Chapter 1
An Introduction to Digital Face Manipulation

Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, Aythami Morales, and Javier Ortega-Garcia

Abstract Digital manipulation has become a thriving topic in the last few years, especially after the popularity of the term DeepFakes. This chapter introduces the prominent digital manipulations with special emphasis on the facial content due to their large number of possible applications. Specifically, we cover the principles of six types of digital face manipulations: *(i)* entire face synthesis, *(ii)* identity swap, *(iii)* face morphing, *(iv)* attribute manipulation, *(v)* expression swap (a.k.a. face reenactment or talking faces), and *(vi)* audio- and text-to-video. These six main types of face manipulation are well established by the research community, having received the most attention in the last few years. In addition, we highlight in this chapter publicly available databases and code for the generation of digital fake content.

1.1 Introduction

Traditionally, the number and realism of digital face manipulations have been limited by the lack of sophisticated editing tools, the domain expertise required, and the complex and time-consuming process involved [2–4]. For example, an early work in this topic [5] was able to modify the lip motion of a subject speaking using a

The present chapter is an updated adaptation of the journal article [1].

R. Tolosana (✉) · R. Vera-Rodriguez · J. Fierrez · A. Morales · J. Ortega-Garcia
Universidad Autonoma de Madrid, Madrid, Spain
e-mail: ruben.tolosana@uam.es

R. Vera-Rodriguez
e-mail: ruben.vera@uam.es

J. Fierrez
e-mail: julian.fierrez@uam.es

A. Morales
e-mail: aythami.morales@uam.es

J. Ortega-Garcia
e-mail: javier.ortega@uam.es

C. Rathgeb et al. (eds.), *Handbook of Digital Face Manipulation and Detection*,
Advances in Computer Vision and Pattern Recognition,
https://doi.org/10.1007/978-3-030-87664-7_1

different audio track, by making connections between the sounds of the audio track and the shape of the subject's face. However, from the original manual synthesis techniques up to now, many things have rapidly evolved. Nowadays, it is becoming increasingly easy to automatically synthesise non-existent faces or manipulate a real face (a.k.a. bonafide presentation [6]) of one subject in an image/video, thanks to: *(i)* the accessibility to large-scale public data and *(ii)* the evolution of deep learning techniques that eliminate many manual editing steps such as Autoencoders (AE) and Generative Adversarial Networks (GAN) [7, 8]. As a result, open software and mobile applications such as ZAO[1] and FaceApp[2] have been released opening the door to anyone to create fake images and videos, without any experience in the field.

In this context of digital face manipulation, there is one term that has recently dominated the panorama of social media [9, 10], becoming at the same time a great public concern [11]: DeepFakes.

In general, the popular term DeepFakes is referred to all digital fake content created by means of deep learning techniques [1, 12]. It was originated after a Reddit user named "deepfakes" claimed in late 2017 to have developed a machine learning algorithm that helped him to swap celebrity faces into porn videos [13]. The most harmful usages of DeepFakes include fake pornography, fake news, hoaxes, and financial fraud [14]. As a result, the area of research traditionally dedicated to general media forensics [15–18], is being invigorated and is now dedicating growing efforts for detecting facial manipulation in image and video [19, 20].

In addition, part of these renewed efforts in fake face detection are built around past research in biometric presentation attack detection (a.k.a. spoofing) [21–23] and modern data-driven deep learning [24–27]. Chapter 2 provides an introductory overview of face manipulation in biometric systems.

The growing interest in fake face detection is demonstrated through the increasing number of workshops in top conferences [28–32], international projects such as MediFor funded by the Defense Advanced Research Project Agency (DARPA), and competitions such as the Media Forensics Challenge (MFC2018)[3] launched by the National Institute of Standards and Technology (NIST), the Deepfake Detection Challenge (DFDC)[4] launched by Facebook, and the recent DeeperForensics Challenge.[5]

In response to those increasingly sophisticated and realistic manipulated content, large efforts are being carried out by the research community to design improved methods for face manipulation detection [1, 12]. Traditional fake detection methods in media forensics have been commonly based on: *(i)* in-camera, the analysis of the intrinsic "fingerprints" (patterns) introduced by the camera device, both hardware and software, such as the optical lens [33], colour filter array and interpolation [34, 35], and compression [36, 37], among others, and *(ii)* out-camera, the analysis of

[1] https://apps.apple.com/cn/app/id1465199127.

[2] https://apps.apple.com/gb/app/faceapp-ai-face-editor/id1180884341.

[3] https://www.nist.gov/itl/iad/mig/media-forensics-challenge-2018.

[4] https://www.kaggle.com/c/deepfake-detection-challenge.

[5] https://competitions.codalab.org/competitions/25228.

the external fingerprints introduced by editing software, such as copy-paste or copy-move different elements of the image [38, 39], reduce the frame rate in a video [40, 41], etc. Chapter 3 provides an in-depth literature review of traditional multimedia forensics before the deep learning era.

However, most of the features considered in traditional fake detection methods are highly dependent on the specific training scenario, being therefore not robust against unseen conditions [2, 16, 26]. This is of special importance in the era we live in as most media fake content is usually shared on social networks, whose platforms automatically modify the original image/video, for example, through compression and resize operations [19, 20].

This first chapter is an updated adaptation of the journal article presented in [1], and serves in this book as an introductory part of the most popular digital manipulations with special emphasis to the facial content due to the large number of possible harmful applications, e.g., the generation of fake news that would provide misinformation in political elections and security threats [42, 43], among others. Specifically, we cover in Sect. 1.2 six types of digital face manipulations: *(i)* entire face synthesis, *(ii)* identity swap, *(iii)* face morphing, *(iv)* attribute manipulation, *(v)* expression swap (a.k.a. face reenactment or talking faces), and *(vi)* audio- and text-to-video. These six main types of face manipulation are well established by the research community, receiving most attention in the last few years. Finally, we provide in Sect. 1.3 our concluding remarks.

1.2 Types of Digital Face Manipulations

1.2.1 Entire Face Synthesis

This manipulation creates entire non-existent face images. These techniques achieve astonishing results, generating high-quality facial images with a high level of realism for the observer. Fig. 1.1 shows some examples for entire face synthesis generated using StyleGAN. This manipulation could benefit many different sectors such as the video game and 3D-modelling industries, but it could also be used for harmful applications such as the creation of very realistic fake profiles on social networks in order to generate misinformation.

Entire face synthesis manipulations are created through powerful GANs. In general, a GAN consists of two different neural networks that contest with each other in a minimax game: the Generator G that captures the data distribution and creates new samples, and the Discriminator D that estimates the probability that a sample comes from the training data (real) rather than G (fake). The training procedure for G is to maximise the probability of D making a mistake, creating, therefore, high-quality fake samples. After the training process, D is discarded and G is used to create fake content. This concept has been exploited in the last years for the entire face synthesis, improving the realism of the manipulations as can be seen in Fig. 1.1.

One of the first popular approaches in this sense was ProGAN [44]. The key idea was to improve the synthesis process growing G and D progressively, i.e., starting from a low resolution, and adding new layers that model increasingly fine details as training progresses. Experiments were performed using the CelebA database [45], showing promising results for the entire face synthesis. The code of the ProGAN architecture is publicly available in GitHub.[6] Later on, Karras et al. proposed an enhanced version named StyleGAN [46] that considered an alternative G architecture motivated by the style transfer literature [47]. StyleGAN proposes an alternative generator architecture that leads to an automatically learned, unsupervised separation of high-level attributes (e.g., pose and identity when trained on human faces) and stochastic variation in the generated images (e.g., freckles, hair), and it enables intuitive, scale-specific control of the synthesis. Examples of this type of manipulations are shown in Fig. 1.1, using CelebA-HQ and FFHQ databases for the training of the StyleGAN [44, 46]. The code of the StyleGAN architecture is publicly available in GitHub.[7]

Finally, one of the prominent GAN approaches is StyleGAN2 [48], and StyleGAN2 with adaptive discriminator augmentation (StyleGAN2-ADA) [49]. Training a GAN using too little data typically leads to D overfitting, causing training to diverge. StyleGAN2-ADA proposes an adaptive discriminator augmentation mechanism that significantly stabilises training in limited data regimes. The approach does not require changes to loss functions or network architectures, and is applicable both when training from scratch and when fine-tuning an existing GAN on another dataset. The authors demonstrated that good results are possible to achieve by using only a few thousand training images. The code of the StyleGAN2-ADA architecture is publicly available in GitHub.[8]

Based on these GAN approaches, different databases are publicly available for research on the entire face synthesis manipulation. Table 1.1 summarises the main publicly available databases in the field, highlighting the specific GAN approach considered in each of them. It is interesting to remark that each fake image may be characterised by a specific GAN fingerprint just like natural images are identified by a device-based fingerprint (i.e., PRNU). In fact, these fingerprints seem to be dependent not only of the GAN architecture, but also to the different instantiations of it [50, 51].

In addition, as indicated in Table 1.1, it is important to note that public databases only contain the fake images generated using the GAN architectures. In order to be able to perform real/fake detection experiments on this digital manipulation group, researchers need to obtain real face images from other public databases such as CelebA [45], FFHQ [46], CASIA-WebFace [53], VGGFace2 [54], or MegaFace2 [55] among many others.

[6] https://github.com/tkarras/progressive_growing_of_gans.

[7] https://github.com/NVlabs/stylegan.

[8] https://github.com/NVlabs/stylegan2-ada-pytorch.

Table 1.1 Entire face synthesis: Publicly available databases

Database	Real images	Fake images
100K-Generated-Images (2019) [46]	–	100,000 (StyleGAN)
10K-Faces (2019) [52]	–	10,000 (–)
DFFD (2019) [24]	–	100,000 (StyleGAN) 200,000 (ProGAN)
iFakeFaceDB (2019) [26]	–	250,000 (StyleGAN) 80,000 (ProGAN)
100K-Generated-Images (2020) [48]	–	100,000 (StyleGAN2)
100K-Generated-Images (2020) [49]	–	100,000 (StyleGAN2-ADA)

Fig. 1.1 Real and fake examples of the **Entire face synthesis** manipulation group. Real images are extracted from http://www.whichfaceisreal.com/ and fake images from https://thispersondoesnotexist.com

We provide next a short description of each public database. In [46], Karras et al. released a set of 100,000 synthetic face images, named 100K-Generated-Images.[9] This database was generated using their proposed StyleGAN architecture, which was trained using the FFHQ dataset [46].

Another public database is 10K-Faces [52], containing 10,000 synthetic images for research purposes. In this database, contrary to the 100K-Generated-Images database, the network was trained using photos of models, considering face images from a more controlled scenario (e.g., with a flat background). Thus, no strange artefacts created by the GAN architecture are included in the background of the images. In addition, this dataset considers other interesting aspects such as ethnicity and gender diversity, as well as other metadata such as age, eye colour, hair colour and length, and emotion.

[9] https://github.com/NVlabs/stylegan.

(a) Fake (b) Fake after GANprintR

Fig. 1.2 Examples of a fake image created using StyleGAN and its improved version after removing the GAN-fingerprint information with GANprintR [26]

Recently, Dang et al. introduced in [24] a new database named Diverse Fake Face Dataset (DFFD).[10] Regarding the entire face synthesis manipulation, the authors created 100,000 and 200,000 fake images through the pre-trained ProGAN and StyleGAN models, respectively.

Neves et al. presented in [26] the iFakeFaceDB database. This database comprises 250,000 and 80,000 synthetic face images originally created through StyleGAN and ProGAN, respectively. As an additional feature in comparison to previous databases, and in order to hinder fake detectors, in this database, the fingerprints produced by the GAN architectures were removed through an approach named GANprintR (GAN fingerprint Removal), while keeping very realistic appearance. Figure 1.2 shows an example of a fake image directly generated with StyleGAN and its improved version after removing the GAN-fingerprint information. As a result of the GANprintR step, iFakeFaceDB presents a higher challenge for advanced fake detectors compared with the other databases.

Finally, we highlight the two popular 100K-Generated-Images public databases released by Karras et al. [48, 49], based on the prominent StyleGAN2 and StyleGAN2-ADA architectures. The corresponding fake databases trained using the FFHQ dataset [46] can be found in their GitHub.[11,12]

This section has described the main aspects of the entire face synthesis manipulation. For a complete understanding of fake detection techniques on this face manipulation, we refer the reader to Chap. 9.

1.2.2 Identity Swap

This manipulation consists of replacing the face of one subject in a video (source) with the face of another subject (target). Unlike the entire face synthesis, where manipulations are carried out at image level, in identity swap the objective is to generate realistic fake videos. Figure 1.3 shows some visual image examples extracted

[10] http://cvlab.cse.msu.edu/dffd-dataset.html.

[11] https://github.com/NVlabs/stylegan2.

[12] https://github.com/NVlabs/stylegan2-ada.

Fig. 1.3 Real and fake examples of the **Identity Swap** manipulation group. Face images are extracted from videos of Celeb-DF database [56]

from videos of Celeb-DF database [56]. In addition, very realistic videos of this type of manipulation can be seen on Youtube.[13] Many different sectors could benefit from this type of manipulation, in particular, the film industry.[14] However, on the other side, it could also be used for bad purposes such as the creation of celebrity pornographic videos, hoaxes, and financial fraud, among many others.

Two different approaches are usually considered for identity swap manipulations: *(i)* classical computer graphics-based techniques such as FaceSwap,[15] and *(ii)* novel deep learning techniques known as DeepFakes, e.g., the recent ZAO mobile application,[16] and the popular FaceSwap[17] and DeepFaceLab[18] software tools. In general, for each frame of the source video, the following stages are considered in the generation process of the identity swap video [57]: *(i)* face detection and cropping, *(ii)* extraction of intermediate representations, *(iii)* synthesis of a new face based on some driving signal (e.g., another face), and finally *(iv)* blending the generated face of the target subject into the source video, as shown in Fig. 1.3. For each of these stages, many possibilities could be considered to improve the quality of the fake videos. We

[13] https://www.youtube.com/watch?v=UlvoEW7l5rs.
[14] https://www.youtube.com/c/Shamook/featured.
[15] https://github.com/MarekKowalski/FaceSwap.
[16] https://apps.apple.com/cn/app/id1465199127.
[17] https://github.com/deepfakes/faceswap.
[18] https://github.com/iperov/DeepFaceLab.

Table 1.2 Identity swap: Publicly available databases

Database	Real videos	Fake videos
1st generation		
UADFV (2018) [58]	49 (Youtube)	49 (FakeApp)
DeepfakeTIMIT (2018) [11]	–	620 (faceswap-GAN)
FaceForensics++ (2019) [20]	1000 (Youtube)	1000 (FaceSwap) 1000 (DeepFake)
2nd generation		
DeepFakeDetection (2019) [66]	363 (Actors)	3068 (DeepFake)
Celeb-DF (2019) [56]	890 (Youtube)	5639 (DeepFake)
DFDC Preview (2019) [59]	1131 (Actors)	4119 (Multiple)
DFDC (2020) [67]	23,654 (Actors)	104,500 (Multiple)
DeeperForensics-1.0 (2020) [60]	50,000 (Actors)	1000 (DeepFake)
WildDeepfake (2020) [61]	3805 (Internet)	3509 (DeepFake)

describe next the main aspects considered in publicly available fake databases. For more details about the generation process, we refer the reader to Chaps. 4, and 14.

Since publicly available fake databases such as the UADFV database [58], up to the latest Celeb-DF, DFDC, DeeperForensics-1.0, and WildDeepfake databases [56, 59–61], many visual improvements have been carried out, increasing the realism of fake videos. As a result, identity swap databases can be divided into two different generations. Table 1.2 summarises the main details of each public database, grouped in each generation.

Three different databases are grouped in the first generation. UADFV was one of the first public databases [58]. This database comprises 49 real videos from Youtube, which were used to create 49 fake videos through the FakeApp mobile application,[19] swapping in all of them the original face with the face of Nicolas Cage. Therefore, only one identity is considered in all fake videos. Each video represents one individual, with a typical resolution of 294 × 500 pixels, and 11.14 s on average.

Korshunov and Marcel introduced in [11] the DeepfakeTIMIT database. This database comprises 620 fake videos of 32 subjects from the VidTIMIT database [62]. Fake videos were created using the public GAN-based face-swapping algorithm.[20] In that approach, the generative network is adopted from CycleGAN [63], using the weights of FaceNet [64]. The method Multi-Task Cascaded Convolution Networks is used for more stable detections and reliable face alignment [65]. Besides, the Kalman filter is also considered to smooth the bounding box positions over frames and eliminate jitter on the swapped face. Regarding the scenarios considered in DeepfakeTIMIT, two different qualities are considered: *(i)* low quality (LQ) with images of 64 × 64 pixels, and *(ii)* high quality (HQ) with images of 128 × 128

[19] https://www.malavida.com/en/soft/fakeapp/.

[20] https://github.com/shaoanlu/faceswap-GAN.

pixels. Additionally, different blending techniques were applied to the fake videos regarding the quality level.

One of the most popular databases is FaceForensics++ [20]. This database was introduced in early 2019 as an extension of the original FaceForensics database [68], which was focussed only on expression swap. FaceForensics++ contains 1000 real videos extracted from Youtube. Regarding the identity swap fake videos, they were generated using both computer graphics and DeepFake approaches (i.e., learning approach). For the computer graphics approach, the authors considered the publicly available FaceSwap algorithm[21] whereas for the DeepFake approach, fake videos were created through the DeepFake FaceSwap GitHub implementation.[22] The FaceSwap approach consists of face alignment, Gauss–Newton optimization and image blending to swap the face of the source subject to the target subject. The DeepFake approach, as indicated in [20], is based on two autoencoders with a shared encoder that is trained to reconstruct training images of the source and the target face, respectively. A face detector is used to crop and align the images. To create a fake image, the trained encoder and decoder of the source face are applied to the target face. The autoencoder output is then blended with the rest of the image using Poisson image editing [69]. Regarding the figures of the FaceForensics++ database, 1000 fake videos were generated for each approach. Later on, a new dataset named DeepFakeDetection, grouped inside the 2nd generation due to its higher realism, was included in the FaceForensics++ framework with the support of Google [66]. This dataset comprises 363 real videos from 28 paid actors in 16 different scenes. Additionally, 3068 fake videos are included in the dataset based on DeepFake FaceSwap GitHub implementation. It is important to remark that for both FaceForensics++ and DeepFakeDetection databases different levels of video quality are considered, in particular: *(i)* RAW (original quality), *(ii)* HQ (constant rate quantization parameter equal to 23), and *(iii)* LQ (constant rate quantization parameter equal to 40). This aspect simulates the video processing techniques usually applied in social networks.

Several databases have been recently released, including them in the 2nd generation due to their higher realism. Li et al. presented in [56] the Celeb-DF database. This database aims to provide fake videos of better visual qualities, similar to the popular videos that are shared on the Internet,[23] in comparison to previous databases that exhibit low visual quality for the observer with many visible artefacts. Celeb-DF consists of 890 real videos extracted from Youtube, and 5639 fake videos, which were created through a refined version of a public DeepFake generation algorithm, improving aspects such as the low resolution of the synthesised faces and colour inconsistencies.

Facebook in collaboration with other companies and academic institutions such as Microsoft, Amazon, and the MIT launched at the end of 2019 a new challenge named the Deepfake Detection Challenge (DFDC) [59]. They first released a preview dataset consisting of 1131 real videos from 66 paid actors, and 4119 fake videos.

[21] https://github.com/MarekKowalski/FaceSwap.

[22] https://github.com/deepfakes/faceswap.

[23] https://www.youtube.com/channel/UCKpH0CKltc73e4wh0_pgL3g.

Later on, they released the complete DFDC dataset comprising over 100K fake videos using 8 different face- swapping methods such as autoencoders, StyleGAN and morphable-mask models [67].

Another interesting database is DeeperForensics-1.0 [60]. The first version of this database (1.0) comprises 60K videos (50K real videos and 10K fake videos). Real videos were recorded in a professional indoor environment using 100 paid actors and ensuring variability in gender, age, skin colour, and nationality. Regarding fake videos, they were generated using a newly proposed end-to-end face-swapping framework based on Variational Autoencoders. In addition, extensive real-world perturbations (up to 35 in total) such as JPEG compression, Gaussian blur, and change of colour saturation were considered. All details of DeeperForensics-1.0 database, together with the corresponding competition, are described in Chap. 14.

Finally, Zi et al. presented in [61] WildDeepfake, a challenging real-world database for DeepFake detection. This database comprises 7314 videos (3805 and 3509 real and fake videos, respectively) collected completely from the internet. Contrary to previous databases, WildDeepfake claims to contain a higher diversity in terms of scenes and people in each scene, and also in facial expressions.

To conclude this section, we discuss at a higher level the key differences among fake databases of the 1st and 2nd generations. In general, fake videos of the 1st generation are characterised by: *(i)* low-quality synthesised faces, *(ii)* different colour contrast among the synthesised fake mask and the skin of the original face, *(iii)* visible boundaries of the fake mask, *(iv)* visible facial elements from the original video, *(v)* low pose variations, and *(vi)* strange artefacts among sequential frames. Also, they usually consider controlled scenarios in terms of camera position and light conditions. Many of these aspects have been successfully improved in databases of the 2nd generation, not only at visual level, but also in terms of variability (in-the-wild scenarios). For example, the recent DFDC database considers different acquisition scenarios (i.e., indoors and outdoors), light conditions (i.e., day, night, etc.), distances from the subject to the camera, and pose variations, among others. Figure 1.4 graphically summarises the weaknesses present in identity swap databases of the 1st generation and the improvements carried out in the 2nd generation. Finally, it is also interesting to remark the larger number of fake videos included in the databases of the 2nd generation.

This section has described the main aspects of the identity swap digital manipulation. For a complete understanding of the generation process and fake detection techniques, we refer the reader to Chaps. 4, 5, and 10–14.

Fig. 1.4 Graphical representation of the weaknesses present in **Identity Swap** databases of the 1st generation and the improvements carried out in the 2nd generation, not only at visual level, but also in terms of variability (in-the-wild scenarios). Fake images are extracted from: UADFV and FaceForensics++ (1st generation) [20, 58]; Celeb-DF and DFDC (2nd generation) [56, 59]

1.2.3 Face Morphing

Face morphing is a type of digital face manipulation that can be used to create artificial biometric face samples that resemble the biometric information of two or more individuals [70, 71]. This means that the new morphed face image would be successfully verified against facial samples of these two or more individuals creating a serious threat to face recognition systems [72, 73]. Figure 1.5 shows an example of the face morphing digital manipulation adapted from [70]. It is worth noting that face morphing is mainly focussed on creating fake samples at the image level, not

(a) Subject 1 (b) Face Morphing c) Subject 2

Fig. 1.5 Example for **a Face morphing** image (**b**) of subject 1 (**a**) and subject 2 (**c**). This figure has been adapted from [70]

video such as identity swap manipulations. In addition, as shown in Fig. 1.5, frontal view faces are usually considered.

There has been recently a large amount of research in the field of face morphing. Comprehensive surveys have been published in [70, 74] including both morphing techniques and also morphing attack detectors. In general, the following three consecutive stages are considered in the generation process of face morphing images: *(i)* determining correspondences between the face images of the different subjects. This is usually carried out by extracting landmark points, e.g., eyes, nose tips, mouth, etc.; *(ii)* the real face images of the subjects are distorted until the corresponding elements (landmarks) of the samples are geometrically aligned; and *(iii)* the colour values of the warped images are merged, referred to as blending. Finally, postprocessing techniques are usually considered to correct strange artefacts caused by pixel/region-based morphing [75, 76].

Prominent benchmarks have been recently presented in the field of face morphing. Raja et al. has recently presented an interesting framework in order to address serious open issues in the field such as independent benchmarking, generalizability challenges and considerations to age, gender, and ethnicity [77]. As a result, the authors have presented a new sequestered dataset and benchmark[24] for facilitating the advancements of morphing attack detection. The database comprises morphed and real images constituting 1800 photographs of 150 subjects. Morphing images are generated using 6 different algorithms, presenting a wide variety of possible approaches.

In this line, NIST has recently launched the FRVT MORPH evaluation.[25] This is an ongoing evaluation designed to obtain an assessment on morph detection capability with two separate tasks: *(i)* algorithmic capability to detect face morphing (morphed/blended faces) in still photographs, and *(ii)* face recognition algorithm resis-

[24] https://biolab.csr.unibo.it/fvcongoing/UI/Form/BenchmarkAreas/BenchmarkAreaDMAD.aspx.

[25] https://pages.nist.gov/frvt/html/frvt_morph.html.

tance against morphing. The evaluation is updated as new algorithms and datasets are added.

Despite these recent evaluations, we would like to highlight the lack of public databases for research. To the best of our knowledge, the only publicly available database is the AMSL Face Morph Image dataset[26] [78]. This is mainly produced due to most face morphing databases are created from existing face databases. As a result, the licenses can not be easily transferred which often prevents sharing.

This section has briefly described the main aspects of face morphing. For a complete understanding of the digital generation and fake detection techniques, we refer the reader to Chaps. 2, 6, 15, and 16.

1.2.4 Attribute Manipulation

This manipulation, also known as face editing or face retouching, consists of modifying some attributes of the face such as the colour of the hair or the skin, the gender, the age, adding glasses, etc. [79]. This manipulation process is usually carried out through GAN such as the StarGAN approach proposed in [80]. One example of this type of manipulation is the popular FaceApp mobile application. Consumers could use this technology to try on a broad range of products such as cosmetics and makeup, glasses, or hairstyles in a virtual environment. Figure 1.6 shows some examples for the attribute manipulation generated using FaceApp [81].

Despite the success of GAN-based frameworks for face attribute manipulations [80, 82–88], few databases are publicly available for research in this area, to the best of our knowledge. The main reason is that the code of most GAN approaches are publicly available, so researchers can easily generate their own fake databases as they like. Therefore, this section aims to highlight the latest GAN approaches in the field, from older to closer in time, providing also the link to their corresponding codes.

In [86], the authors introduced the Invertible Conditional GAN (IcGAN)[27] for complex image editing as the union of an encoder used jointly with a conditional GAN (cGAN) [89]. This approach provides accurate results in terms of attribute manipulation. However, it seriously changes the face identity of the subject.

Lample et al. proposed in [83] an encoder-decoder architecture that is trained to reconstruct images by disentangling the salient information of the image and the attribute values directly in the latent space.[28] However, as it happens with the IcGAN approach, the generated images may lack some details or present unexpected distortions.

[26] https://omen.cs.uni-magdeburg.de/disclaimer/index.php.

[27] https://github.com/Guim3/IcGAN.

[28] https://github.com/facebookresearch/FaderNetworks.

Fig. 1.6 Real and fake examples of the **Attribute Manipulation** group. Real images are extracted from http://www.whichfaceisreal.com/ and fake images are generated using FaceApp

An enhanced approach named StarGAN[29] was proposed in [80]. Before the StarGAN approach, many studies had shown promising results in image-to-image translations for two domains in general. However, few studies had focussed on handling more than two domains. In that case, a direct approach would be to build different models independently for every pair of image domains. StarGAN proposed a novel approach able to perform image-to-image translations for multiple domains using only a single model. The authors trained a conditional attribute transfer network via attribute-classification loss and cycle consistency loss. Good visual results were achieved compared with previous approaches. However, it sometimes includes undesired modifications from the input face image such as the colour of the skin.

Almost at the same time He et al. proposed in [82] attGAN,[30] a novel approach that removes the strict attribute-independent constraint from the latent representation, and just applies the attribute-classification constraint to the generated image to guarantee the correct change of the attributes. AttGAN provides state-of-the-art results on realistic attribute manipulation with other facial details well preserved.

One of the latest approaches proposed in the literature is STGAN[31] [84]. In general, attribute manipulation can be tackled by incorporating an encoder-decoder or GAN. However, as commented Liu et al. [84], the bottleneck layer in the encoder-decoder usually provides blurry and low quality manipulation results. To improve this, the authors presented and incorporated selective transfer units with an encoder-decoder for simultaneously improving the attribute manipulation ability and the image quality. As a result, STGAN has recently outperformed the state of the art in attribute manipulation.

Finally, we would like to highlight two recent attribute manipulation approaches that are currently achieving also very realistic visual results: RelGAN and SSC-

[29] https://github.com/yunjey/stargan/blob/master/README.md.

[30] https://github.com/LynnHo/AttGAN-Tensorflow.

[31] https://github.com/csmliu/STGAN.

GAN [90, 91]. RelGAN improves multi-domain image-to-image translation, whereas SSCGAN injects the target attribute information into multiple style skip connection paths between the encoder and decoder in order to incorporate global facial statistics.

Despite the fact that the code of the most attribute manipulation approaches are publicly available, the lack of public databases and experimental protocols results crucial when comparing among different manipulation detection approaches, otherwise it is not possible to perform a fair comparison among studies. Up to now, to the best of our knowledge, the DFFD database [24] seems to be the only public database that considers this type of facial manipulations. This database comprises 18,416 and 79,960 fake images generated through FaceApp and StarGAN approaches, respectively.

This section has briefly described the main aspects of the face attribute manipulation. For a complete understanding of this digital manipulation group, we refer the reader to Chap. 17.

1.2.5 Expression Swap

This manipulation, also known as face reenactment, consists of modifying the facial expression of the subject. Although different manipulation techniques are proposed in the literature, e.g., at image level through popular GAN architectures [84], in this group we focus on the most popular techniques Face2Face and NeuralTextures [92, 93], which replaces the facial expression of one subject in a video with the facial expression of another subject. Figure 1.7 shows some visual examples extracted from FaceForensics++ database [20]. This type of manipulation could be used with serious consequences, e.g., the popular video of Mark Zuckerberg saying things he never said.[32]

To the best of our knowledge, the only available database for research in this area is FaceForensics++ [20], an extension of FaceForensics [68].

Initially, the FaceForensics database was focussed on the Face2Face approach [93]. This is a computer graphics approach that transfers the expression of a source video to a target video while maintaining the identity of the target subject. This was carried out through manual keyframe selection. Concretely, the first frames of each video were used to obtain a temporary face identity (i.e., a 3D model), and track the expression over the remaining frames. Then, fake videos were generated by transferring the source expression parameters of each frame (i.e., 76 Blendshape coefficients) to the target video. Later on, the same authors presented in FaceForensics++ a new learning approach based on NeuralTextures [92]. This is a rendering approach that uses the original video data to learn a neural texture of the target subject, including a rendering network. In particular, the authors considered in their implementation a patch-based GAN-loss as used in Pix2Pix [94]. Only the facial

[32] https://www.bbc.com/news/technology-48607673.

Fig. 1.7 Real and fake examples of the **Expression Swap** manipulation group. Images are extracted from videos of FaceForensics++ database [20]

expression corresponding to the mouth was modified. It is important to remark that all data is available on the FaceForensics++ GitHub.[33] In total, there are 1000 real videos extracted from Youtube. Regarding the manipulated videos, 2000 fake videos are available (1000 videos for each considered fake approach). In addition, it is important to highlight that different video quality levels are considered, in particular: *(i)* RAW (original quality), *(ii)* HQ (constant rate quantization parameter equal to 23), and *(iii)* LQ (constant rate quantization parameter equal to 40). This aspect simulates the video processing techniques usually applied in social networks.

In addition to the Face2Face and NeuralTexture techniques considered in expression swap manipulations at video level, different approaches have been recently proposed to change the facial expression in both images and videos. A very popular approach was presented in [95]. Averbuch-Elor et al. proposed a technique to automatically animate a still portrait using a video of a different subject, transferring the expressiveness of the subject of the video to the target portrait. Unlike Face2Face and NeuralTexture approaches that require videos from both input and target faces, in [95] just an image of the target is needed. In this line, recent approaches have been presented achieving astonishing results in both one-shot and few-shot learning [96–98].

[33] https://github.com/ondyari/FaceForensics.

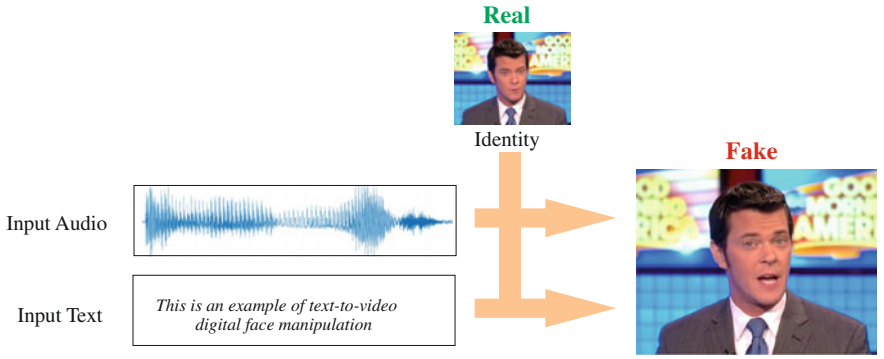

Fig. 1.8 Real and fake example of the **Audio-to-Video and Text-to-Video** face manipulation group

1.2.6 Audio-to-Video and Text-to-Video

A related topic to expression swap is the synthesis of video from audio or text. Figure 1.8 shows an example for the audio- and text-to-video face manipulation. These types of video face manipulations are also known as lip-sinc DeepFakes [99] or audio-driven facial reenactment [100]. Popular examples can be seen on the Internet.[34]

Regarding the synthesis of fake videos from audio (audio-to-video), Suwajanakorn et al. presented in [101] an approach to synthesise high-quality videos of a subject (Obama in this case) speaking with accurate lip sync. For this, they used as input to their approach many hours of previous videos of the subject together with a new audio recording. In their approach, they employed a recurrent neural network (based on Long Short-Term Memory, LSTM) to learn the mapping from raw audio features to mouth shapes. Then, based on the mouth shape at each frame, they synthesised high-quality mouth texture, and composited it with 3D pose alignment to create the new video to match the input audio track, producing photorealistic results.

In [102], Song et al. proposed an approach based on a novel conditional recurrent generation network that incorporates both image and audio features in the recurrent unit for temporal dependency, and also a pair of spatial-temporal discriminators for better image/video quality. As a result, their approach can model both lip and mouth together with expression and head pose variations as a whole, achieving much more realistic results. The source code is publicly available in GitHub.[35] Also, in [103], Song et al. presented a dynamic method not assuming a subject-specific rendering network like in [101]. In their approach, they are able to generate very realistic fake videos by carrying out a 3D face model reconstruction from the input video plus a recurrent network to translate the source audio into expression parameters. Finally,

[34] https://www.youtube.com/watch?v=VWMEDacz3L4.

[35] https://github.com/susanqq/Talking_Face_Generation.

they introduced a novel video rendering network and a dynamic programming method to construct a temporally coherent and photorealistic video. Video results are shown on the Internet.[36]

Another interesting approach was presented in [104]. Zhou et al. proposed a novel framework called Disentangled Audio-Visual System (DAVS), which generates high-quality talking face videos using disentangled audio-visual representation. Both audio and video speech information can be employed as input guidance. The source code is available in GitHub.[37]

Regarding the synthesis of fake videos from text (text-to-video), Fried et al. proposed in [105] a method that takes as input a video of a subject speaking and the desired text to be spoken, and synthesises a new video in which the subject's mouth is synchronised with the new words. In particular, their method automatically annotates an input talking-head video with phonemes, visemes, 3D face pose and geometry, reflectance, expression, and scene illumination per frame. Finally, a recurrent video generation network creates a photorealistic video that matches the edited transcript. Examples of the fake videos generated with this approach are publicly available.[38]

Finally, we would like to highlight the work presented in [100], named Neural Voice Puppetry. Thies et al. proposed an approach to synthesise videos of a target actor with the voice of any unknown source actor or even synthetic voices that can be generated utilising standard text-to-speech approaches, achieving astonishing visual results.[39]

To the best of our knowledge, there are no publicly available databases and benchmarks related to audio- and text-to-video fake detection content. Research on this topic is usually carried out through the synthesis of in-house data using publicly available implementations like the ones described in this section.

This section has briefly described the main aspects of the audio- and text-to-video face manipulation. For a complete understanding of this digital manipulation group, we refer the reader to Chap. 8.

1.3 Conclusions

This chapter has served as an introduction of the most popular digital face manipulations in the literature. In particular, we have covered six manipulation groups: *(i)* entire face synthesis, *(ii)* identity swap, *(iii)* face morphing, *(iv)* attribute manipulation, *(v)* expression swap (a.k.a. face reenactment or talking faces), and *(vi)* audio- and text-to-video. For each of them, we have described the main principles, publicly available databases, and code for the generation of digital fake content.

[36] https://wywu.github.io/projects/EBT/EBT.html.

[37] https://github.com/Hangz-nju-cuhk/Talking-Face-Generation-DAVS.

[38] https://www.ohadf.com/projects/text-based-editing/.

[39] https://justusthies.github.io/posts/neural-voice-puppetry/.

For more details about digital face manipulation and fake detection techniques, we refer the reader to Parts II and III of the present book. Finally, Part IV describes further topics, trends, and challenges in the field of digital face manipulation and detection.

Acknowledgements This work has been supported by projects: PRIMA (H2020-MSCA-ITN-2019-860315), TRESPASS-ETN (H2020-MSCA-ITN-2019-860813), BIBECA (MINECO/FEDER RTI2018-101248-B-I00), and COST CA16101 (MULTI-FORESEE).

References

1. Tolosana R, Vera-Rodriguez R, Fierrez J, Morales A, Ortega-Garcia J (2020) DeepFakes and beyond: a survey of face manipulation and fake detection. Inform Fusion 64:131–148
2. Farid H (2009) Image forgery detection. IEEE Signal Process Mag 26(2):16–25
3. Milani S, Fontani M, Bestagini P, Barni M, Piva A, Tagliasacchi M, Tubaro S (2012) An overview on video forensics. APSIPA Trans Signal Inform Process 1
4. Piva A (2013) An overview on image forensics. ISRN Signal Process
5. Bregler C, Covell M, Slaney M (1997) Video rewrite: driving visual speech with audio. Comput Graph 31(2):353–361
6. Information Technology-Biometric Presentation Attack Detection-Part 3: Testing and Reporting. Technical report, ISO/IEC JTC1 SC37 Biometrics (2017)
7. Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y (2014) Generative adversarial nets. In: Proceedings of advances in neural information processing systems
8. Kingma DP, Welling M (2013) Auto-encoding variational bayes. In: Proceedings of international conference on learning representations
9. Cellan-Jones R (2019) Deepfake videos double in nine months. https://www.bbc.com/news/technology-49961089
10. Citron D (2019) How DeepFake undermine truth and threaten democracy. https://www.ted.com
11. Korshunov P, Marcel S (2018) Deepfakes: a new threat to face recognition? Assessment and detection. arXiv preprint arXiv:1812.08685
12. Verdoliva L (2020) Media forensics and DeepFakes: an overview. IEEE J Sel Top Signal Process 14:910–932
13. BBC Bitesize: Deepfakes: what are they and why would i make one? (2019). https://www.bbc.co.uk/bitesize/articles/zfkwcqt
14. Kietzmann J, Lee LW, McCarthy IP, Kietzmann TC (2020) Deepfakes: trick or treat? Business Horizons 63(2):135–146
15. Korus P (2017) Digital image integrity-a survey of protection and verification techniques. Digital Signal Process 71:1–26
16. Rocha A, Scheirer W, Boult T, Goldenstein S (2011) Vision of the unseen: current trends and challenges in digital image and video forensics. ACM Comput Surv 43(4):1–42
17. Stamm M, Liu K (2010) Forensic detection of image manipulation using statistical intrinsic fingerprints. IEEE Trans Inform Forensics Secur 5(3):492–506
18. Swaminathan A, Wu M, Liu KJR (2008) Digital image forensics via intrinsic fingerprints. IEEE Trans Inform Forensics Secur 3(1):101–117
19. Cozzolino D, Rössler A, Thies J, Nießner M, Verdoliva L (2020) ID-Reveal: identity-aware DeepFake video detection. arXiv preprint arXiv:2012.02512
20. Rössler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Nießner M (2019) FaceForensics++: learning to detect manipulated facial images. In: Proceedinsg of IEEE/CVF international conference on computer vision

21. Galbally J, Marcel S, Fierrez J (2014) Biometric anti-spoofing methods: a survey in face recognition. IEEE Access 2:1530–1552
22. Hadid A, Evans N, Marcel S, Fierrez J (2015) Biometrics systems under spoofing attack: an evaluation methodology and lessons learned. IEEE Signal Process Mag
23. Marcel S, Nixon M, Fierrez J, Evans N (2019) Handbook of biometric anti-spoofing, 2nd edn
24. Dang H, Liu F, Stehouwer J, Liu X, Jain A (2020) On the detection of digital face manipulation. In: Proceedings of IEEE/CVF conference on computer vision and pattern recognition
25. Hernandez-Ortega J, Tolosana R, Fierrez J, Morales A (2021) DeepFakesON-Phys: DeepFakes detection based on heart rate estimation. In: Proceedings of 35th AAAI conference on artificial intelligence workshops
26. Neves JC, Tolosana R, Vera-Rodriguez R, Lopes V, Proença H, Fierrez J (2020) GANprintR: improved fakes and evaluation of the state of the art in face manipulation detection. IEEE J Sel Top Signal Process 14(5):1038–1048
27. Tolosana R, Romero-Tapiador S, Fierrez J, Vera-Rodriguez R (2021) DeepFakes evolution: analysis of facial regions and fake detection performance. In: Proceedings of international conference on pattern recognition workshops
28. Barni M, Battiato S, Boato G, Farid H, Memon N (2020) Multimedia forensics in the wild. In: International conference on pattern recognition. https://iplab.dmi.unict.it/mmforwild/
29. Biggio B, Korshunov P, Mensink T, Patrini G, Rao D, Sadhu A (2019) Synthetic realities: deep learning for detecting audio visual fakes. In: International conference on machine learning. https://sites.google.com/view/audiovisualfakes-icml2019/
30. Gregory S, Canton C, Leal-Taixé L, Bregler C, Farid H, Nießner M, Escalera S, Delp E, McCloskey S, Guyon I, Basharat A, Thies J, Verdoliva L, Escalante HJ, Scharfenberg C, Rössler A, Wan J, Cozzolino D, Guodong G (2020) Workshop on media forensics. In: Conference on computer vision and pattern recognition. https://sites.google.com/view/wmediaforensics2020/home
31. Raja K, Damer N, Chen C, Dantcheva A, Czajka A, Han H, Ramachandra R (2020) Workshop on Deepfakes and presentation attacks in biometrics. In: Winter conference on applications of computer vision. https://sites.google.com/view/wacv2020-deeppab/
32. Verdoliva L, Bestagini P (2019) Multimedia forensics. ACM Multimed. https://acmmm.org/tutorials/#tut3
33. Yerushalmy I, Hel-Or H (2011) Digital image forgery detection based on lens and sensor aberration. Int J Comput Vis 92(1):71–91
34. Cao H, Kot AC (2009) Accurate detection of demosaicing regularity for digital image forensics. IEEE Trans Inform Forensics Secur 4(4):899–910
35. Popescu AC, Farid H (2005) Exposing digital forgeries in color filter array interpolated images. IEEE Trans Signal Process 53(10):3948–3959
36. Chen YL, Hsu CT (2011) Detecting recompression of jpeg images via periodicity analysis of compression artifacts for tampering detection. IEEE Trans Inform Forensics Secur 6(2):396–406
37. Lin Z, He J, Tang X, Tang C (2009) Fast, automatic and fine-grained tampered JPEG image detection via DCT coefficient analysis. Pattern Recogn 42(11):2492–2501
38. Amerini I, Ballan L, Caldelli R, Bimbo A, Serra G (2011) A sift-based forensic method for copy-move attack detection and transformation recovery. IEEE Trans Inform Forensics Secur 6(3):1099–1110
39. Cozzolino D, Poggi G, Verdoliva L (2015) Splicebuster: a new blind image splicing detector. In: Proceedings of IEEE international workshop on information forensics and security, pp 1–6
40. Gironi A, Fontani M, Bianchi T, Piva A, Barni M (2014) A video forensic technique for detecting frame deletion and insertion. In: Proceedings of IEEE international conference on acoustics, speech and signal processing, pp 6226–6230
41. Wu Y, Jiang X, Sun T, Wang W (2014) Exposing video inter-frame forgery based on velocity field consistency. In: Proceedinsg of IEEE international conference on acoustics, speech and signal processing, pp 2674–2678

42. Allcott H, Gentzkow M (2017) Social media and fake news in the 2016 election. J Econ Perspect 31(2):211–236
43. Lazer DM, Baum MA, Benkler Y, Berinsky AJ, Greenhill KM, Menczer F, Metzger MJ, Nyhan B, Pennycook G, Rothschild D et al (2018) The science of fake news. Science 359(6380):1094–1096
44. Karras T, Aila T, Laine S, Lehtinen J (2018) Progressive growing of GANs for improved quality, stability, and variation. In: Proceedings of international conference on learning representations
45. Liu Z, Luo P, Wang X, Tang X (2015) Deep learning face attributes in the wild. In: Proceedings of IEEE/CVF international conference on computer vision
46. Karras T, Laine S, Aila T (2019) A style-based generator architecture for generative adversarial networks. In: Proceedings of IEEE/CVF conference on computer vision and pattern recognition
47. Huang X, Belongie S (2017) Arbitrary style transfer in real-time with adaptive instance normalization. In: Proceedings of IEEE/CVF international conference on computer vision
48. Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T (2020) Analyzing and improving the image quality of StyleGAN. In: Proceedings of IEEE/CVF conference on computer vision and pattern recognition
49. Karras T, Aittala M, Hellsten J, Laine S, Lehtinen J, Aila T (2020) Training Generative adversarial networks with limited data. arXiv preprint arXiv:2006.06676
50. Albright M, McCloskey S (2019) Source generator attribution via inversion. In: Proceedings of conference on computer vision and pattern recognition workshops
51. Marra F, Gragnaniello D, Verdoliva L, Poggi G (2019) Do GANs leave artificial fingerprints? In: Proceedings of IEEE conference on multimedia information processing and retrieval, pp 506–511
52. 100,000 faces generated by AI (2018). https://generated.photos/
53. Yi D, Lei Z, Liao S, Li S (2014) Learning face representation from scratch. arXiv preprint arXiv:1411.7923
54. Cao Q, Shen L, Xie W, Parkhi O, Zisserman A (2018) VGGFace2: a dataset for recognising faces across pose and age. In: Proceedings of international conference on automatic face & gesture recognition
55. Nech A, Kemelmacher-Shlizerman I (2017) Level playing field for million scale face recognition. In: Proceedings of IEEE/CVF conference on computer vision and pattern recognition
56. Li Y, Yang X, Sun P, Qi H, Lyu S (2020) Celeb-DF: a large-scale challenging dataset for DeepFake forensics. In: Proceedings of IEEE/CVF conference on computer vision and pattern recognition
57. Mirsky Y, Lee W (2021) The creation and detection of Deepfakes: a survey. ACM Comput Surv 54(1):1–41
58. Li Y, Chang M, Lyu S (2018) In Ictu Oculi: exposing AI generated fake face videos by detecting eye blinking. In: Proceedings of international workshop on information forensics and security
59. Dolhansky B, Howes R, Pflaum B, Baram N, Ferrer C (2019) The Deepfake detection challenge (DFDC) preview dataset. arXiv preprint arXiv:1910.08854
60. Jiang L, Wu W, Li R, Qian C, Loy CC (2020) DeeperForensics-1.0: a large-scale dataset for real-world face forgery detection. In: Proceedings of IEEE/CVF conference on computer vision and pattern recognition (2020)
61. Zi B, Chang M, Chen J, Ma X, Jiang YG (2020) WildDeepfake: a challenging real-world dataset for deepfake detection. In: Proceedings of ACM international conference on multimedia
62. Sanderson C, Lovell B (2009) Multi-region probabilistic histograms for robust and scalable identity inference. In: Proceedings of international conference on biometrics
63. Zhu J, Park T, Isola P, Efros A (2017) Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of international conference on computer vision (2017)

64. Schroff F, Kalenichenko D, Philbin J (2015) Facenet: a unified embedding for face recognition and clustering. In: Proceedings of IEEE/CVF conference on computer vision and pattern recognition
65. Zhang K, Zhang Z, Li Z, Qiao Y (2016) Joint face detection and alignment using multitask cascaded convolutional networks. IEEE Signal Process Lett 23(10):1499–1503
66. Google AI: Contributing data to Deepfake detection research (2019). https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html
67. Dolhansky B, Bitton J, Pflaum B, Lu J, Howes R, Wang M, Ferrer CC (2020) The DeepFake detection challenge dataset. arXiv preprint arXiv:2006.07397
68. Rössler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Nießner M (2018) FaceForensics: a large-scale video dataset for forgery detection in human faces. arXiv preprint arXiv:1803.09179
69. Pérez P, Gangnet M, Blake A (2003) Poisson image editing. ACM Trans Graph 22(3):313–318
70. Scherhag U, Rathgeb C, Merkle J, Breithaupt R, Busch C (2019) Face recognition systems under morphing attacks: a survey. IEEE Access 7:23012–23026
71. Wolberg G (1998) Image morphing: a survey. Vis Comput 14(8–9):360–372
72. Gomez-Barrero M, Rathgeb C, Scherhag U, Busch C (2017) Is your biometric system robust to morphing attacks? In: Proceedings of IEEE international workshop on biometrics and forensics
73. Korshunov P, Marcel S (2019) Vulnerability of face recognition to deep morphing. arXiv preprint arXiv:1910.01933
74. Venkatesh S, Raghavendra R, Raja K, Busch C (2021) Face morphing attack generation & detection: a comprehensive survey. IEEE Trans Technol Soc
75. Weng Y, Wang L, Li X, Chai M, Zhou K (2013) Hair interpolation for portrait morphing. Comput Graph Forum 32:79–84
76. Zhang H, Venkatesh S, Ramachandra R, Raja K, Damer N, Busch C (2021) MIPGAN-generating strong and high quality morphing attacks using identity prior driven GAN. arXiv preprint arXiv:2009.01729
77. Raja K, Ferrara M, Franco A, Spreeuwers L, Batskos I, de Wit F, Gomez-Barrero M, Scherhag U, Fischer D, Venkatesh S, Singh JM, Li G, Bergeron L, Isadskiy S, Ramachandra R, Rathgeb C, Frings D, Seidel U, Knopjes F, Veldhuis R, Maltoni D, Busch C (2020) Morphing attack detection-database. Evaluation platform and benchmarking, IEEE Trans Inform Forensics Secur
78. Neubert T, Makrushin A, Hildebrandt M, Kraetzer C, Dittmann J (2018) Extended StirTrace benchmarking of biometric and forensic qualities of morphed face images. IET Biometrics 7(4):325–332
79. Gonzalez-Sosa E, Fierrez J, Vera-Rodriguez R, Alonso-Fernandez F (2018) Facial soft biometrics for recognition in the wild: recent works, annotation and COTS evaluation. IEEE Trans Inform Forensics Secur 13(8):2001–2014
80. Choi Y, Choi M, Kim M, Ha J, Kim S, Choo J (2018) StarGAN: unified generative adversarial networks for multi-domain image-to-image translation. In: Proceedings of IEEE/CVF conference on computer vision and pattern recognition
81. FaceApp (2017). https://apps.apple.com/cn/app/id1465199127
82. He Z, Zuo W, Kan M, Shan S, Chen X (2019) AttGAN: facial attribute editing by only changing what you want. IEEE Trans Image Process
83. Lample G, Zeghidour N, Usunier N, Bordes A, Denoyer L, Ranzato M (2017) Fader networks: manipulating images by sliding attributes. In: Proceedings of advances in neural information processing systems
84. Liu M, Ding Y, Xia M, Liu X, Ding E, Zuo W, Wen S (2019) STGAN: a unified selective transfer network for arbitrary image attribute editing. In: Proceedings of IEEE/CVF conference on computer vision and pattern recognition (2019)
85. Li M, Zuo W, Zhang D (2016) Deep identity-aware transfer of facial attributes. arXiv preprint arXiv:1610.05586

86. Perarnau G, Weijer JVD, Raducanu B, Álvarez J (2016) Invertible conditional GANs for image editing. In: Proceedings of advances in neural information processing systems workshops
87. Shen W, Liu R (2017) Learning residual images for face attribute manipulation. In: Proceedings of conference on computer vision and pattern recognition
88. Xiao T, Hong J, Ma J (2018) ELEGANT: exchanging latent encodings with GAN for transferring multiple face attributes. In: Proceedings of European conference on computer vision
89. Mirza M, Osindero S (2014) Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784
90. Chu W, Tai Y, Wang C, Li J, Huang F, Ji R (2020) SSCGAN: facial attribute editing via style skip connections. In: Proceedings of European conference on computer vision
91. Wu PW, Lin YJ, Chang CH, Chang EY, Liao SW (2019) RelGAN: multi-domain image-to-image translation via relative attributes. In: Proceedings of IEEE/CVF international conference on computer vision
92. Thies J, Zollhöfer M, Nießner M (2019) Deferred neural rendering: image synthesis using neural textures. ACM Trans Graph 38(66):1–12
93. Thies J, Zollhofer M, Stamminger M, Theobalt C, Nießner M (2016) Face2face: real-time face capture and reenactment of RGB videos. In: Proceedings of IEEE/CVF conference on computer vision and pattern recognition
94. Isola P, Zhu J, Zhou T, Efros A (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings of conference on computer vision and pattern recognition
95. Averbuch-Elor H, Cohen-Or D, Kopf J, Cohen MF (2017) Bringing portraits to life. ACM Trans Graph 36(6):196
96. Ha S, Kersner M, Kim B, Seo S, Kim D (2020) Marionette: few-shot face reenactment preserving identity of unseen targets. In: Proceedings of AAAI conference on artificial intelligence
97. Siarohin A, Lathuilière S, Tulyakov S, Ricci E, Sebe N (2019) First order motion model for image animation. In: Proceedings of advances in neural information processing systems
98. Zakharov E, Shysheya A, Burkov E, Lempitsky V (2019) Few-shot adversarial learning of realistic neural talking head models. In: Proceedings of IEEE/CVF international conference on computer vision
99. Agarwal S, Farid H, Fried O, Agrawala M (2020) Detecting deep-fake videos from phoneme-viseme mismatches. In: Proceedings of workshop on media forensics, CVPRw
100. Thies J, Elgharib M, Tewari A, Theobalt C, Nießner M (2020) Neural voice puppetry: audio-driven facial reenactment. In: Proceedings of European conference on computer vision
101. Suwajanakorn S, Seitz S, Kemelmacher-Shlizerman I (2017) Synthesizing Obama: Learning Lip Sync From Audio. ACM Transactions on Graphics 36(4):1–13
102. Song Y, Zhu J, Li D, Wang A, Qi H (2019) Talking face generation by conditional recurrent adversarial network. In: Proceedings of international joint conference on artificial intelligence
103. Song L, Wu W, Qian C, He R, Loy C (2020) Everybody's Talkin': let me talk as you want. arXiv preprint arXiv:2001.05201
104. Zhou H, Liu Y, Liu Z, Luo P, Wang X (2019) Talking face generation by adversarially disentangled audio-visual representation. In: Proceedings of AAAI conference on artificial intelligence
105. Fried O, Tewari A, Zollhöfer M, Finkelstein A, Shechtman E, Goldman DB, Genova K, Jin Z, Theobalt C, Agrawala M (2019) Text-based editing of talking-head video. ACM Trans Graph 38(4)