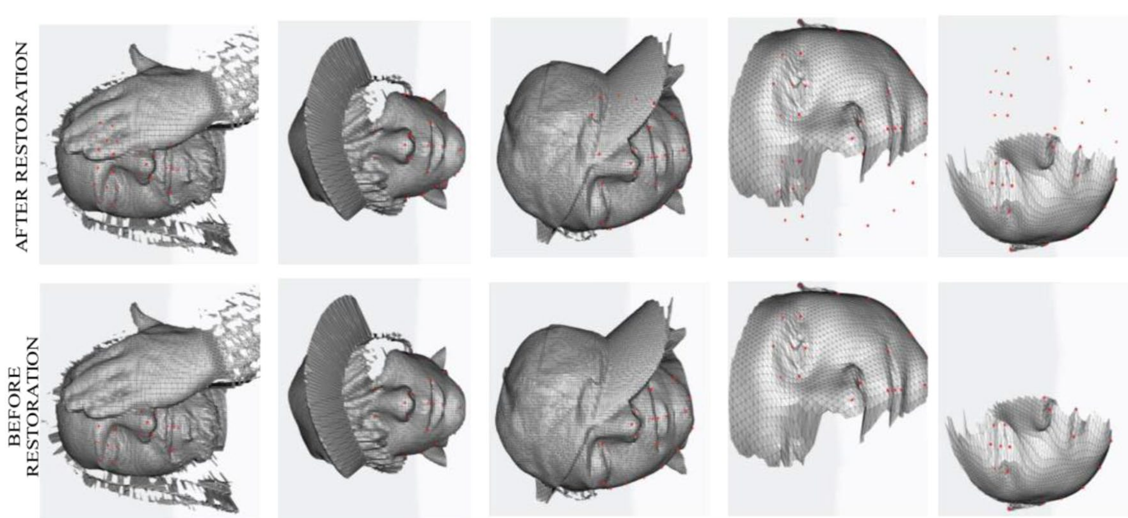


Fig. 20 3D Face reconstruction based on facial landmarks [9]

(SLGAN). They used controllable GAN to enable the user with adjustment of makeup shading (see Fig. 22).

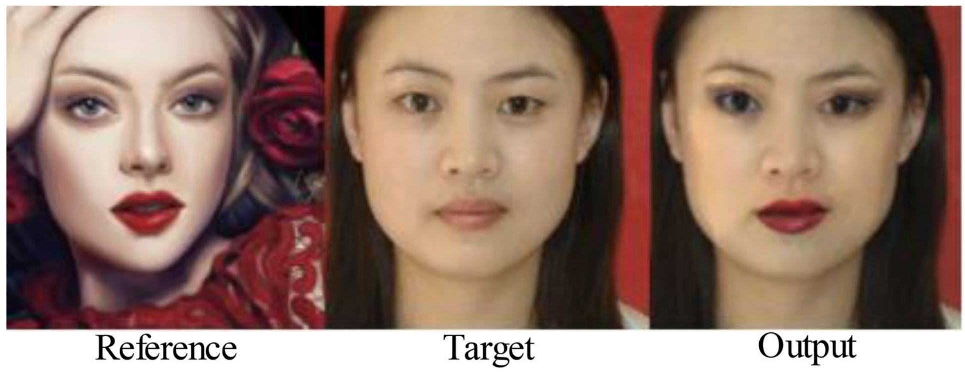


Fig. 21 MakeupBag based output for applying makeup from reference to target face [154]

7.1.3 Expression Transfer

Expression transfer is an active problem, especially with the advent of GANs. Wu et al. [157] proposed ReenactGAN, a

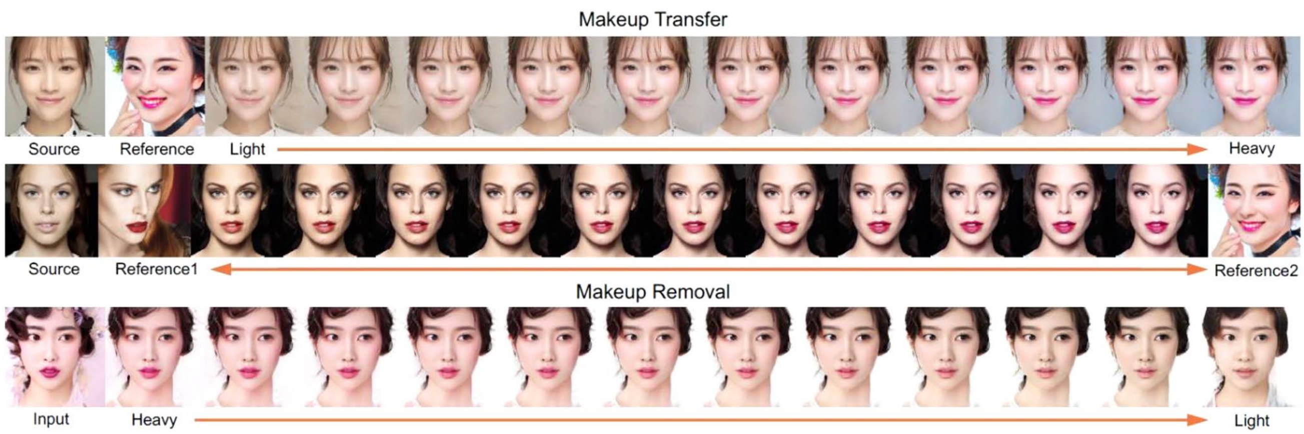


Fig. 22 GAN based makeup transfer and removal [156]

Fig. 23 Expression transfer using ReenactGAN [157]

method capable of transferring the person's expressions from source video to target video. They employed the encoder-decoder based model for doing the transformation of the face from source to target. The transformer used three loss functions for evaluation, viz. cycle-loss, adversarial loss, and shape constrain loss. Donald Trump images reenacting the expressions are depicted in Fig. 23.

Deep fakes are a matter of concern where the facial expression and context are different. Nirkin et al. [158] proposed a deep fake detection method to detect identity manipulations and face swaps. In deep fake images, the face regions are manipulated by targeting the face to change by variation in the context. Tolosana et al. [159] worked on a survey for four kinds of deep fake methods, including full face synthesis, identity swapping, face attribute manipulation, and expression swapping.

7.1.4 Age Prediction

Due to deep fakes and generative adversarial networks [140, 142], the faces can be deformed to other ages, as seen in Fig. 24. With this, the challenge of age prediction of a person goes beyond imagination, especially in fake faces on identity cards or social networking platforms.

Fang et al. [141] proposed a GAN-based technique for face age simulation. The proposed Triple-GAN model used the triple translation loss for the modelling of age pattern interrelation. They employed an encoder-decoder based generator and discriminator for age classification. Kumar et al. [160] employed reinforcement learning over the latent space based on the GAN model [161]. They used Markov Decision Process (MDP) for doing semantic manipulation. Pham et al. [162] proposed a semi-supervised GAN technique to generate realistic face images. They synthesised the face images using the real data and the target age while training the network. Zhu et al. [163] used the attention-based conditional GAN technique to target high-fidelity in the synthesised face images.

7.2 Future Challenges

Unsupervised learning in 3D face reconstruction is an open problem. Work has been presented lately by [164] to work around symmetric deformable objects in 3D. In this paper, some future possibilities for 3D face reconstruction such as lips reconstruction, teeth and tongue capturing, eyes and eyelids capturing, hairstyle, and full head reconstruction have been discussed in detail. These challenges have been laid out for the researchers working in the domain of 3D face reconstruction.

7.2.1 Lips Reconstruction

The lips are one of the most critical components of the mouth area. Various celebrities get surgeries on lips ranging from lip lift surgery, lip reduction surgery, and lip augmentation surgery [165, 166]. Heidekrueger et al. [165] surveyed the preferred lip ratio for females. It was concluded that gender, age, profession, and country might affect the preferences of lower lip ratio.

Baudoin et al. [166] conducted a review on upper lip aesthetics. Various treatment options ranging from fillers to dermabrasion and surgical excision were examined. Zollhofer et al. [1] discussed lip reconstruction as one application for 3D face reconstruction, as shown in Fig. 25. In [167], the lips' video reconstructed the rolling, stretching, and bending of lips.

7.2.2 Teeth and Tongue Capturing

In literature, few works have worked on capturing the interior of the mouth. Reconstructing teeth and tongue in GAN-based 2D faces is a difficult task. A beard or moustache can

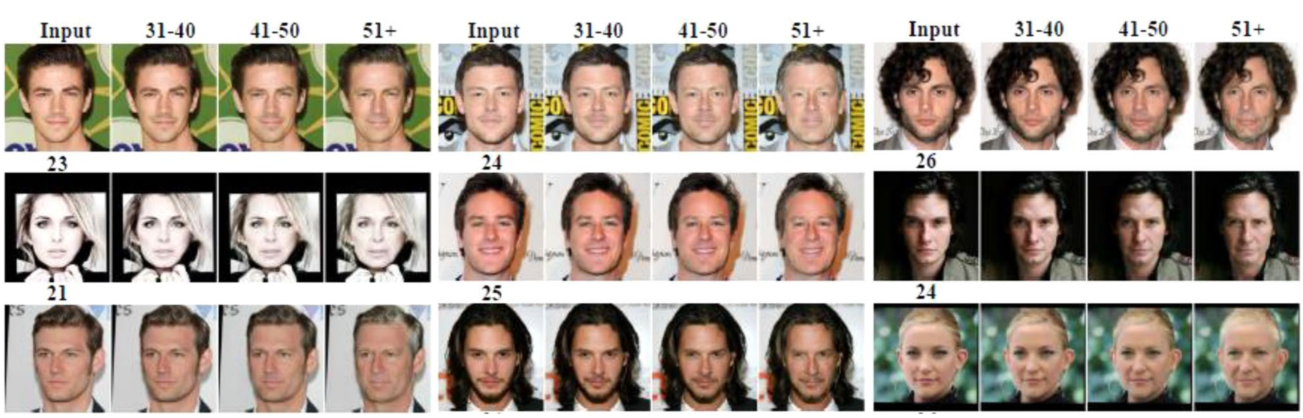


Fig. 24 Results of progressive face aging GAN [142]

Fig. 25 High-quality lip shapes for reconstruction [1]

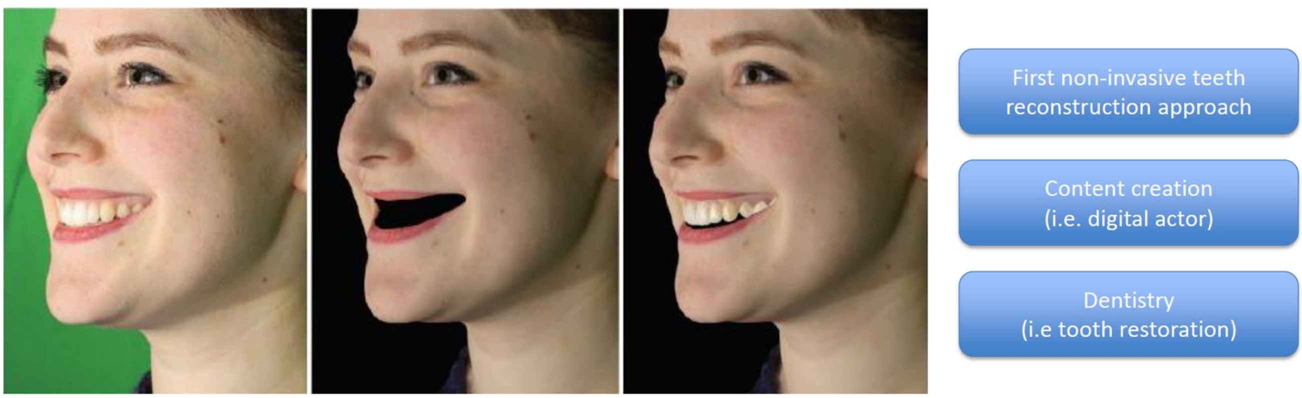


Fig. 26 Teeth reconstruction with its applications [168]

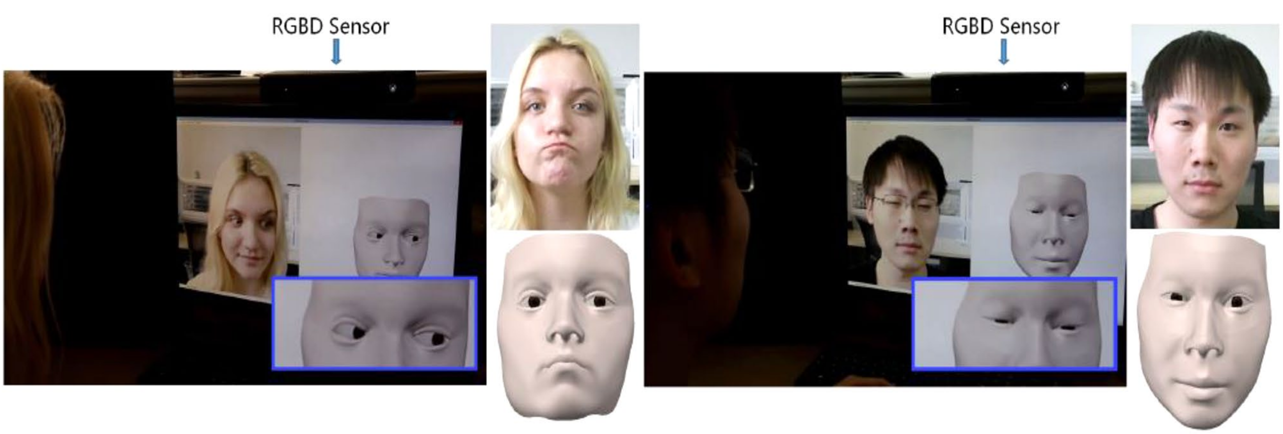


Fig. 27 Eyelid tracking based on semantic edges [169]

make it difficult to capture the teeth and tongue. In [163] a statistical model was discussed.. There are different applications for reconstructing the teeth area, viz. production of content for digital avatars and dentistry for facial geometry-based tooth restoration (see Fig. 26).

7.2.3 Eyes and Eyelids Capturing

Wang et al. [170] showed 3D eye gaze estimation and facial reconstruction from an RGB video. Wen et al. [169] presented a technique for tracking and reconstructing 3D eyelids

Fig. 28 3D Hair Synthesis using volumetric VAE [172]

in real-time (see Fig. 27). This approach is combined with a face and eyeball tracking system to achieve a full face with detailed eye regions. In [171], Bi-directional LSTM was employed for eyelids tracking.

7.2.4 Hair Style Reconstruction

Hair style reconstruction is a challenging task on 3D faces. A volumetric variational autoencoder based 3D hair synthesis [172] is shown in Fig. 28. Ye et al. [173] proposed a hair strand reconstruction model based on the encoder-decoder technique. It generated a volumetric vector field using the hairstyle-based oriented map. They used a mixture of CNN layers, skip connections, fully connected layers, and the deconvolution layers while generating the architecture in encoder-decoder format. Structure and content loss was used as the evaluation metric during the training process.

7.2.5 Complete Head Reconstruction

The reconstruction of the 3D human head is an active area of research. He et al. [174] presented a full head data-driven 3D face reconstruction. The input image and reconstructed result with a side-view texture were generated (see Fig. 29). They employed the albedo parameterised model for complementing the head texture map. A convolution network was used for face and hair region segmentation. There are various applications of human head reconstruction in virtual reality as well as avatar generation.

Table 7 presents the challenges and future directions along with their target problems.

8 Conclusion

This paper presents a detailed survey with extensive study for 3D face reconstruction techniques. Initially, six types of reconstruction techniques have been discussed. The observation is that scalability is the biggest challenge for 3D face problems because 3D faces do not have sufficiently large datasets publicly available. Most of the researchers have worked on RGB-D images. With deep learning, working on a mesh image or the voxel image has hardware constraints. The current and future challenges related to 3D face reconstruction in the real world have been discussed. This domain is an open area of research ranging with many challenges,

Fig. 29 Full head reconstruction [174]

Table 7 Challenges and future research directions for 3D face reconstruction

Challenges	Target Problem	Technique Used	References
Occlusion Removal	Forensics and surveillance	GAN, VAE, BiLSTM, and Triplet Loss	[2, 8, 9]
Makeup Removal	Online meetings, forensics, and cosmetics	Controllable GAN	[154–156]
Expression Transfer	Animation and dubbing	Encoder-Decoder based GAN	[157–159]
Age Prediction	Photography, fashion, and robotics	Conditional GAN	[141, 160–163]
Lips Reconstruction	Surgery and AI in medicines	Surgery-based	[165, 166]
Teeth and Tongue Capturing	3D modeling	GAN	[168]
Eyes and Eyelids Capturing	Proctored examinations	BiLSTM	[169–171]
Hair Style	Cosmetics and hair style industry	CNN, Autoencoder	[172, 173]
Complete Head	Augmented reality and Virtual reality	CNN	[174]

especially with the capabilities of GANs and deep fakes. The study is unexplored in lips reconstruction, the interior of mouth reconstruction, eyelids reconstruction, hair styling for various hair, and complete head reconstruction.

Declarations

Conflict of interest On behalf of all authors, the corresponding author states that there is no conflict of interest.

References

1. Zollhöfer M, Thies J, Garrido P et al (2018) State of the art on monocular 3D face reconstruction, tracking, and applications. Comput Graph Forum 37(2):523–550. https://doi.org/10.1111/cgf.13382
2. Sharma S, Kumar V (2020) Voxel-based 3D face reconstruction and its application to face recognition using sequential deep learning. Multimed Tools Appl 79:17303–17330. https://doi.org/10.1007/s11042-020-08688-x
3. Cloud Vision API | Google Cloud. https://cloud.google.com/vision/docs/face-tutorial. Accessed 12 Jan 2021
4. AWS Marketplace: Deep Vision API. https://aws.amazon.com/marketplace/pp/Deep-Vision-AI-Inc-Deep-Vision-API/B07JHXVZ4M. Accessed 12 Jan 2021
5. Computer Vision | Microsoft Azure. https://azure.microsoft.com/en-in/services/cognitive-services/computer-vision/. Accessed 12 Jan 2021
6. Koujan MR, Dochev N, Roussos A (2020) Real-Time Monocular 4D Face Reconstruction using the LSFM models. preprint arXiv:2006.10499.
7. Behzad M, Vo N, Li X, Zhao G (2021) Towards reading beyond faces for sparsity-aware 4D affect recognition. Neurocomputing 458:297–307
8. Sharma S, Kumar V (2020) Voxel-based 3D occlusion-invariant face recognition using game theory and simulated annealing. Multimedia Tools and Applications 79(35):26517–26547
9. Sharma S, Kumar V (2021) 3D landmark-based face restoration for recognition using variational autoencoder and triplet loss. IET Biometrics 10(1):87–98. https://doi.org/10.1049/bme2.12005
10. Tu X, Zhao J, Xie M et al (2020) 3D face reconstruction from a single image assisted by 2D face images in the Wild. IEEE

Trans Multimed 23:1160–1172. https://doi.org/10.1109/TMM.2020.2993962
11. Bulat A, Tzimiropoulos G How far are we from solving the 2D & 3D Face Alignment problem? (and a dataset of 230,000 3D facial landmarks). In: Proceedings of the IEEE international conference on computer vision (ICCV), pp 1021–1030
12. Zhu X, Lei Z, Liu X, et al (2016) Face alignment across large poses: a 3D solution. computer vision and pattern recognition (CVPR), pp 146–155
13. Gu S, Bao J, Yang H, et al (2019) Mask-guided portrait editing with conditional gans. In: Proc IEEE comput soc conf comput vis pattern recognit 2019-June:3431–3440. doi: https://doi.org/10.1109/CVPR.2019.00355
14. Guo Y, Wang H, Hu Q et al (2020) Deep learning for 3D point clouds: A survey. IEEE Trans Pattern Anal Mach Intell 43(12):4338–4364. https://doi.org/10.1109/tpami.2020.3005434
15. Ye M, Shen J, Lin G et al (2021) Deep learning for person re-identification: a survey and outlook. IEEE Trans Pattern Anal Mach Intell 8828:1–1. https://doi.org/10.1109/tpami.2021.3054775
16. Tran L, Liu X Nonlinear 3D Face Morphable Model. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 7346–7355
17. Egger B, Smith WAP, Tewari A et al (2020) 3D morphable face models—past, present, and future. ACM Trans Graph 39(5):1–38. https://doi.org/10.1145/3395208
18. Blanz V, Vetter T (1999) Face recognition based on fitting a 3D morphable model. IEEE Trans Pattern Anal Mach Intell 25(9):1063–1074
19. Booth J, Roussos A, Ponniah A et al (2018) Large scale 3D morphable models. Int J Comput Vis 126:233–254. https://doi.org/10.1007/s11263-017-1009-7
20. Cao C, Weng Y, Zhou S et al (2014) FaceWarehouse: A 3D facial expression database for visual computing. IEEE Trans Vis Comput Graph 20:413–425. https://doi.org/10.1109/TVCG.2013.249
21. Gerig T, Morel-Forster A, Blumer C, et al (2018) Morphable face models - An open framework. In: Proceedings 13th IEEE int conf autom face gesture recognition, FG. 75–82. https://doi.org/10.1109/FG.2018.00021
22. Huber P, Hu G, Tena R, et al. (2016) A multiresolution 3d morphable face model and fitting framework. InProceedings of the 11th joint conference on computer vision, imaging and computer graphics theory and applications, pp 79–86. SciTePress.
23. Li T, Bolkart T et al (2017) Learning a model of facial shape and expression from 4D scans. ACM Trans Graphics 36(6):1–17. https://doi.org/10.1145/3130800.3130813
24. Lin J, Yuan Y, Shao T, Zhou K (2020) Towards high-fidelity 3D face reconstruction from in-the-wild images using graph convolutional networks. Comput Vision Pattern Recognition (CVPR). https://doi.org/10.1109/cvpr42600.2020.00593
25. Paysan P, Knothe R, Amberg B, et al (2009) A 3D face model for pose and illumination invariant face recognition. In: 6th IEEE international conference on advanced video and signal based surveillance, AVSS 2009. pp 296–301
26. Kim D, Hernandez M, Choi J, Medioni G (2018) Deep 3D face identification. IEEE international joint conference on biometrics (IJCB), IJCB 2017 2018-January:133–142. https://doi.org/10.1109/BTAS.2017.8272691
27. Gecer B, Ploumpis S, Kotsia I, Zafeiriou S (2019) Ganfit: Generative adversarial network fitting for high fidelity 3D face reconstruction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition:1155–1164. https://doi.org/10.1109/CVPR.2019.00125
28. Kim H, Garrido P, Tewari A et al (2018) Deep video portraits. ACM Trans Graphics 37:1–14. https://doi.org/10.1145/3197517.3201283
29. Maninchedda F, Oswald MR, Pollefeys M (2017) Fast 3D reconstruction of faces with glasses. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (CVPR). https://doi.org/10.1109/CVPR.2017.490
30. Zhang S, Yu H, Wang T et al (2018) Dense 3D facial reconstruction from a single depth image in unconstrained environment. Virtual Reality 22(1):37–46. https://doi.org/10.1007/s10055-017-0311-6
31. Jiang L, Wu X, Kittler J (2018) Pose invariant 3D face reconstruction. 1–8. arXiv preprint arXiv:1811.05295
32. Wu F, Li S, Zhao T et al (2019) Cascaded regression using landmark displacement for 3D face reconstruction. Pattern Recogn Lett 125:766–772. https://doi.org/10.1016/j.patrec.2019.07.017
33. Kollias D, Cheng S, Ververas E et al (2020) Deep neural network augmentation: generating faces for affect analysis. Int J Comput Vision 128:1455–1484. https://doi.org/10.1007/s11263-020-01304-3
34. 4DFAB: A Large Scale 4D Facial Expression Database for Biometric Applications | DeepAI. https://deepai.org/publication/4dfab-a-large-scale-4d-facial-expression-database-for-biometric-applications. Accessed 14 Oct 2020
35. Lyu J, Li X, Zhu X, Cheng C (2020) Pixel-Face: A Large-Scale, High-Resolution Benchmark for 3D Face Reconstruction. arXiv preprint arXiv:2008.12444
36. Zhu Z, Luo P, Wang X, Tang X (2013) Deep learning identity-preserving face space. In: Proceedings of the IEEE international conference on computer vision. institute of electrical and electronics engineers inc., pp 113–120
37. Tang Y, Salakhutdinov R, Hinton G (2012) Deep Lambertian Networks. arXiv preprint arXiv:1206.6445
38. Richardson E, Sela M, Kimmel R (2016) 3D face reconstruction by learning from synthetic data. In: Proceedings - 2016 4th international conference on 3D vision, 3DV 2016. Institute of electrical and electronics engineers inc., pp 460–467
39. Richardson E, Sela M, Or-El R, Kimmel R (2017) Learning detailed face reconstruction from a single image. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 1259–1268
40. Laine S, Karras T, Aila T, et al. (2016) Facial performance capture with deep neural networks. arXiv preprint arXiv:1609.06536, 3
41. Nair V, Susskind J, Hinton GE (2008) Analysis-by-synthesis by learning to invert generative black boxes. In: International conference on artificial neural networks, pp 971–981
42. Peng X, Feris RS, Wang X, Metaxas DN (2016) A recurrent encoder-decoder network for sequential face alignment. In: European conference on computer vision, pp 38–56.
43. Zulqarnain Gilani S, Mian A (2018) Learning from millions of 3D scans for large-scale 3D face recognition. Proceedings IEEE Comput soc conf comput vis pattern recognit, pp 1896–1905. https://doi.org/10.1109/CVPR.2018.00203
44. Thies J, Elgharib M, Tewari A, et al (2019) Neural voice puppetry: audio-driven facial reenactment. In: European conference on computer vision, pp 716–731
45. Li X, Hu G, Zhu J et al (2020) Learning symmetry consistent deep CNNs for face completion. IEEE Trans Image Proc 29:7641–7655. https://doi.org/10.1109/TIP.2020.3005241
46. Han X, Hou K, Du D et al (2020) CaricatureShop: personalized and photorealistic caricature sketching. IEEE Trans Vis Comput Graphics 26:2349–2361. https://doi.org/10.1109/TVCG.2018.2886007
47. Moschoglou S, Ploumpis S, Nicolaou MA et al (2020) 3DFaceGAN: adversarial nets for 3D face representation, generation, and translation. Int J Comput Vision 128(10):2534–2551. https://doi.org/10.1007/s11263-020-01329-8

48. Feng M, Zulqarnain Gilani S, Wang Y, et al (2018) 3D face reconstruction from light field images: a model-free approach. Lect Notes Comput Science (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 11214 LNCS: 508–526. https://doi.org/10.1007/978-3-030-01249-6_31
49. Anbarjafari G, Haamer RE, LÜSi I, et al (2019) 3D face reconstruction with region based best fit blending using mobile phone for virtual reality based social media. Bull Polish Acad Sci Tech Sci. 67: 125–132. https://doi.org/10.24425/bpas.2019.127341
50. Kim H, Zollhöfer M, Tewari A, et al (2018) InverseFaceNet: deep monocular inverse face rendering. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4625–4634
51. Jackson AS, Bulat A, Argyriou V, Tzimiropoulos G (2017) Large Pose 3D Face Reconstruction from a single image via direct volumetric CNN regression. In: Proceedings IEEE int conf comput Vis 2017-Octob:1031–1039. https://doi.org/10.1109/ICCV.2017.117
52. Eigen D, Puhrsch C, Fergus R (2014) Depth map prediction from a single image using a multi-scale deep network. arXiv preprint arXiv:1406.2283
53. Saxena A, Chung SH, Ng AY (2008) 3-D depth reconstruction from a single still image. Int J Comput Vis 76:53–69. https://doi.org/10.1007/s11263-007-0071-y
54. Tulsiani S, Zhou T, Efros AA, Malik J (2017) Multi-view supervision for single-view reconstruction via differentiable ray consistency. In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 2626–2634
55. Tatarchenko M, Dosovitskiy A, Brox T (2017) Octree generating networks: efficient convolutional architectures for high-resolution 3D outputs. In: Proceedings of the IEEE international conference on computer vision, pp 2088–2096
56. Roth J, Tong Y, Liu X (2016) Adaptive 3D face reconstruction from unconstrained photo collections, In: Proceedings of the IEEE conference on computer vision and pattern recognition, pp 4197–4206
57. Kemelmacher-Shlizerman I, Seitz SM (2011) Face reconstruction in the wild. In: Proceedings of the IEEE international conference on computer vision. pp 1746–1753
58. Xing Y, Tewari R, Mendonça PRS (2019) A self-supervised bootstrap method for single-image 3D face reconstruction. Proc - 2019 IEEE Winter Conf Appl Comput Vision. WACV 2019:1014–1023. https://doi.org/10.1109/WACV.2019.00113
59. Kemelmacher-Shlizerman I, Basri R (2011) 3D face reconstruction from a single image using a single reference face shape. IEEE Trans Pattern Anal Mach Intell 33:394–405. https://doi.org/10.1109/TPAMI.2010.63
60. Sengupta S, Lichy D, Kanazawa A et al (2020) SfSNet: learning shape, reflectance and illuminance of faces in the wild. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/TPAMI.2020.3046915
61. Jiang L, Zhang J, Deng B et al (2018) 3D face reconstruction with geometry details from a single image. IEEE Trans Image Process 27:4756–4770. https://doi.org/10.1109/TIP.2018.2845697
62. 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
63. Liu F, Zeng D, Li J, Zhao Q, jun (2017) On 3D face reconstruction via cascaded regression in shape space. Front Inf Technol Electron Eng 18:1978–1990. https://doi.org/10.1631/FITEE.1700253
64. Tewari A, Zollhöfer M, Kim H, et al (2017) MoFA: Model-based deep convolutional face autoencoder for unsupervised monocular reconstruction. In: Proceedings - 2017 IEEE Int Conf Comput Vis Work ICCVW 2017 2018-Janua:1274–1283. https://doi.org/10.1109/ICCVW.2017.153
65. Krizhevsky A, Sutskever I, Hinton GE (2012) ImageNet classification with deep convolutional neural networks. Adv Neural Inf Process Syst 25:1097–1105
66. Visual Geometry Group - University of Oxford. http://www.robots.ox.ac.uk/~vgg/data/vgg_face/. Accessed 13 Oct 2020
67. Dou P, Shah SK, Kakadiaris IA (2017) End-to-end 3D face reconstruction with deep neural networks. In: Proceedings- 30th IEEE conf comput vis pattern recognition, CVPR 1503–1512. https://doi.org/10.1109/CVPR.2017.164
68. Han X, Gao C, Yu Y (2017) DeepSketch2Face: a deep learning based sketching system for 3D face and caricature modeling. ACM Trans Graph 36:1–12. https://doi.org/10.1145/3072959.3073629
69. Hsu GS, Shie HC, Hsieh CH, Chan JS (2018) Fast landmark localization with 3D component reconstruction and CNN for cross-pose recognition. IEEE Trans Circuits Syst Video Technol 28:3194–3207. https://doi.org/10.1109/TCSVT.2017.2748379
70. Cao X, Chen Z, Chen A et al (2018) Sparse photometric 3D face reconstruction guided by morphable models. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. https://doi.org/10.1109/CVPR.2018.00487
71. Tran AT, Hassner T, Masi I et al (2018) Extreme 3D face reconstruction: seeing through occlusions. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. https://doi.org/10.1109/CVPR.2018.00414
72. Feng ZH, Huber P, Kittler J, et al (2018) Evaluation of dense 3D reconstruction from 2D face images in the wild. In: Proceedings - 13th IEEE int conf autom face gesture recognition, FG 2018 780–786. https://doi.org/10.1109/FG.2018.00123
73. Feng Y, Wu F, Shao X, et al (2018) Joint 3d face reconstruction and dense alignment with position map regression network. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 11218 LNCS:557–574. https://doi.org/10.1007/978-3-030-01264-9_33
74. Liu F, Zhu R, Zeng D et al (2018) Disentangling features in 3d face shapes for joint face reconstruction and recognition. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. https://doi.org/10.1109/CVPR.2018.00547
75 Chinaev N, Chigorin A, Laptev I (2019) MobileFace: 3D face reconstruction with efficient CNN regression. In: Leal-Taixé Laura, Roth Stefan (eds) Computer Vision – ECCV 2018 Workshops: Munich, Germany, September 8-14, 2018, Proceedings, Part IV. Springer International Publishing, Cham, pp 15–30. https://doi.org/10.1007/978-3-030-11018-5_3
76. Deng Y, Yang J, Xu S, et al (2019) Accurate 3D face reconstruction with weakly-supervised learning: From single image to image set. IEEE Comput Soc Conf Comput Vis Pattern Recognit Work 2019-June:285–295. https://doi.org/10.1109/CVPRW.2019.00038
77. Yuan X, Park IK (2019) Face de-occlusion using 3D morphable model and generative adversarial network. In: Proceedings IEEE int conf comput vis 2019-Octob:10061–10070. https://doi.org/10.1109/ICCV.2019.01016
78. Luo Y, Tu X, Xie M (2019) Learning robust 3D face reconstruction and discriminative identity representation. 2019 2nd IEEE int conf inf commun signal process ICICSP 2019 317–321. https://doi.org/10.1109/ICICSP48821.2019.8958506
79. Gecer B, Lattas A, Ploumpis S et al (2019) Synthesizing coupled 3D face modalities by trunk-branch generative adversarial networks. European conference on computer vision. Springer, Cham, pp 415–433
80. Chen Y, Wu F, Wang Z et al (2019) Self-supervised Learning of Detailed 3D Face Reconstruction. IEEE Trans Image Process 29:8696–8705

81. Large-scale CelebFaces Attributes (CelebA) Dataset. http://mmlab.ie.cuhk.edu.hk/projects/CelebA.html. Accessed 13 Oct 2020
82. Labelled Faces in the Wild (LFW) Dataset | Kaggle. https://www.kaggle.com/jessicali9530/lfw-dataset. Accessed 13 Oct 2020
83. Ren W, Yang J, Deng S, et al (2019) Face video deblurring using 3D facial priors. In: Proceedings IEEE int conf comput vis 2019-Octob:9387–9396. https://doi.org/10.1109/ICCV.2019.00948
84. Jourabloo A, Liu X (2015) Pose-invariant 3D face alignment. In Proceedings of the IEEE international conference on computer vision. pp 3694–3702
85. Cheng S, Kotsia I, Pantic M, et al. (2018) 4DFAB: a large scale 4D facial expression database for biometric applications. https://arxiv.org/pdf/1712.01443v2.pdf. Accessed 14 Oct 2020
86. Liu F, Zhao Q, Liu X, Zeng D (2020) Joint face alignment and 3D face reconstruction with application to face recognition. IEEE Trans Pattern Anal Mach Intell 42:664–678. https://doi.org/10.1109/TPAMI.2018.2885995
87. Ye Z, Yi R, Yu M, et al (2020) 3D-CariGAN: an end-to-end solution to 3D caricature generation from face photos. 1–17. arXiv preprint arXiv:2003.06841
88. Huo J, Li W, Shi Y, et al. (2017) Webcaricature: a benchmark for caricature recognition. arXiv preprint arXiv:1703.03230
89. Lattas A, Moschoglou S, Gecer B, et al (2020) AvatarMe: realistically renderable 3D facial reconstruction "In-the-Wild." 757–766. https://doi.org/10.1109/cvpr42600.2020.00084
90. Cai H, Guo Y, Peng Z, Zhang J (2021) Landmark detection and 3D face reconstruction for caricature using a nonlinear parametric model. Graphical Models 115:101103. https://doi.org/10.1016/j.gmod.2021.101103
91. Deng Y, Yang J, Chen D, et al (2020) Disentangled and controllable face image generation via 3D imitative-contrastive learning. https://doi.org/10.1109/cvpr42600.2020.00520
92. Li K, Yang J, Jiao N, et al (2020) Adaptive 3D face reconstruction from a single image. 1–11. arXiv preprint arXiv:2007.03979
93 Chaudhuri B, Vesdapunt N, Shapiro L, Wang B (2020) Personalized Face Modeling for Improved Face Reconstruction and Motion Retargeting. In: Vedaldi A, Bischof H, Brox T, Frahm J-M (eds) Computer Vision – ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part V. Springer International Publishing, Cham, pp 142–160. https://doi.org/10.1007/978-3-030-58558-7_9
94. Shang J, Shen T, Li S, et al (2020) Self-supervised monocular 3D face reconstruction by occlusion-aware multi-view geometry consistency. In: computer vision–ECCV 2020: 16th European conference, Glasgow, UK, August 23–28, 2020, proceedings, Part XV 16 (pp. 53–70). Springer International Publishing
95. Cai X, Yu H, Lou J, et al (2020) 3D facial geometry recovery from a depth view with attention guided generative adversarial network. arXiv preprint arXiv:2009.00938
96. Xu S, Yang J, Chen D, et al (2020) Deep 3D portrait from a single image. 7707–7717. https://doi.org/10.1109/cvpr42600.2020.00773
97. Zhang J, Lin L, Zhu J, Hoi SCH (2021) Weakly-supervised multi-face 3D reconstruction. 1–9. arXiv preprint arXiv:2101.02000
98. Köstinger M, Wohlhart P, Roth PM, Bischof H (2011) Annotated facial landmarks in the wild: a large-scale, real-world database for facial landmark localization. Proc IEEE Int Conf Comput Vis. https://doi.org/10.1109/ICCVW.2011.6130513
99. ICG - AFLW. https://www.tugraz.at/institute/icg/research/team-bischof/lrs/downloads/aflw/. Accessed 14 Oct 2020
100. Tu X, Zhao J, Jiang Z et al (2019) 3D face reconstruction from a single image assisted by 2D face images in the wild. IEEE Trans Multimed. https://doi.org/10.1109/TMM.2020.2993962
101. Moschoglou S, Papaioannou A, Sagonas C, et al (2017) AgeDB: the first manually collected, in-the-wild age database. In: proceedings of the IEEE conference on computer vision and pattern recognition workshops, pp 51–59
102. Morphace. https://faces.dmi.unibas.ch/bfm/main.php?nav=1-1-0&id=details. Accessed 14 Oct 2020
103. Savran A, Alyüz N, Dibeklioğlu H et al (2008) Bosphorus Database for 3D Face Analysis. European workshop on biometrics and identity management. Springer, Berlin, Heidelberg, pp 47–56
104. 3D facial expression database - Binghamton University. http://www.cs.binghamton.edu/~lijun/Research/3DFE/3DFE_Analysis.html. Accessed 13 Oct 2020
105. Center for Biometrics and Security Research. http://www.cbsr.ia.ac.cn/english/3DFace Databases.asp. Accessed 14 Oct 2020
106. Yi D, Lei Z, Liao S, Li SZ (2014) Learning face representation from scratch. arXiv preprint arXiv:1411.7923
107. Celebrities in Frontal-Profile in the Wild. http://www.cfpw.io/. Accessed 14 Oct 2020
108. Yang H, Zhu H, Wang Y, et al (2020) FaceScape: a large-scale high quality 3D face dataset and detailed riggable 3D face prediction. 598–607
109. FaceWarehouse. http://kunzhou.net/zjugaps/facewarehouse/. Accessed 13 Oct 2020
110. Phillips PJ, Flynn PJ, Scruggs T, et al. (2005) Overview of the face recognition grand challenge. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05), 1: 947–954
111. MORENO, A. (2004) GavabDB : a 3d face database. In: Proceedings 2nd COST275 work biometrics internet, 2004 75–80
112. Le V, Brandt J, Lin Z, et al (2012) Interactive facial feature localization. Lect Notes Comput Sci (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics) 7574 LNCS:679–692. https://doi.org/10.1007/978-3-642-33712-3_49
113. IJB-A Dataset Request Form | NIST. https://www.nist.gov/itl/iad/image-group/ijb-dataset-request-form. Accessed 14 Oct 2020
114. Min R, Kose N, Dugelay JL (2014) KinectfaceDB: a kinect database for face recognition. IEEE Trans Syst Man, Cybern Syst 44:1534–1548. https://doi.org/10.1109/TSMC.2014.2331215
115. Belhumeur PN, Jacobs DW, Kriegman DJ, Kumar N (2011) Localizing parts of faces using a consensus of exemplars. Proc IEEE Comput Soc Conf Comput Vis Pattern Recognit. https://doi.org/10.1109/CVPR.2011.5995602
116. Bagdanov AD, Del Bimbo A, Masi I (2011) The florence 2D/3D hybrid face dataset. In: Proceedings of the 2011 joint ACM workshop on Human gesture and behavior understanding - J-HGBU '11. ACM Press, New York, New York, USA, p 79
117. Notre Dame CVRL. https://cvrl.nd.edu/projects/data/#nd-2006-data-set. Accessed 13 Oct 2020
118. Laboratory for Image and Video Engineering - The University of Texas at Austin. http://live.ece.utexas.edu/research/texas3dfr/. Accessed 14 Oct 2020
119. Le HA, Kakadiaris IA (2017) UHDB31: A dataset for better understanding face recognition across pose and illumination variation. In: 2017 IEEE international conference on computer vision workshops (ICCVW). IEEE, pp 2555–2563
120. Colombo A, Cusano C, Schettini R (2011) UMB-DB: a database of partially occluded 3D faces. In: Proceedings of the IEEE international conference on computer vision. pp 2113–2119
121. Parkhi OM, Vedaldi A, Zisserman A, (2015) Deep Face Recognition. pp 1–12
122. Sanderson C (2002) The VidTIMIT Database. (No. REP_WORK). IDIAP

123. Son Chung J, Nagrani A, Zisserman A, (2018) VoxCeleb2: deep speaker recognition. arXiv preprint arXiv:1806.05622
124. YouTube Faces Database : Main. https://www.cs.tau.ac.il/~wolf/ytfaces/. Accessed 14 Oct 2020
125. 300-VW | Computer Vision Online. https://computervisiononline.com/dataset/1105138793. Accessed 13 Oct 2020
126. i·bug - resources - 300 Faces In-the-Wild Challenge (300-W), ICCV 2013. https://ibug.doc.ic.ac.uk/resources/300-W/. Accessed 14 Oct 2020
127. Vijayan V, Bowyer K, Flynn P (2011) 3D twins and expression challenge. In: Proceedings of the IEEE international conference on computer vision. pp 2100–2105
128. AI + X: Don't Switch Careers, Add AI - YouTube. https://www.youtube.com/watch?v=4Ai7wmUGFNA. Accessed 5 Feb 2021
129. Cao C, Hou Q, Zhou K (2014) Displaced dynamic expression regression for real-time facial tracking and animation. In: ACM transactions on graphics. Association for computing machinery, pp 1–10
130. Bouaziz S, Wang Y, Pauly M (2013) Online modeling for real-time facial animation. ACM Trans Graph 32:1–10. https://doi.org/10.1145/2461912.2461976
131. Garrido P, Valgaerts L, Sarmadi H et al (2015) VDub: modifying face video of actors for plausible visual alignment to a dubbed audio track. Comput Graph Forum 34:193–204. https://doi.org/10.1111/cgf.12552
132. Thies J, Zollhöfer M, Stamminger M, et al Face2Face: real-time face capture and reenactment of RGB videos
133. MIT Introduction to Deep Learning | 6.S191 - YouTube. https://www.youtube.com/watch?v=5tvmMX8r_OM. Accessed 8 Feb 2021
134. Garrido P, Valgaerts L, Wu C, Theobalt C (2013) Reconstructing detailed dynamic face geometry from monocular video. ACM Trans Graph 32:1–10. https://doi.org/10.1145/2508363.2508380
135. Viswanathan S, Heisters IES, Evangelista BP, et al. (2021) Systems and methods for generating augmented-reality makeup effects. U.S. Patent 10,885,697
136. Nam H, Lee J, Park JI (2020) Interactive Pixel-unit AR Lip Makeup System Using RGB Camera. J Broadcast Eng 25(7):1042–51
137. Siegl C, Lange V, Stamminger M, et al FaceForge: markerless non-rigid face multi-projection mapping
138. Face replacement in video using a still image and Face Tools - After Effects tutorial - YouTube. https://www.youtube.com/watch?v=x7T5jiUpUiE. Accessed 6 Feb 2021
139. Antipov G, Baccouche M and Dugelay JL, (2017), Face aging with conditional generative adversarial networks. In: IEEE international conference on image processing (ICIP), pp. 2089–2093
140. Shi C, Zhang J, Yao Y et al (2020) CAN-GAN: conditioned-attention normalized GAN for face age synthesis. Pattern Recognit Lett 138:520–526. https://doi.org/10.1016/j.patrec.2020.08.021
141. Fang H, Deng W, Zhong Y, Hu J (2020) Triple-GAN: Progressive face aging with triple translation loss. In: IEEE comput soc conf comput vis pattern recognit work 2020-June:3500–3509. https://doi.org/10.1109/CVPRW50498.2020.00410
142. Huang Z, Chen S, Zhang J, Shan H (2020) PFA-GAN: progressive face aging with generative adversarial network. IEEE Trans Inf Forensics Secur. https://doi.org/10.1109/TIFS.2020.3047753
143. Liu S, Li D, Cao T et al (2020) GAN-based face attribute editing. IEEE Access 8:34854–34867. https://doi.org/10.1109/ACCESS.2020.2974043
144. Yadav D, Kohli N, Vatsa M, et al (2020) Age gap reducer-GAN for recognizing age-separated faces. In: 25th international conference on pattern recognition (ICPR), pp 10090–10097
145. Sharma N, Sharma R, Jindal N (2020) An improved technique for face age progression and enhanced super-resolution with generative adversarial networks. Wirel Pers Commun 114:2215–2233. https://doi.org/10.1007/s11277-020-07473-1
146. Liu L, Yu H, Wang S et al (2021) Learning shape and texture progression for young child face aging. Sig Proc Image Commun 93:116127. https://doi.org/10.1016/j.image.2020.116127
147. Nirkin Y, Keller Y, Hassner T (2019) FSGAN: subject agnostic face swapping and reenactment. In: Proceedings of the IEEE/CVF international conference on computer vision, pp 7184–7193
148. Tripathy S, Kannala J, Rahtu E (2020) ICface: interpretable and controllable face reenactment using GANs. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision, pp 3385–3394
149. Ha S, Kersner M, Kim B, et al (2019) MarioNETte: few-shot face reenactment preserving identity of unseen targets. arXiv 34:10893–10900
150. Zhang J, Zeng † Xianfang, Wang M, et al (2020) FReeNet: multi-identity face reenactment. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp. 5326–5335.
151. Zeng X, Pan Y, Wang M, et al (2020) Realistic face reenactment via self-supervised disentangling of identity and pose. arXiv 34:12757–12764
152. Ding X, Raziei Z, Larson EC, et al (2020) Swapped face detection using deep learning and subjective assessment. EURASIP Journal on Information Security, pp 1–12
153. Zukerman J, Paglia M, Sager C, et al (2019) Video manipulation with face replacement. U.S. Patent 10,446,189
154. Hoshen D (2020) MakeupBag: Disentangling makeup extraction and application. arXiv preprint rXiv:2012.02157
155. Li Y, Huang H, Yu J, et al (2020) Cosmetic-aware makeup cleanser. arXiv preprint arXiv:2004.09147
156. Horita D, Aizawa K (2020) SLGAN: style- and latent-guided generative adversarial network for desirable makeup transfer and removal. arXiv preprint arXiv:2009.07557
157. Wu W, Zhang Y, Li C, et al (2018) ReenactGAN: learning to reenact faces via boundary transfer. In: Proceedings of the European conference on computer vision (ECCV), pp 603–619
158. Nirkin Y, Wolf L, Keller Y, Hassner T (2020) DeepFake detection based on the discrepancy between the face and its context. arXiv preprint arXiv:2008.12262.
159. Tolosana R, Vera-Rodriguez R, Fierrez J et al (2020) DeepFakes and beyond: a survey of face manipulation and fake detection. Inf Fusion 64:131–148
160. Shubham K, Venkatesh G, Sachdev R, et al (2020) Learning a deep reinforcement learning policy over the latent space of a pre-trained GAN for semantic age manipulation. In: 2021 international joint conference on neural networks (IJCNN), pp 1–8. IEEE
161. Karras T, Aila T, Laine S, Lehtinen J (2017) Progressive growing of GANs for improved quality, stability, and variation. arXiv preprint arXiv:1710.10196
162. Pham QTM, Yang J, Shin J (2020) Semi-supervised facegan for face-age progression and regression with synthesized paired images. Electron 9:1–16. https://doi.org/10.3390/electronics9040603
163. Zhu H, Huang Z, Shan H, Zhang J (2020) LOOK GLOBALLY , AGE LOCALLY : FACE AGING WITH AN ATTENTION MECHANISM Haiping Zhu Zhizhong Huang Hongming Shan Shanghai Key Lab of Intelligent Information Processing , School of Computer Science , Fudan University , China , 200433. ICASSP 2020 - 2020 IEEE Int Conf Acoust Speech Signal Process 1963–1967
164. Wu S, Rupprecht C, Vedaldi A (2021) Unsupervised learning of probably symmetric deformable 3D objects from images in

the wild. IEEE Trans Pattern Anal Mach Intell. https://doi.org/10.1109/TPAMI.2021.3076536
165. Heidekrueger PI, Juran S, Szpalski C et al (2017) The current preferred female lip ratio. J Cranio-Maxillofacial Surg 45:655–660. https://doi.org/10.1016/j.jcms.2017.01.038
166. Baudoin J, Meuli JN, di Summa PG et al (2019) A comprehensive guide to upper lip aesthetic rejuvenation. J Cosmet Dermatol 18:444–450
167. Garrido P, Zollhöfer M, Wu C et al (2016) Corrective 3D reconstruction of lips from monocular video. ACM Trans Graph 35:1–11. https://doi.org/10.1145/2980179.2982419
168. Wu C, Bradley D, Garrido P et al (2016) Model-based teeth reconstruction. ACM Trans Graph 35(6):220–221. https://doi.org/10.1145/2980179.2980233
169. Wen Q, Xu F, Lu M, Yong JH (2017) Real-time 3D eyelids tracking from semantic edges. ACM Trans Graph 36:1–11. https://doi.org/10.1145/3130800.3130837
170. Wang C, Shi F, Xia S, Chai J (2016) Realtime 3D eye gaze animation using a single RGB camera. ACM Trans Graph 35:1–14. https://doi.org/10.1145/2897824.2925947
171. Zhou X, Lin J, Jiang J, Chen S (2019) Learning a 3D gaze estimator with improved itracker combined with bidirectional LSTM. In: Proceedings - IEEE international conference on multimedia and expo. IEEE Computer Society, pp 850–855
172. Li H, Hu L, Saito S (2020) 3D hair synthesis using volumetric variational autoencoders. ACM Transactions on Graphics (TOG) 37(6):1–12
173. Ye Z, Li G, Yao B, Xian C (2020) HAO-CNN: filament-aware hair reconstruction based on volumetric vector fields. Comput Animat Virtual Worlds 31:e1945. https://doi.org/10.1002/cav.1945
174. He H, Li G, Ye Z et al (2019) Data-driven 3D human head reconstruction. Comput Graph 80:85–96. https://doi.org/10.1016/j.cag.2019.03.008