

Article

Digital Face Manipulation Creation and Detection: A Systematic Review

Minh Dang^{1,2,*} and Tan N. Nguyen^{3,*} ¹ Institute of Research and Development, Duy Tan University, Da Nang 550000, Vietnam² Faculty of Information Technology, Duy Tan University, Da Nang 550000, Vietnam³ Department of Architectural Engineering, Sejong University, 209 Neungdong-ro, Gwangjin-gu, Seoul 05006, Republic of Korea

* Correspondence: danglienminh@duytan.edu.vn (M.D.); tnnguyen@sejong.ac.kr (T.N.N.)

† Minh Dang and Tan N. Nguyen are co-first authors. These authors contributed equally to this work.

Abstract: The introduction of publicly available large-scale datasets and advances in generative adversarial networks (GANs) have revolutionized the generation of hyper-realistic facial images, which are difficult to detect and can rapidly reach millions of people, with adverse impacts on the community. Research on manipulated facial image detection and generation remains scattered and in development. This survey aimed to address this gap by providing a comprehensive analysis of the methods used to produce manipulated face images, with a focus on deepfake technology and emerging techniques for detecting fake images. The review examined four key groups of manipulated face generation techniques: (1) attributes manipulation, (2) facial re-enactment, (3) face swapping, and (4) face synthesis. Through an in-depth investigation, this study sheds light on commonly used datasets, standard manipulated face generation/detection approaches, and benchmarking methods for each manipulation group. Particular emphasis is placed on the advancements and detection techniques related to deepfake technology. Furthermore, the paper explores the benefits of analyzing deepfake while also highlighting the potential threats posed by this technology. Existing challenges in the field are discussed, and several directions for future research are proposed to tackle these challenges effectively. By offering insights into the state of the art for manipulated face image detection and generation, this survey contributes to the advancement of understanding and combating the misuse of deepfake technology.



Citation: Dang, M.; Nguyen, T.N. Digital Face Manipulation Creation and Detection: A Systematic Review. *Electronics* **2023**, *12*, 3407. <https://doi.org/10.3390/electronics12163407>

Academic Editors: Soochang Park and Hosung Park

Received: 24 July 2023

Revised: 7 August 2023

Accepted: 8 August 2023

Published: 10 August 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: deepfake; image forensics; face manipulation; GAN; deep learning

1. Introduction

The dominance of cost-effective and advanced mobile devices, such as smartphones, mobile computers, and digital cameras, has led to a significant surge in multimedia content within cyberspace. These multimedia data encompass a wide range of formats, including images, videos, and audio. Fueling this trend, the dynamic and ever-evolving landscape of social media has become the ideal platform for individuals to effortlessly and quickly share their captured multimedia data with the public, contributing to the exponential growth of such content. A representative example of this phenomenon is Facebook, a globally renowned social networking site, which purportedly processes approximately 105 terabytes of data every 30 min and scans about 300 million photos each day (Source: <https://techcrunch.com/2012/08/22/how-big-is-facebooks-data-2-5-billion-pieces-of-content-and-500-terabytes-ingested-every-day/> (accessed on 10 December 2021)).

With the advent of social networking services (SNSs), there has been a remarkable increase in the demand for altering multimedia data, such as photos on Instagram or videos on TikTok, to attract a larger audience. In the past, the task of manipulating multimedia data was daunting for regular users, primarily due to the barriers posed by professional graphics editor applications like Adobe and the GNU Image Manipulation Program (GIMP), as

well as the time-consuming editing process. However, recent advancements in technology have significantly simplified the multimedia data manipulation process, yielding more realistic outputs. Notably, the rapid progress in deep learning (DL) technology has introduced sophisticated architectures, including generative adversarial networks (GANs) [1] and autoencoders (AEs) [2]. These cutting-edge techniques enable users to effortlessly create genuine faces with identities that do not exist or produce highly realistic video face manipulations without the need for manual editing.

AEs, which first emerged in 2017, are what the research community commonly refers to as deepfakes. AEs quickly gained attention when they were utilized to synthesize adult videos using the faces of famous Hollywood actors and politicians. Subsequently, a wave of face manipulation applications, such as FaceApp and FaceSwap, flooded the scene. To make the matter worse, the introduction of a smart undressing app called Deepnude in June 2019 sent shock waves across the world [3]. It has become increasingly challenging for regular users to filter out manipulated content, as multimedia data can spread like wildfire on the internet, leading to severe consequences like election manipulation, warmongering scenarios, and defamation. Moreover, the situation has worsened with the recent proliferation of powerful, advanced, and user-friendly mobile manipulation apps, including FaceApp [4], Snapchat [5], and FaceSwap [6], making it even more difficult to authenticate and verify the integrity of images and videos.

To address the escalating threat of progressively advancing and realistic manipulated facial images, the research community has dedicated substantial efforts to introducing innovative approaches that can efficiently and effectively identify signs of manipulated multimedia data [7]. The growing interest in digital face manipulation identification is evident in the increasing number of (1) papers at top conferences; (2) global research programs like Media Forensics (MediFor) backed by the Defense Advanced Research Project Agency (DARPA) [8]; and (3) global artificial intelligence (AI) competitions, such as the Deepfake Detection Challenge (DFDC) (<https://www.kaggle.com/c/deepfake-detection-challenge> (accessed on 10 December 2021)) organized by Facebook, the Open Media Forensics Challenge (OpenMFC) (<https://www.nist.gov/itl/iad/mig/open-media-forensics-challenge> (accessed on 10 December 2021)) backed by the National Institute of Standards and Technology (NIST), and the Trust Media Challenge launched by the National University of Singapore (<https://trustedmedia.aisingapore.org/> (accessed on 10 December 2021)).

Traditional approaches for identifying manipulated images commonly rely on camera and external fingerprints. Camera fingerprints refer to intrinsic fingerprints injected by digital cameras, while external fingerprints result from editing software. Previous manipulation detection methods based on camera fingerprints have utilized various properties such as optical lens characteristics [9], color filter array interpolation [10], and compression techniques [11]. On the other hand, existing manipulation detection approaches based on external fingerprints aim to identify copy-paste fingerprints in different parts of the image [12], frame rate reduction [13], and other features. While these approaches have achieved good performance, most of the features used in the training process are hand-crafted and heavily reliant on specific settings, making them less effective when applied to testing data in unseen conditions [14]. Currently, external fingerprints are considered more important than camera fingerprints due to the prevalence of manipulated media being uploaded and shared on social media sites, which automatically modify uploaded images and videos through processes such as compression and resizing operations [15].

In contrast to traditional approaches that heavily depend on manual feature engineering, DL entirely circumvents this process by automatically mapping and extracting essential features through a deep structure comprising multiple convolutional layers [16]. DL has demonstrated remarkable performance in image-related tasks [17], leading to a significant surge in the development of DL-based face manipulation detection models recently [18,19].

1.1. Relevant Surveys

Table 1 describes the key contributions of the latest comprehensive surveys that have examined various aspects of face manipulation. The increasing number of surveys in recent times reflects the growing interest and significance of the digital face manipulation topic.

Table 1. Summary of previous digital face manipulation reviews, including references, publication year, and primary contributions.

ID	Ref.	Year	Contributions
1	[20]	2022	<ul style="list-style-type: none"> Reviewed the latest deepfake creation and detection methods. Presented existing deepfake datasets. Explored the current challenges and trends in deepfake creation and detection.
2	[21]	2022	<ul style="list-style-type: none"> Analyzed existing deepfake generation methods and categorized them into five main categories. Explored the identification of deepfakes in face images and videos, considering different approaches and their performance. Described the current trends in deepfake datasets.
3	[22]	2021	<ul style="list-style-type: none"> Assessed the latest deepfake tools and techniques. Discussed the latest efficient strategies to address deepfake challenges. Explored the related challenges and future deepfake trends.
4	[23]	2021	<ul style="list-style-type: none"> Investigated the latest methods for generating and detecting deepfake images. Focused on summarizing and analyzing the architectures of deepfake creation and detection techniques. Presented future directions to enhance the architecture of deepfake models.
5	[14]	2021	<ul style="list-style-type: none"> Focused on showcasing the latest research on deepfake video detection and generation processes. Analyzed the robustness and generalization of deepfake video detection and generation models. Provided an overview of existing benchmarks for deepfake video creation.
6	[24]	2020	<ul style="list-style-type: none"> Provided a comprehensive review of standard image forgery methods. Described open data sources for image manipulation identification research. Focused on DL-based image manipulation identification.
7	[25]	2020	<ul style="list-style-type: none"> Offered a comprehensive review of current face manipulation methods, including deepfake, and strategies for identifying such manipulations. Divided existing facial manipulation techniques into four main groups. Reviewed publicly available datasets and crucial benchmarks for face manipulation and manipulated face detection.
8	[26]	2020	<ul style="list-style-type: none"> Conducted a comprehensive analysis of the identification of manipulated images and videos. Emphasized the emerging deepfake phenomenon from the forensic analyst's viewpoint. Highlighted the drawbacks of the latest forensic software, addressed urgent issues, explored upcoming challenges, and provided recommendations for future research directions.
9	[27]	2020	<ul style="list-style-type: none"> Defined deepfake and provided an overview of its fundamental technology. Categorized deepfake approaches and discussed the risks and opportunities associated with each group. Introduced a framework to address deepfake risks effectively.
10	[17]	2020	<ul style="list-style-type: none"> Reviewed the applications and developments of recent DL-based face synthesis and semantic manipulations. Discussed future perspectives to enhance face perception further.
11	[28]	2020	<ul style="list-style-type: none"> Differentiated and explained the relationships between image forgery, image manipulation, and image tampering. Provided motivations and explanations for various tampering detection methods. Explored standard benchmark datasets.

Since the term “deepfake” was introduced, deepfake technology has been prominently featured in recent headlines, leading to various studies on the topic of image manipulation. Notably, two recent studies discussed some of the latest developments in deepfake creation and detection [20,21]. In 2021, three surveys were released, with two of them focusing on the detection of deepfake content [14,22]. The remaining study conducted by Mirsky et al. [23] delved into the latest approaches to effectively generate and detect deepfake images. In 2020, two comprehensive surveys concerning general digital image manipulation were introduced [24,26]. While Thakur et al. explored DL-based image forgery detection [24], Verdoliva et al. conducted a thorough review of detecting manipulated images and videos [26]. In the same year, both Tolosana et al. and Kietzmann et al. offered insights into different aspects of deepfake [25,27], including the definition of deepfake [27] and publicly available deepfake datasets, along with crucial evaluation methods [25]. Additionally, Abdolahnejad et al. showcased the applications and advancements of current DL-based face synthesis methods in 2020 [17], while Zheng et al. analyzed the distinctive features of image forgery, image manipulation, and image tampering during the same year [28].

1.2. Contributions

The substantial number of studies on the face manipulation topic reflects a growing interest in the subject across various sectors and disciplines. This survey stands out from previous reviews due to the following distinct characteristics and contributions:

1. This review categorizes and discusses the latest studies that have achieved state-of-the-art results in face manipulation generation and detection.
2. Covering over 160 studies on the creation and detection of face manipulation, this survey categorizes them into smaller subtopics and provides a comprehensive discussion.
3. The review focuses on the emerging field of DL-based deepfake content generation and detection.
4. The survey uncovers challenges, addresses open research questions, and suggests upcoming research trends in the domain of digital face manipulation generation and detection.

The rest of the manuscript is organized as follows. Section 2 provides the essential background on face manipulation. Subsequently, Section 3 discusses various types of digital face manipulation and the associated datasets. Sections 4–7 delve into the critical aspects of each type of face manipulation, encompassing detection methods, identification techniques, and benchmark results. Moving forward, Section 8 elaborates on the evolution of deepfake technology. In Section 9, we shed light on open issues and future trends concerning face manipulation creation and detection. Finally, the paper ends with Section 10, presenting our concluding remarks.

1.3. Research Scope and Collection

This section addresses the scope of this work, the review methodology, and the process of selecting relevant papers.

1.3.1. Scope

The main scope of this survey is face manipulation, especially deepfake technology. Initially, previous face manipulation studies are categorized into four main categories: face swapping, facial re-enactment, face attribute manipulation, and face synthesis. Subsequently, explanations are provided for benchmark datasets, standard generation, and detection methods for each manipulation category. Finally, the survey discusses the challenges and trends for each category.

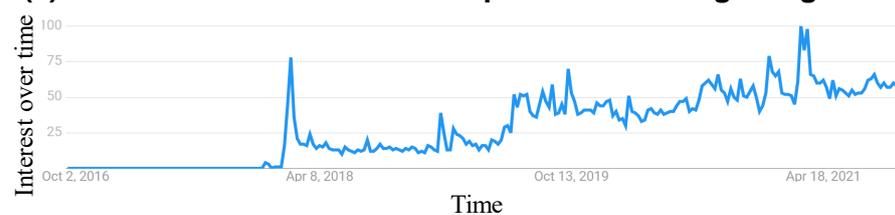
1.3.2. Research Collection

To efficiently and effectively collect related face manipulation papers, various sources were utilized for research paper collection. Initially, approximately 70 research papers of interest were gathered from two popular GitHub repositories on face manipulation

(<https://github.com/clpeng/Awesome-Face-Forgery-Generation-and-Detection> (accessed on 15 December 2021)) and the deepfake topic (<https://github.com/Billy1900/Awesome-DeepFake-Learning> (accessed on 15 December 2021)). Additionally, the latest deepfake publications from two well-known scientific portals, namely Google Scholar and Scopus, as well as arXiv, were queried using predefined keywords, including “DeepFake”, “fake face image”, “face-swapping”, “face synthesis”, “GAN”, “manipulated face”, “face forgery”, and “tampered face”. Furthermore, face manipulation is a prominent topic at top-tier conferences and workshops. Consequently, related publications from these events, published within the last three years, were also collected. It is noteworthy that the generation approaches were primarily identified based on the techniques mentioned in related detection studies for each category.

After completing the paper collection process, a total of 160 papers published between June 2008 and 2022 were downloaded from various scholarly search engines. The majority of these papers were related to the generation and detection of face manipulation, as well as the emergence of deepfake technology. As shown in Figure 1a, interest in face manipulation has seen a significant surge since the appearance of realistic fake videos towards the end of 2017. Figure 1b illustrates the distribution of research publications across various publication types, including papers from top references, journals, and other sources. Overall, the number of publications on the face manipulation topic has exhibited a stable growth over the years.

(a) Interest over time for the DeepFake term using Google trends



(b) Distribution of published scientific articles over time

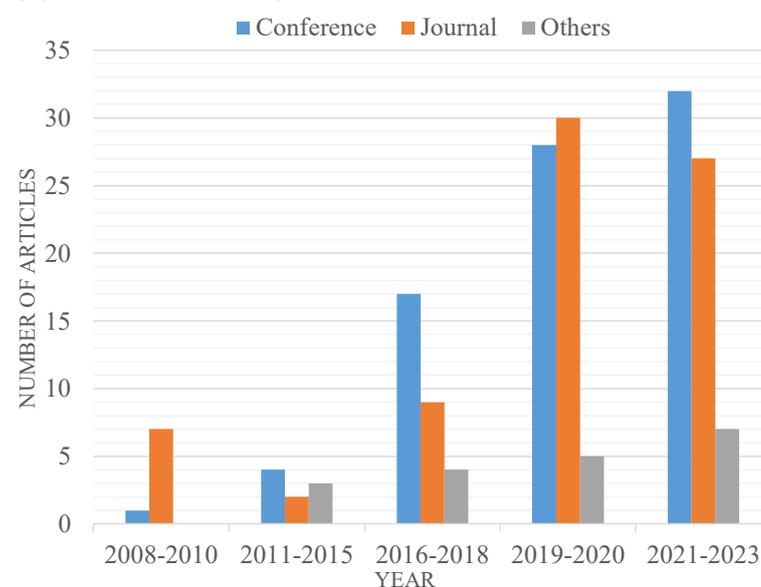


Figure 1. Evidence of the growing interest in the face manipulation topic is demonstrated through (a) Google trends for the keyword “face manipulation” and (b) the distribution of face-manipulation-related studies collected in this study over the years. Note: the Google interest over time is graphed relative to the highest point on the chart worldwide between October 2016 and October 2021.

In the realm of face manipulation generation methods, a substantial amount of research has originated from CVPR and arXiv. Approximately one-third of the downloaded publications were published at high-ranking conferences and in top journals, suggesting that most creation techniques underwent a rigorous review process, lending credibility to their findings. Similarly, for the face manipulation detection topic, the majority of articles were sourced from CVPR and other relevant conferences. In summary, in the face manipulation detection domain, a significant number of studies have been derived from arXiv and other sources, while research from top conferences and journals constitutes approximately one-third of the collected papers. Notably, as depicted in Figure 1b, around 65% of the studies published during the last three years have exhibited a growing research interest in face manipulation generation and detection topics.

2. Background

Image manipulation dates back to as early as 1860, when a picture of southern politician John C. Calhoun was realistically altered by replacing the original head with that of US President Abraham Lincoln [29]. In the past, image forgery was achieved through two standard techniques: image splicing and copy-move forgery, wherein objects were manipulated within an image or between two images [30]. To improve the visual appearance and perspective coherence of the forged image while eliminating visual traces of manipulation, additional post-processing steps, such as lossy JPEG compression, color adjustment, blurring, and edge smoothing, were implemented [31].

In addition to conventional image manipulation approaches, recent advancements in CV and DL have facilitated the emergence of various novel automated image manipulation methods, enabling the production of highly realistic fake faces [32]. Notably, hot topics in this domain include the automatic generation of synthetic images and videos using algorithms like GANs and AEs, serving various purposes, such as realistic and high-resolution human face synthesis [17] and human face attribute manipulation [33,34]. Among these, deepfake stands out as one of the trending applications of GANs, capturing significant public attention in recent years.

Deepfake is a technique used to create highly realistic and deceptive digital media, particularly manipulated videos and images, using DL algorithms [35]. The term “deepfake” is derived from the terms “deep learning” and “fake”. It involves using artificial intelligence, particularly deep neural networks, to manipulate and alter the content of an existing video or image by superimposing someone’s face onto another person’s body or changing their facial expressions [36]. Deepfake technology has evolved rapidly, and its sophistication allows for the creation of highly convincing fake videos that are challenging to distinguish from genuine footage. This has raised concerns about its potential misuse, as it can be employed for various purposes, including spreading misinformation, creating fake news, and fabricating compromising content [23,27]. For example, in May 2019, a distorted video of US House Speaker Nancy Pelosi was meticulously altered to deceive viewers into believing that she was drunk, confused, and slurring her words [37]. This manipulated video quickly went viral on various social media platforms and garnered over 2.2 million views within just two days. This incident served as a stark reminder of how political disinformation can be easily propagated and exploited through the widespread reach of social media, potentially clouding public understanding and influencing opinions.

Another related term, “cheap fake”, involves audio-visual manipulations produced using more affordable and accessible software [38]. These techniques include basic cutting, speeding, photoshopping, slowing, recontextualizing, and splicing, all of which alter the entire context of the message delivered in existing footage.

3. Types of Digital Face Manipulation and Datasets

3.1. Digitally Manipulated Face Types

Previous studies on digital facial manipulation can be classified into four primary categories based on the degree of manipulation. Figure 2 provides visual descriptions

of each facial manipulation category, ranging from high-risk to low-risk in terms of the potential impact on the public. The high risk associated with face swapping and facial re-enactment arises from the fact that malicious individuals can exploit these techniques to create fraudulent identities or explicit content without consent. Such concerns are rapidly increasing, and if left unchecked, they could lead to widespread abuse.

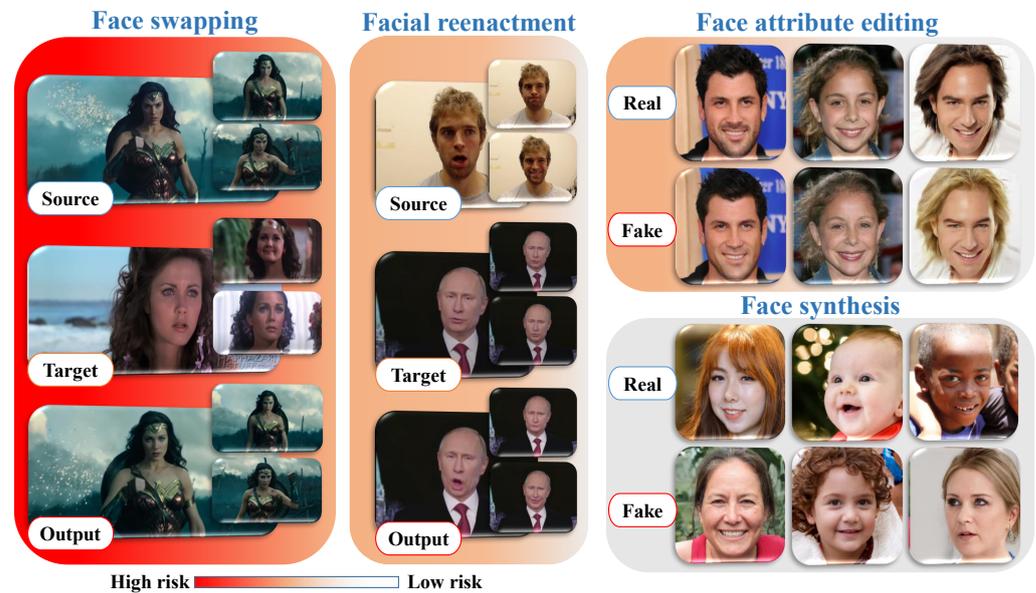


Figure 2. Four primary categories of face manipulation, including face swapping, facial re-enactment, face attribute editing, and face synthesis. Note: the gradient color bar on the bottom left of the image visualizes the risk levels based on the survey outcomes.

- **Face synthesis** encompasses a series of methods that utilize efficient GANs to generate human faces that do not exist, resulting in astonishingly realistic facial images. Figure 2 introduces various examples of entire face synthesis created using the PGGAN structure [39]. While face synthesis has revolutionized industries like gaming and fashion [40], it also carries potential risks, as it can be exploited to create fake identities on social networks for spreading false information.
- **Face swapping** involves a collection of techniques used to replace specific regions of a person's face with corresponding regions from another face to create a new composite face. Presently, there are two main methods for face swapping: (i) traditional CV-based methods (e.g., FaceSwap), and (ii) more sophisticated DL-based methods (e.g., deepfake). Figure 2 illustrates highly realistic examples of this type of manipulation. Despite its applications in various industrial sectors, particularly film production, face swapping poses the highest risk of manipulation due to its potential for malevolent use, such as generating pornographic deepfakes, committing financial fraud, and spreading hoaxes.
- **Face attribute editing** involves using generative models, including GANs and variational autoencoders (VAEs), to modify various facial attributes, such as adding glasses [33], altering skin color and age [34], and changing gender [33]. Popular social media platforms like TikTok, Instagram, and Snapchat feature examples of this manipulation, allowing users to experiment with virtual makeup, glasses, hairstyles, and hair color transformations in a virtual environment.
- **Facial re-enactment** is an emerging topic in conditional face synthesis, aimed at two main concurrent objectives: (1) transferring facial expressions from a source face to a target face, and (2) retaining the features and identity of the target face. This type of manipulation can have severe consequences, as demonstrated by the popular fake video of former US President Barack Obama speaking words that were not real [41].

3.2. Datasets

To generate fake images, researchers often utilize authentic images from public face datasets, including CelebA [34], FFHQ [42], CASIAWebFace [43], and VGGFace2 [44]. Essential details about each of these public datasets are provided in Table 2.

Table 2. Publicly available face datasets used for performing face image manipulation.

Name	Year	Source	Number of Images	Reference
VGGFace2	2018	Google Images	3.31 M	[44]
CelebFaces Attributes (CelebA)	2015	Internet	500 K	[34]
Flickr-Faces-HQ dataset (FFHQ)	2014	Flickr	70 K	[42]
CASIA-WebFace	2014	Internet	500 K	[43]
Labeled Faces in the Wild (LFW)	2007	Internet	13 K	[45]

Note: “M” indicates million, and “K” stands for thousand.

3.2.1. Face Synthesis and Face Attribute Editing

Despite the significant progress in GAN-based algorithms [33,46], to the best of our knowledge, few benchmark datasets are available for these topics. This scarcity is mainly attributed to the fact that most GAN frameworks can be easily re-implemented, as their codes are accessible online [47]. As a result, researchers can either download GAN-specific datasets directly or generate their fake datasets effortlessly.

Table 3 presents well-known publicly available datasets for studies on face synthesis and face attribute editing manipulation. Interestingly, each synthetic image is characterized by a specific GAN fingerprint, akin to the device-based fingerprint (fixed pattern noise) found in images captured by camera sensors. Furthermore, most of the mentioned datasets consist of synthetic images generated using GAN models. Therefore, researchers interested in conducting face synthesis generation experiments need to utilize authentic face images from other public datasets, such as VGGFace2 [44], FFHQ [42], CelebA [34], and CASIAWebFace [43].

In general, most datasets in the table are relevant because they are associated with well-known GAN frameworks like StyleGAN [48] and PGGAN [39]. In 2019, Karras et al. introduced the 100K-Generated-Images dataset [48], consisting of approximately 100,000 automatically generated face images using the StyleGAN structure applied to the FFHQ dataset [42]. The unique architecture of StyleGAN enabled it to automatically separate high-level attributes, such as pose and identity (human faces), while also handling stochastic variations in the created images, such as skin color, beards, hair, and freckles. This allowed the model to perform scale-specific mixing operations and achieve impressive image generation results.

Another publicly available dataset is 100K-Faces [49], comprising 100,000 synthesized face images created using the PGGAN model at a resolution of 1024 by 1024 pixels. Compared to the 100K-Generated-Images dataset, the StyleGAN model in the 100K-Faces dataset was trained using about 29,000 images from a controlled scenario with a simple background. This resulted in the absence of strange artifacts in the image backgrounds created by the StyleGAN model.

Recently, Dang et al. introduced the DFFD dataset [50], containing 200,000 synthesized face images using the pre-trained StyleGAN model [48] and 100,000 images using PGGAN [39]. Finally, the iFakeFaceDB dataset was released by Neves et al. [51], comprising 250,000 and 80,000 fake face images generated by StyleGAN [48] and PGGAN [39], respectively. An additional challenging feature of the iFakeFaceDB dataset is that GAN-printR [51] was used to eliminate the fingerprints introduced by the GAN architectures while maintaining a natural appearance in the images.

Table 3. Well-known public datasets for each type of face manipulation.

Type	Name	Year	Source	T:O	Dataset Size	Consent	Reference
Face synthesis and face attribute editing	Diverse Fake Face Dataset (DFFD)	2020	StyleGAN and PGGAN		300 K images	✓	[50]
	iFakeFaceDB	2020	StyleGAN and PGGAN		330 K images	✓	[51]
	100K-Generated-Images	2019	StyleGAN		100 K images	✓	[48]
	PGGAN	2018	PGGAN		100 K images	✓	[39]
	100K-Faces	2018	StyleGAN		100 K images	✓	[49]
Face swapping and facial re-enactment	ForgeryNIR	2022	CASIA NIR-VIS 2.0 [52]	1:4.1	50 K identities	×	[53]
	Face Forensics in the Wild (<i>FFIW</i> _{10K})	2021	YouTube	1:1	40 K videos	×	[54]
	OpenForensics	2021	Google open images	1:0.64	115 K BBox/masks	✓	[55]
	ForgeryNet	2021	CREMAD [56], RAVDESS [57], VoxCeleb2 [58], and AVSpeech [59]	1:1.2	221 K videos	✓	[60]
	Korean Deepfake Detection (KoDF)	2021	Actors	1:0.35	238 K videos	✓	[61]
	Deepfake Detection Challenge (DFDC)	2020	Actors	1:0.28	128 K videos	✓	[62]
	Deepfake Detection Challenge Preview (DFDC_P)	2020	Actors	1:0.28	5K videos	✓	[63]
	DeeperForensics-1.0	2020	Actors	1:5	60 K videos	×	[64]
	A large-scale challenging dataset for deepfake (Celeb-DF)	2020	YouTbe	1:0.1	6.2 K videos	×	[65]
	WildDeepfake	2020	Internet		707 videos	×	[66]
	FaceForensics++	2019	YouTube	1:4	5 K videos	Partly	[67]
	Google DFD	2019	Actors	1:0.1	3 K videos	✓	[68]
	UADFV	2019	YouTube	1:1	98 videos	×	[69]
	Media Forensic Challenge (MFC)	2019	Internet		100 K images 4 K videos	×	[70]
	Deepfake-TIMIT	2018	YouTube	Only fake	620 videos	×	[71]

Note: T:O indicates the ratio between tampered images and original images.

3.2.2. Face Swapping and Facial Re-Enactment

Table 3 presents a collection of datasets commonly used for conducting face swapping and facial re-enactment identification. Some small datasets, such as WildDeepfake [66], UADFV [69], and Deepfake-TIMIT [71], are early versions and contain less than 500 unique faces. For instance, the WildDeepfake dataset [66] consists of 3805 real face sequences and 3509 fake face sequences originating from 707 fake videos. The Deepfake-TIMIT database has 640 fake videos created using Faceswap-GAN [69]. Meanwhile, the UADFV dataset [71] contains 98 videos, with half of them generated by FakeAPP.

In contrast, more recent generations of datasets have exponentially increased in size. FaceForensics++ (FF++) [67] is considered the first large-scale benchmark for deepfake detection, consisting of 1000 pristine videos from YouTube and 4000 fake videos created by four different deepfake algorithms: deepfake [72], Face2Face [73], FaceSwap [74], and NeuralTextures [75]. The Deepfake Detection (DFD) [68] dataset, sponsored by Google, contains an additional 3000 fake videos, and video quality is evaluated in three categories: (1) RAW (uncompressed data), (2) HQ (constant quantization parameter of 23), and (3) LQ (constant quantization parameter of 40). Celeb-DF [65] is another well-known deepfake dataset, comprising a vast number of high-quality synthetic celebrity videos generated using an advanced data generation procedure.

Facebook introduced one of the biggest deepfake datasets, DFDC [62], with an earlier version called DFDC Preview (DFDC-P) [63]. Both DFDC and DFDC-P present significant challenges, as they contain various extremely low-quality videos. More recently, DeeperForensics1.0 [64] was published, modifying the original FF++ videos with a novel end-to-end face-swapping technique. Additionally, OpenForensics [55] was introduced as one of the first datasets designed for deepfake detection and segmentation, considering that most of the abovementioned datasets were proposed for performing deepfake classification. Figure 3 displays two sample images from each of the five well-known face synthesis datasets.



Figure 3. Comparison of fake images extracted from various deepfake datasets. From left to right, FaceForensics++ [67], DFDC [63], DeeperForensics [64], Celeb-DF [65], and OpenForensics [55].

While the number of public datasets has gradually increased due to advancements in face manipulation generation and detection, it is evident that Celeb-DF, FaceForensics++, and UADFV are currently among the most standard datasets. These datasets boast a vast collection of videos with varying categories that are appropriately formatted. However, there is a difference in the number of classes between the datasets. For example, the UADFV database is relatively simple and contains only two classes: pristine and fake. In contrast, the FaceForensics++ dataset is more complex, involving different types of video manipulation techniques and encompassing five main classes.

One common issue among existing deepfake datasets is that they were generated by splitting long videos into multiple short ones, leading to many original videos sharing

similar backgrounds. Additionally, most of these databases have a limited number of unique actors. Consequently, synthesizing numerous fake videos from the original videos may result in machine learning models struggling to generalize effectively, even after being trained on such a large dataset.

4. Face Synthesis

Entire face synthesis has emerged as an intriguing yet challenging topic in the fields of CV and ML. It encompasses a set of methods that aim to create photo-realistic images of faces that do not exist by utilizing a provided semantic domain. This technology has recently gained significance as a critical pre-processing phase for mainstream face recognition [71,76] and serves as an extraordinary testament to AI’s capabilities in utilizing complex probability distributions. GANs have become the most standard technology for performing face synthesis. GANs belong to a class of unsupervised learning models that utilize real human face images as training data and attempt to generate new face images from the same distribution.

Figure 4 illustrates an architecture-variant tree graph for GANs from 2014 to 2020, as described in this section. Notably, there are various interconnections between different GAN architectures. The relationship between GAN models and digital face manipulation has led to significant advancements in the quality and realism of generated manipulated face images. GAN models like DCGAN [1], WGAN [77], PGGAN [39], and StyleGAN [48] have been applied to facial attribute editing, deepfakes, face swapping, and facial re-enactment. For instance, cGAN allows controlled face image synthesis based on specific attributes, enabling targeted facial attribute editing [78]. CycleGAN enables unpaired image-to-image translation, facilitating face-to-face translation and facial attribute transfer without paired training data [79]. StyleGAN, on the other hand, produces high-quality manipulated face images with better control over specific features, resulting in realistic deepfake faces and facial re-enactment [48]. Additionally, advancements in GAN models and related techniques have contributed to a more comprehensive understanding of image manipulation methods, fostering the development of robust countermeasures to detect manipulated face images and ensure the integrity and trustworthiness of digital media [80].

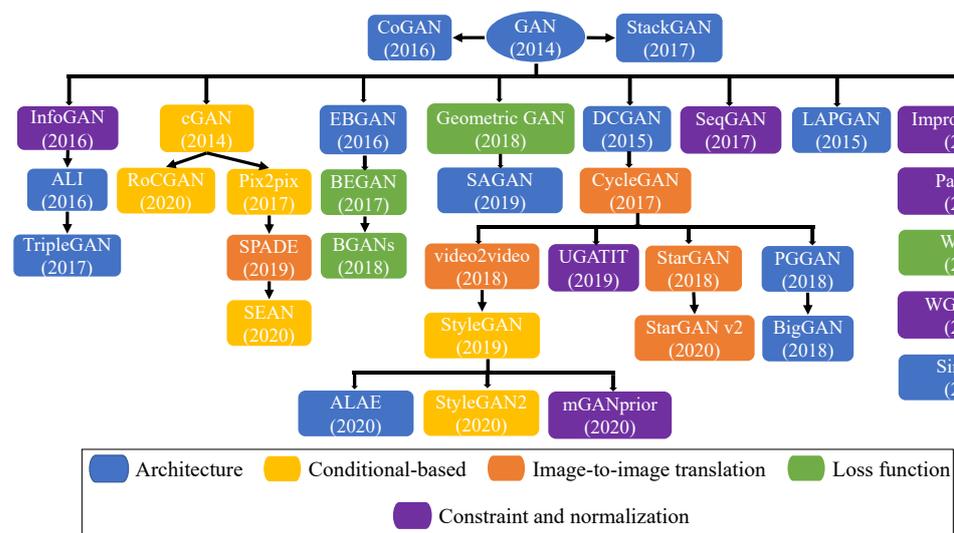


Figure 4. Tree graph depicting the relationships among various GAN models.

4.1. Generation Techniques

A generative adversarial network (GAN) is a type of DL model introduced by Ian Goodfellow and his colleagues in 2014 [81]. GANs consist of two neural networks, the generator

and the discriminator, which are trained simultaneously through a competitive process. Figure 5 illustrates the GAN process.

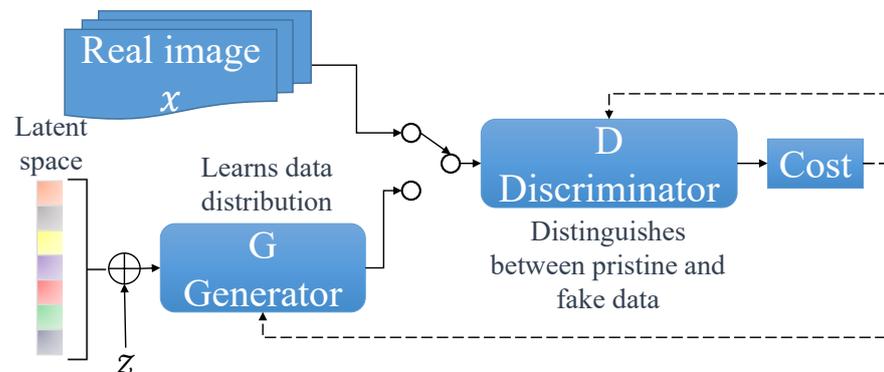


Figure 5. Key components of a generative adversarial network (GAN).

The generator's main task is to create realistic synthetic data, such as images or text, from random noise. It tries to generate data that are indistinguishable from real examples in the training dataset. On the other hand, the discriminator acts as a detective, attempting to distinguish between real and fake data provided by the generator. During the training process, the generator continuously improves its ability to produce more realistic data as it receives feedback from the discriminator. At the same time, the discriminator becomes more adept at accurately classifying real and fake data. The following equation describes the process:

$$\min_G \max_D V(D, G) = \mathbb{E}_{x \sim p_{\text{data}}(x)} [\log D(x)] + \mathbb{E}_{z \sim p_z(z)} [\log(1 - D(G(z)))] \quad (1)$$

where $V(D, G)$ represents the value of the objective function, which is the measure of how well the GAN is performing. The higher the value, the better the discriminator is at distinguishing real data from generated data, and, consequently, the better the generator is at producing realistic data. \mathbb{E} is the expected value of the logarithm of the discriminator's output when fed with real data samples (x) from the true data distribution $p_{\text{data}}(x)$. z represents random noise samples drawn from a prior distribution $p_z(z)$.

- Generative modeling in GANs aims to learn a generator distribution that closely resembles the target data distribution $p_{\text{data}(x)}$. Instead of explicitly assigning a probability to each variable x in the data distribution, GAN constructs a generator network G that creates samples by converting a noise variable $z \sim p_z(z)$ into a sample $G(z)$ from the generator distribution.
- Discrimination modeling in GANs seeks to generate an adversarial discriminator network D that can effectively discriminate between samples from the generator's distribution P_G and the true data distribution P_{data} . The discriminator's role is to distinguish between real data samples and those generated by the generator, thus creating a competitive learning dynamic.

The first extension of the GAN structure, which integrated a convolutional neural network (CNN) into a GAN, were deep convolutional GANs, or DCGANs [1]. Unlike previous GAN architectures, a DCGAN utilizes convolutional and transposed convolutional layers in the generator and discriminator, respectively, without fully connected and max-pooling layers. A DCGAN is an unsupervised learning approach that employs the discriminator's learned features as a feature extractor for a classification model. Its generator can easily manipulate various semantic properties in generated images using arithmetic vectors. DCGANs significantly improved image generation quality, which laid the groundwork for digital face manipulation tasks. Researchers used DCGANs to generate realistic faces, and their application was later extended to create manipulated faces by modifying specific attributes or facial features [82,83].

In the early generations of GANs, researchers faced the challenge of balancing the generator and discriminator during training to avoid mode dropping. Mode dropping refers to the issue where the generator fails to learn certain details from the evaluator, resulting in incomplete learning [84]. Therefore, many studies have been conducted to stabilize GAN training process [85,86]. One significant advancement in addressing this problem was the Wasserstein GAN (WGAN). The WGAN improves model stability during training by introducing a loss function that corresponds to the quality of the generated images [77]. Instead of using a discriminator to predict whether an image is real or fake, the WGAN replaces it with a critic that evaluates the realness or fakeness of an input image. The WGAN served as a foundation for various other studies. For example, WGAN gradient penalty (WGAN-GP) [77] addressed some issues of the WGAN, such as convergence failures and poor sample generation due to weight clipping. WGAN-GP focused on penalizing the norm of the gradient of the discriminator using the input. This technique demonstrated robust modeling performance and stability across different architectures.

One additional limitation of the initial GAN structures was their ability to generate only relatively low-resolution images [87], such as 64×64 or 128×128 pixels. As a result, several studies have proposed new architectures aimed at efficiently generating high-quality, detailed images of at least 1024×1024 pixels, making it challenging to distinguish between genuine and fake images. Notably, Progressive Growing of GANs (PGGAN) has gained recognition for introducing an efficient method to produce high-resolution outputs [39]. PGGAN achieves this by maintaining a symmetric mapping of the generator and discriminator during training and progressively adding layers to enhance the generator's image quality and the discriminator's capability. This approach accelerates the training process and significantly stabilizes the GAN. However, some output images may appear unrealistic during the PGGAN testing phase. Recently, BigGAN [88] aimed to create large, high-fidelity outputs by combining several current best approaches. It focused on generating class-conditional images and significantly increased the number of parameters and batch size to achieve its objectives.

4.2. Detection Techniques

In recent years, there have been a huge number of studies focusing on detecting face synthesis. Table 4 presents a comprehensive discussion of the most important research in this area, including crucial information such as datasets, methods, classifiers, and results. The best performances for each benchmark dataset are highlighted in bold, providing insights into the top-performing approaches.

It is worth noting that various evaluation metrics, such as the equal error rate (EER) and area under the curve (AUC), were utilized in these studies, which can sometimes make it challenging to directly compare different research findings. Additionally, some researchers have analyzed and leveraged the GAN structure itself to identify the artifacts between generated faces and genuine faces, further enhancing the detection capabilities.

Table 4. Comparison of current face synthesis identification studies.

Features	Classifier	Dataset	Performance	Reference
Frequency discrepancy	CNN	ProGAN	ACC = 90.7%	[89]
		StarGAN	ACC = 94.4%	
Frequency discrepancy	CNN	StyleGAN	ACC = 99.9%	[90]
		PGGAN	ACC = 97.4%	
Deep	CNN	Own (PCGAN and BEGAN)	ACC = 98.0%	[91]
Deep	CNN + attention mechanism	DFFD (ProGAN, StyleGAN)	AUC = 100% EER = 0.1%	[50]

Table 4. Cont.

Features	Classifier	Dataset	Performance	Reference
Deep	CNN	100K-Faces (StyleGAN)	EER = 0.3%	[51]
		iFakeFaceDB	EER = 4.5%	
Multi-level	CNN	Own (reconstruction)	ACC = 94%	[92]
Textural	CNN	StyleGAN	ACC = 95.51%	[93]
		PGGAN	ACC = 92.28%	
Layer-wise neuron	Support vector machine (SVM)	Own (InterFaceGAN, StyleGAN)	ACC = 84.7%	[94]
Steganalysis	CNN	100K-Faces (StyleGAN)	EER = 1.23%	[95]
Landmark	SVM	Own (PCGAN)	ACC = 94.13%	[96]
GAN	SVM	NIST MFC2018	AUC = 70.0%	[97]

Note: The best performance on each benchmark dataset is highlighted in bold. AUC indicates area under the curve, ACC is accuracy, and EER represents equal error rate.

As shown in Table 4, numerous face synthesis identification studies relied on conventional forensic-based methods. They examined datasets to detect differences between genuine images and those generated by GAN models, manually extracting these features. For example, McCloskey et al. demonstrated that the treatment of color between the camera sensor and GAN-generated images was distinguishable [97]. They extracted color features from the NIST MFC2018 dataset and used a support vector machine (SVM) model to classify whether an input image was real or fake, achieving the highest recorded performance of a 70% AUC. While these methods showed good performance on the collected dataset, they are susceptible to various attacks, such as compression, and may not scale well with larger training datasets, as often observed in shallow-learning-based models.

Another intriguing approach in this category was proposed by Yang et al. [96]. They focused on analyzing facial landmark points to identify GAN-generated images. The authors found that facial parts in images produced by GAN networks differed from those in the original images due to the loss of global constraints. Consequently, they extracted facial landmark features and fed them into an SVM model, achieving the highest accuracy of 94.13% on the PGGAN database. However, most of these approaches become ineffective when dealing with images generated by more advanced and complicated GANs. Moreover, they are sensitive to simple perturbation attacks such as compression, cropping, blurring, or noise. Consequently, the models need to be retrained for each specific scenario.

Currently, most face synthesis identification research is based on DL, harnessing the advantages of CNNs to automatically learn the differences between genuine and fake images. DL-based approaches have achieved state-of-the-art results on well-known publicly available datasets. For instance, Dang et al. utilized a lightweight CNN model to efficiently learn GAN-based artifacts for detecting computer-generated face images [91]. In another approach, Chen et al. customized the Xception model to locally detect GAN-generated face images [92]. They incorporated customized modules such as dilated convolution, a feature pyramid network (FPN), and squeeze and excitation (SE) to capture multi-level and multi-scale features of the local face regions. The experimental results demonstrated that their approach achieved higher performance compared to existing architectures that relied on the global features of the generated faces.

Inspired by steganalysis, Nataraj et al. introduced a detection system that combined pixel co-occurrence matrices and a CNN [95]. The robustness of the detection model was tested using fake images created by CycleGAN and StarGAN, achieving the highest equal error rate (EER) of 12.3%. Additionally, recent studies have shown that current GAN-based models struggle to reproduce high-frequency Fourier spectrum decay features, and researchers have utilized this characteristic to perform GAN-generated image identification. For instance, Jeong et al. achieved state-of-the-art detection performance on images created

by ProGAN and StarGAN models [89] with an accuracy of 90.7% and 94.4%, respectively. Moreover, attention mechanisms have been explored to enhance face manipulation detection during the training process. Dang et al. utilized attention mechanisms to evaluate the feature maps used for training the CNN model [50]. Through experiments, the authors demonstrated that the extracted attention maps accurately emphasized the GAN-generated artifacts. This approach achieved state-of-the-art performance compared to other methods, highlighting the potential of novel attention mechanisms in face manipulation detection.

Although the listed DL-based models have shown promising performance in controlled conditions, they face challenges in identifying unseen types of synthesis data. To address this, Marra et al. recently introduced a multi-task incremental learning manipulation identification approach that could identify unseen groups of images generated by a GAN without compromising the quality of the output images [98]. The model utilized representation learning (iCaRL) and incremental classifier techniques, allowing it to adapt to new classes by simultaneously learning classifiers and feature representations. The evaluation process involved assessing five well-known GAN models, and the XceptionNet-based model demonstrated accurate detection of new GAN-generated images. Finally, for completeness, we also include references to some crucial research related to detecting general GAN-generated image manipulations, not limited to face manipulation. Interested readers can refer to Mi et al. [99] and Marra et al. [100].

5. Face Attribute Editing

The investigation of face attributes, which represent various semantic features of a face such as “sex”, “hair color”, and “emotion”, has received significant attention due to its wide-ranging applications in fields like face recognition [101], facial image search [102], and expression parsing [103]. Altering face attributes, such as hairstyle, hair color, or the presence of a beard, can drastically change a person’s identity. For instance, an image captured during a person’s youth can appear entirely different from one taken during adulthood, all due to the “age” attribute. Consequently, face attribute editing, also known as face attribute manipulation, holds great importance for numerous applications.

For example, face attribute editing enables people to virtually design their fashion styles by visualizing how they would appear with different face attributes [104]. Furthermore, it enhances the robustness of face recognition by providing facial images with varying ages of a specific identity. Just like face synthesis, GAN architectures are commonly employed for face attribute editing. Some popular GAN models used for this purpose are StarGAN [105] and STGAN [106].

5.1. Generation Techniques

In recent years, GANs have emerged as the standard DL architecture for image generation. While previous GAN models could generate random plausible images for a given dataset, controlling specific properties of the output images proved challenging [107]. To address this limitation, conditional GANs (cGANs) were introduced as an extension of the GAN, incorporating information constraints to generate images based on conditional information y targeting a specific type [78]. For digital face manipulation, cGANs allowed the generation of manipulated face images with desired attributes or expressions, opening avenues for facial attribute editing and facial expression synthesis [108,109]. Among the efforts to perform GAN-based face attribute editing using cGANs, the Invertible cGAN (IcGAN) stands out [110]. IcGAN combines a cGAN with an encoder, where the encoder maps an original image to the latent space and a conditional representation to reconstruct and modify face attributes. While IcGAN demonstrates appropriate and robust face attribute manipulation, it completely alters a person’s identity. Additionally, IcGAN’s limitations include the inability to perform face attribute editing across more than two domains, requiring the processing of multiple models independently for each pair of image domains, making the process tedious and cumbersome [111].

To address the need for a more efficient face attribute manipulation approach, StarGAN [105] was proposed. StarGAN's structure comprised a generator and a discriminator that could be simultaneously trained on multiple different-domain databases, enabling image-to-image translations for various domains. The model incorporated a conditional attribute transfer model involving cycle consistency loss and attribute classification loss, resulting in visually impressive outcomes compared to previous studies. However, it could sometimes produce strange modifications, such as undesired skin color changes. To further address these challenges, a novel GANimation model was introduced, which anatomically encoded consistent face deformations using action units (AUs) [112]. This allowed the management of the magnitude of activation for each AU separately and the production of a more diverse set of facial expressions compared to StarGAN, which was limited by the dataset's content. StarGAN allowed the manipulation of multiple facial attributes in a single generator, which contributed to more versatile and efficient facial attribute editing [113,114].

An emerging trend in face attribute editing is real-time and interactive manipulation, with two representative GAN models leading the way: MaskGAN [115] and StyleMapGAN [116], both outperforming existing state-of-the-art GAN models. MaskGAN, a recent proposal by Lee et al. [115], is a novel face attribute editing model based on interactive semantic masks. It enables flexible face attribute manipulation while preserving high fidelity. The authors also introduced CelebAMask-HQ, a high-quality dataset containing over 30,000 fine-grained face mask annotations. In a different approach, Kim et al. introduced StyleMapGAN [116] with a novel spatial-dimensions-enabled intermediate latent space and spatially variant modulation. These improvements enhanced the embedding process while retaining the key properties of GANs.

Existing approaches have tried to discover a latent representation that is attribute-independent to perform extra face attribute editing, but this request for an attribute-independent representation is unreasonable, since facial attributes are inherently relevant [117]. This approach may cause information loss and limit the representation ability of attribute editing. To address this, AttGAN [118] proposed replacing the strict attribute-independent constraint with an attribute classification constraint on the latent representation, ensuring attribute manipulation accuracy. AttGAN can also directly control attribute intensity on various facial attributes. By incorporating the attribute classification constraint, adversarial learning, and reconstruction learning, AttGAN offers a robust face attribute editing framework. However, both StarGAN [105] and AttGAN [118] faced challenges when attempting arbitrary attribute editing using the target attribute vector as input, which could lead to undesired changes and visual degradation. As a solution, STGAN has been recently proposed [106], simultaneously enhancing attribute manipulation ability by carefully considering the gap between source and target attribute vectors as the model's input. Additionally, the introduction of a selective transfer unit significantly improves the quality of attribute editing, resulting in more realistic and accurate face attribute editing results.

Most of the existing research leverages GANs for face attribute editing, but they often alter irrelevant attribute parts of the face. To address this issue, recent works have been introduced. For example, Zhang et al. integrated a spatial attention component into a GAN to modify only the attribute-specific parts while keeping unrelated parts intact [46]. The experimental results on the CelebA and LFW databases demonstrated superior performance compared to previous face attribute editing approaches. In another approach, Xu et al. implemented region-wise style codes to decouple colors and texture, and 3D priors were used to separate pose, expression, illumination, and identity information [33]. Adversarial learning with an identity-style normalization component was then used to embed all the information. They also addressed disentanglement in GANs by introducing disentanglement losses, enabling the recognition of representations that can determine distinctive and informative aspects of data variations. With a different strategy, Collins et al. improved local face attribute alteration in the StyleGAN model by extracting latent object representations and using them during style interpolation [119]. The enhanced StyleGAN

model generated realistic attribute editing parts without introducing any artifacts. Recently, Xu et al. proposed a novel transformer-based controllable facial editing named TransEditor [120], which improved interaction in a dual-space GAN. The experimental results demonstrated that TransEditor outperformed baseline models in handling complicated attribute editing scenarios.

5.2. Detection Techniques

Originally, face attribute editing was primarily implemented for face recognition to assess the robustness of biometric systems against potential factors like occlusions, makeup, or plastic surgery [28]. However, with the advancement in image quality and resolution and the rapid growth of mobile applications supporting face attribute manipulation, such as FaceApp and TikTok, the focus has shifted towards detecting such manipulations. Table 5 provides a comprehensive analysis of the latest detection techniques for face attribute manipulation, including descriptions of datasets, approaches, classifiers, and performance for each study.

Table 5. Comparison of previous face attribute editing identification research.

Features	Classifier	Dataset	Performance	Reference
Steganalysis	CNN	StarGAN	AUC = 93.4%	[95]
Layer-wise neuron	SVM	StarGAN	ACC = 88%	[94]
		STGAN	ACC = 90%	
Expectation-maximization (EM)	CNN	StarGAN	ACC = 93.17%	[121]
		AttGAN	ACC = 92.67%	
		GDWCT	ACC = 88.4%	
		STYLEGAN2	ACC = 99.81%	
Deep	CNN + attention mechanism	FF++	AUC = 99.4% EER = 3.4%	[50]
Deep	CNN	ProGAN	ACC = 99%	[122]
		Adobe Photoshop	AUC = 74.9%	
Facial patches	CNN	ND-IIITD	ACC = 99.65%	[123]
		StarGAN	ACC = 99.83%	
Facial patches	Supervised restricted Boltzmann machine (SRBM)	ND-IIITD	ACC = 87.1%	[124]
		Celebrity	ACC = 96.2%	
Facial patches	Hierarchical CNN	StarGAN	ACC = 99.98%	[125]

Note: The highest performance on each public dataset is highlighted in bold. AUC indicates area under the curve, ACC is accuracy, and EER represents equal error rate.

Based on the same architecture used for entire face synthesis detection, Wang et al. demonstrated the significance of neuron behavior observation in detecting face attribute manipulation [94]. They found that neurons in the later layers captured more abstract features crucial for fake face detection. The study investigated this technique on authentic faces from the CelebA-HQ dataset and manipulated faces generated using StarGAN and STGAN, achieving the best manipulation detection accuracy of 88% and 90% for StarGAN and STGAN, respectively. Steganalysis features have also been explored for face attribute manipulation detection. As discussed in Section 4.2 for the face synthesis topic, Nataraj et al. proposed a GAN-generated image detection approach that extracted color channels from an

image and fed them into a CNN model [95]. They created a new face attribute manipulation dataset using the StarGAN model, achieving the highest classification accuracy of 93.4%.

Given the dominance of DL, many studies have focused on using DL models for attribute manipulation detection. Researchers have either extracted face patches or used entire faces and then fed them into CNN networks. For example, Bharati et al. introduced a robust face attribute manipulation detection framework by training a restricted Boltzmann machine (RBM) network [124]. Two manipulated datasets were created based on the ND-IIITD dataset and a collection of celebrity pictures from the internet. Face patches were extracted and fed into the RBM model to extract discriminative features for differentiating between pristine and fake images. The experimental results on the ND-IIITD and celebrity datasets were 87.1% and 96.2%, respectively. Similarly, Jain et al. utilized a CNN network with six convolutional layers and two fully connected layers to extract fundamental features for detecting synthetically altered images [123]. The proposed network achieved an altered image identification accuracy of 99.7% on the ND-IIITD database, outperforming existing models.

Furthermore, face attribute manipulation analysis using the entire face has been extensively explored in the literature, generally yielding high performance. Tariq et al. presented a DL-based GAN and human-created fake face image classifier that did not rely on any meta-data information [122]. Through an ensemble approach and various pre-processing methods, the model robustly identified GAN-based and manually retouched images with an accuracy of 99% and an AUC score of 74.9%, respectively.

Attention mechanisms have recently garnered significant attention as a hot research topic due to their potential to enhance the detection rate of face attribute editing when integrated with deep feature maps. Dang et al. incorporated attention mechanisms into a CNN model, revealing that the extracted attention maps effectively highlighted the critical regions (genuine face versus fake face) for evaluating the classification model and visualizing the manipulated regions [50]. The experimental results on the novel DFFD dataset demonstrated the system's high performance, achieving an EER close to 1.0% and a 99.9% AUC.

Another intriguing study by Guarnera et al. [121] focused on fake face detection using local features extracted from an expectation–maximization (EM) algorithm. This approach enabled the detection of manipulated images produced by different GAN models. The extracted features effectively addressed the underlying convolutional generative process, supporting the detection of fake images generated by recent GAN architectures, such as StarGAN [105], AttGAN [118], and StyleGAN2 [126]. With an average classification accuracy of over 90%, this work demonstrated that a model could identify images generated by various GANs without any prior knowledge of the GAN training process.

In summary, most face attribute manipulation research has relied on DL, achieving high detection performances of nearly 100% on various GAN-generated datasets and human-generated fake datasets, as presented in Table 5. This indicates that DL effectively learns the GAN-fingerprint artifacts present in fake images. However, as discussed in the face synthesis section, the latest GAN architectures have gradually eliminated such GAN fingerprints from the generated images and improved the fake images' quality simultaneously [127]. This advancement poses a new challenge for the most advanced manipulation detectors, calling for novel approaches to be explored.

6. Face Swapping

Face swapping is a popular technique in CV and image processing that involves replacing the face of one person in an image or video with the face of another. The process typically begins with detecting and aligning the facial features of both individuals to ensure proper positioning [128]. Then, the facial features of the target face are extracted and mapped onto the corresponding regions of the source face. This mapping is achieved through various techniques, such as landmark-based alignment or facial mesh deformation.

By blending the pixels of the two faces, a seamless transition is created, making it appear as though the target face is part of the original image or video.

The application of face swapping has gained widespread attention in recent years, especially in the realm of entertainment and social media [27]. It is commonly used for creating humorous or entertaining content, such as inserting a celebrity's face into a famous movie scene or swapping faces between friends in photos. However, its misuse has raised ethical concerns, as it can be used to create deepfake videos, where individuals' faces are superimposed onto explicit or false contexts, potentially leading to misinformation and harm [26]. As a result, researchers and developers have been working on developing advanced face-swapping detection techniques to combat the spread of fake or malicious content and protect the integrity of visual media.

6.1. Generation Techniques

Face swapping has a rich history in CV tasks, dating back over two decades. The face-swapping pipeline is depicted in Figure 6. The process begins by extracting frames from a source video, followed by localizing faces using a face detection technique. Once the faces are detected, they are converted into masks of the target face, while preserving the facial expressions. The original faces are then replaced with the synthesized facial masks, achieved through the precise alignment of the facial landmarks. Finally, additional trimming and post-processing methods are applied to the generated videos, aiming to enhance the overall output quality [129].

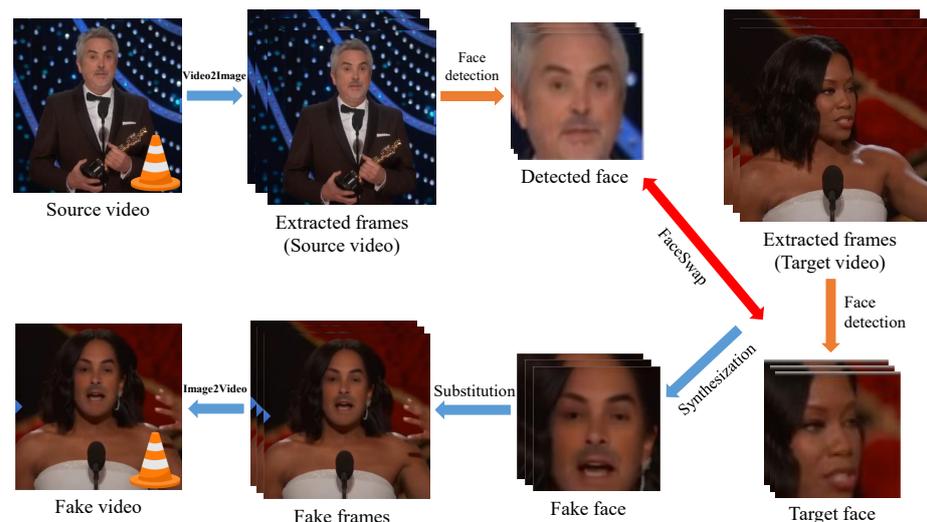


Figure 6. Complete pipeline for face swapping.

The face conversion process, which primarily relies on the AE structure, serves as a crucial component in the face-swapping generation pipeline. Training an AE model with an adequate amount of data from the target individual is essential to produce high-quality outputs and ensure efficient synthesis [130]. Beyond gathering training data, the overall face-swapping pipeline requires minimal human interference. The attractiveness of face swapping and deepfake lies in the quality of the generated fake faces and the simplicity of the generation process, which have garnered significant attention from the research community.

Over time, two generations of face-swapping approaches have emerged. The first generation, prevalent before 2019, was characterized by datasets such as FaceForensics++ [67]. These datasets featured low-resolution, poor-quality, and contrasting videos, resulting in poorly rendered face regions with noticeable GAN artifacts, such as misaligned landmarks, spliced region boundaries, inconsistent orientation, and mismatched skin colors across various videos. In contrast, the second generation of face-swapping datasets, exemplified by DFDC [63] and Celeb-DF [65], exhibited substantial improvements in the face-swapping generation procedure. These datasets presented highly realistic videos with

higher resolutions and fewer visual artifacts, leading to more convincing and visually appealing results.

6.1.1. Traditional Approach

The earliest face-swapping methods required various manual adjustments. For example, Bitouk et al. introduced an end-to-end face replacement framework that automatically replaced an input face with another face selected from a large collection of images [128]. The method involved face detection and alignment on the input image, followed by the selection of candidate face images based on pose and appearance similarity. The color, lighting, and pose of the selected face images were then updated to mimic the conditions of the original faces. However, this approach had two significant weaknesses: it altered the facial expression of the original face, and it lacked supervision over the identity of the output image.

Other works focused specifically on improving the post-processing processes after swapping faces. For instance, Xingjie et al. addressed the issue of the rigid boundary lines that occurred when there was a noticeable difference in color and brightness between the swapped face regions [131]. The authors proposed a novel blending approach using adaptive weight values to efficiently solve the problem with low computational cost. In a more recent work, Chen et al. combined various post-processing methods to enhance face-swapping performance [132]. They implemented a face parsing module to precisely detect the face regions of interest (ROIs) and applied a Poisson image editing module for color correction and boundary processing. The suggested framework was proven to significantly improve the overall performance of the face-swapping task.

The more challenging problem of replacing faces in videos was addressed by Dale et al. [133], as a sequence of images introduces extra complexities involving temporal alignment, evaluating face-swapping quality, and ensuring the temporal consistency of the output video. The framework employed by [133] was intricate, relying on a 3D multi-linear model to track facial quality and 3D geometry for source and target face warping. However, this complexity also resulted in longer processing times and required significant user guidance.

The studies mentioned above mainly relied on a complex multi-stage pipeline, involving face detection, tracking, alignment, and image compositing algorithms. Although they demonstrated acceptable outputs, which were sometimes indistinguishable from real photos, none of them fully addressed the aforementioned challenges.

6.1.2. Deep Learning Approach

During recent decades, DL-based face-swapping methods have demonstrated state-of-the-art performances. For instance, Perov et al. introduced DeepFaceLab, a novel model for performing high-quality face swapping [134]. The framework was flexible and provided a loose coupling structure for experts who wanted to customize the model without writing complex code from scratch. In 2019, Nirkin et al. presented a novel end-to-end recurrent neural network (RNN)-based system for subject-agnostic face swapping that accepted any pair of faces (image or video) without the need to collect data and train a specific model [135]. The proposed system efficiently preserved lighting conditions and skin color. However, it struggled with the texture quality of swapped faces when there was a large angular difference between the target and source faces. In another study, Li et al. introduced a two-stage FaceShifter system [129], where the second stage was trained in a supervised manner to address the face occlusion challenge by recovering abnormal areas.

Although these approaches showed good face-swapping performance, they heavily relied on the backbone architecture. For instance, the encoder of AE networks is responsible for extracting the target face's attributes and identity information so that the decoder can reconstruct the face using that information. This process poses a high risk of losing crucial information during feature extraction [2]. On the other hand, GAN-based structures often fail to deal with the temporal consistency issue and occasionally generate incoherent

outputs due to the backbone [98]. As a result, more recent work has been proposed to address these challenges. For instance, Liu et al. implemented a U-Net model in a novel neural identity carrier to efficiently extract fine-grained facial features across poses [136]. The proposed module prevented the loss of information in the decoder and generated realistic outputs. Taking a different approach, Zhu et al. introduced a novel one-shot face-swapping approach, including a hierarchical representation face encoder and a face transfer (FTM) module [137]. The encoder was implemented in an extended latent space to retain more facial features, while the FTM enabled identity transformation using a nonlinear trajectory without explicit feature disentanglement. Demonstrating an innovative idea, Naruniec et al. emphasized the importance of a multi-way network containing an encoder and multiple decoders to enable multiple face-swapping pairs in a single model [138]. The approach was proven to produce images with higher resolution and fidelity. In another approach, Xu et al. leveraged the learned knowledge of a pre-trained StyleGAN model to enhance the quality of the face-swapping process [139]. Moreover, a landmark-based structure transfer latent module was proposed to efficiently separate identity and pose information. Through extensive experiments, the model outperformed previous face-swapping models in terms of swapping quality and robustness.

Furthermore, DL-based face-swapping systems have demonstrated appealing performance. For example, Chen et al. introduced Simple Swap (SimSwap) [140], a DL-based face-swapping framework that effectively preserved the target face's attributes. The authors proposed a novel weak feature matching loss to enhance attribute preservation capability. Through various experiments, SimSwap demonstrated that it retained the target's face attributes significantly better than existing models. However, DL models often contain millions of parameters and require substantial computational power. Consequently, using them in real-world systems or installing them on smart devices, such as smartphones and tablets, becomes impractical.

To address this limitation, Xu et al. presented a lightweight subject-agnostic face-swapping model by incorporating two dynamic CNNs, including weight modulation and weight prediction, to automatically update the model parameters [141]. The model required only 0.33G floating-point operations per second (FLOPs) per frame, making it capable of performing face swapping on videos using smart devices in real time. This approach significantly reduced the computational burden, allowing for efficient and practical face-swapping applications on mobile devices without compromising performance

6.2. Detection Techniques

As the quality of generated face-swapped data has significantly improved, several techniques have been introduced to detect face swapping robustly and effectively. Table 6 presents different state-of-the-art face-swapping approaches. Currently, most face-swapping studies rely on DL and various kinds of features to identify face-swapped images robustly. Among these, deep features are extensively extracted and utilized due to their effectiveness. For example, Dang et al. combined attention maps with the original deep features to enhance face-swapping detection performance and enable the visualization of the manipulated regions [50]. This approach achieved the best accuracy of 99.4% on the DFFD database. Using a different strategy, Nguyen et al. introduced a CNN network based on a multi-task learning strategy to simultaneously classify manipulated images/videos and segment the fake regions for each query [142]. Recently, Zhou et al. [143] and Rossler et al. [67] focused on using steganalysis features, such as tampered artifacts, camera traits, and local noise residual evidence, along with deep features to improve tampered image detection accuracy. Following the current trend of transformer self-attention learning, Dong et al. proposed an identity consistency transformer face-swapping detection model to identify multiple traits of image degradation [144]. The transformer-based model outperformed previous baseline models and achieved the highest accuracy of 93.7% and 93.57% on the DFDC and DeeperForensics datasets, respectively.

Table 6. Face swapping: comparison of current detection approaches.

Features	Classifier	Dataset	Performance	Reference
Deep	Transformer	FF++	ACC = 98.56%	[144]
		DFDC	ACC = 93.7%	
		DeeperForensics	ACC = 93.57%	
Deep	AE (multi-task learning)	FaceSwap	ACC = 83.71%/EER = 15.7%	[142]
Deep	CNN (attention mechanism)	DFFD	ACC = 99.4%/EER = 3.1%	[50]
Deep	CNN	DFDC	Precision = 93%/Recall = 8.4%	[63]
Steganalysis and deep	CNN SVM	SwapMe	AUC = 99.5%	[143]
		FaceSwap	AUC = 99.9%	
Steganalysis and deep	CNN	FF++ (FaceSwap, LQ)	ACC = 93.7%	[67]
		FF+ (FaceSwap, HQ)	ACC = 98.2%	
		FF+ (FaceSwap, Raw)	ACC = 99.6%	
		FF+ (Deepfake, HQ)	ACC = 98.8%	
		FF+ (Deepfake, Raw)	ACC = 99.5%	
Facial	CNN	UADFV	ACC = 97.4%	[151]
		DeepfakeTIMIT (LQ)	ACC = 99.9%	
		DeepfakeTIMIT (HQ)	ACC = 93.2%	
Facial (source)	CNN	FF++ (Deepfake, -)	AUC = 100%	[145]
		FF++ (FaceSwap, -)	AUC = 100%	
		DFDC	AUC = 94.38%	
		Celeb-DF	AUC = 99.98%	
Facial	Multilayer perceptron (MLP)	Own	AUC = 85.1%	[152]
Facial	Face-cutout + CNN	DFDC	AUC = 92.71%	[146]
		FF++ (LQ)	AUC = 98.77%	
		Celeb-DF	AUC = 99.54%	

Table 6. Cont.

Features	Classifier	Dataset	Performance	Reference
Head pose	SVM	MFC	AUC = 84.3%	[69]
Temporal	CNN + RNN	FF++ (Deepfake, LQ)	ACC = 96.9%	[147]
		FF++ (FaceSwap, LQ)	ACC = 96.3%	
Temporal	CNN	DFDC (HQ)	AUC = 91%	[146]
		DCeleb-DF (HQ)	AUC = 84%	
Spatial and temporal	CNN	FF++ (LQ)	AUC = 90.9%	[148]
		FF++ (HQ)	AUC = 99.2%	
		DeeperForensics	AUC = 90.08%	
Spatial, temporal, and steganalysis	3D CNN	FF++	ACC = 99.83%	[149]
		DeepfakeTIMIT (LQ)	ACC = 99.6%	
		DeepfakeTIMIT (HQ)	ACC = 99.28%	
		Celeb-DF	ACC = 98.7%	
Textural	CNN	FF++ (HQ)	AUC = 99.29%	[153]

Note: The highest performance on each public dataset is highlighted in bold. AUC indicates area under the curve, ACC is accuracy, and EER represents equal error rate.

Facial features have proven to be an effective channel extensively used for detecting face forgeries. Zhao et al. discovered that the inconsistency of facial features in swapped images can be extracted to effectively train ConvNets [145]. Through extensive experiments, their model improved the average state-of-the-art AUC from 96.4% to 98.5%. Additionally, Das et al. proposed Face-Cutout, a data augmentation approach that removed random regions of a face using facial landmarks [146]. The dataset generated using this method allowed the model to focus on the input's important manipulated areas, enhancing model generalizability and robustness. Moreover, temporal features [146,147] and spatial features [148,149] have been extracted to detect face swapping in videos, enabling frame texture extraction and frame content reduction. For example, Trinh et al. recently proposed a lightweight 3D CNN, which fused spatial and temporal features, outperforming the latest deepfake detection studies [148]. Although previous face-swapping detection models achieved state-of-the-art results on benchmark datasets, interpreting the models' output remained challenging for experts. Dong et al. addressed this issue by applying various explainable algorithms to show how existing models detected face-swapping images without compromising the models' performance [150].

7. Facial Re-Enactment

7.1. Generation Techniques

Facial re-enactment, also known as facial expression transfer, is a fascinating technology that involves manipulating and modifying a target person's facial expressions in a video or image by transferring the facial movements from a source person's video or image [154]. The goal of facial re-enactment is to create realistic and convincing visual content in which the target person's face appears to express the same emotions or facial gestures as the source person's face.

This review focuses on two well-known facial re-enactment techniques: Face2Face [73] and NeuralTextures [75], both of which have demonstrated state-of-the-art results in facial re-enactment. To the best of our knowledge, the FaceForensics++ dataset [67] is currently the only benchmark dataset available for evaluating facial re-enactment methods, and it is an updated version of the original FaceForensics dataset [155].

Originally introduced in 2016, Face2Face gained popularity as a real-time facial re-enactment approach that transferred facial expressions from a source video to a target video without altering the target's identity [73]. The process began by reconstructing the face shape of the target subject using a model-based bundle adjustment algorithm. Then, the expressions of both the source and target actors were tracked using a dense analysis-by-synthesis method. The actual re-enactment took place using a deformation transfer module. Finally, the target's face, with transferred expression coefficients, was re-rendered, considering the estimated environment lighting of the target video's background. To synthesize a high-quality mouth shape, an image-based mouth synthesis method was implemented, which involved extracting and warping the most suitable mouth interior from available samples.

Later, the same authors presented NeuralTextures [75], which was a rendering framework that learned the neural texture of the target subject using original video data. The framework was trained using an adversarial loss and a photometric reconstruction loss. It incorporated deferred neural rendering, an end-to-end image synthesis paradigm that combined conventional graphics pipelines with learnable modules. Additionally, neural textures, which are extracted feature maps carrying richer information, were introduced and effectively utilized in the new deferred neural renderer. As a result of its end-to-end nature, the neural textures and deferred neural renderer enabled the system to generate highly realistic outputs, even when the inputs had poor quality.

Contrary to the Face2Face and NeuralTexture methods, which are mainly used for video-level facial re-enactment, several recent methods have been introduced to perform facial re-enactment in both images and videos. An influential method that automatically enlivens a still image to mimic the facial expressions of a driving video was introduced by

Averbuch-Elor et al. [156]. Unlike previous studies that typically require input source and target sequences for facial re-enactment, Averbuch-Elor’s method only needs a single target image. Due to the high potential for practical applications, multiple approaches have been recently presented, providing outstanding performance for both one-shot [157] and few-shot learning scenarios [158]. Lastly, with the smartphone revolution, many popular mobile applications have emerged, most of which rely on standard GAN models. For instance, FaceApp [4] supports changing facial expressions such as smiling, sadness, surprise, happiness, and anger.

7.2. Detection Techniques

To the best of our knowledge, research on facial re-enactment detection has primarily focused on videos. However, single image-level facial expression swap research can often be conducted using the methods discussed in Section 6.2. As a result, this section provides a comprehensive analysis of facial re-enactment detection at the video level, utilizing benchmark databases such as FaceForensics++ [67] and DFDC [62].

Table 7 presents a thorough comparison of various research on the facial re-enactment detection topic. For each work, essential information, including datasets, classifiers, features under consideration, and recorded performances, is provided. The highest results recorded for each dataset are highlighted in bold. It is essential to note that different evaluation metrics were used in some cases, making it challenging to conduct a direct comparison among the research. Furthermore, some studies mentioned in Section 6.2 for face swapping also conducted experiments on facial re-enactment. Therefore, in this section, the facial re-enactment results reported by those studies are discussed as well.

Table 7. Facial re-enactment: comparison of current state-of-the-art identification approaches.

Features	Classifier	Dataset	Performance	Reference
Deep	CNN	FF++ (NeuralTextures, HQ)	ACC = 94.2%	[159]
		FF++ (Face2Face, HQ)	ACC = 95.7%	
Deep	AE	FF++ (Face2Face, RAW)	ACC = 86.6%	[152]
Deep	CNN	DFDC	AUC = 99.4% EER = 3.4%	[50]
Deep	AE	FF++ (Face2Face, HQ)	AUC = 92.7% EER = 7.8%	[142]
Deep	Transformer	FF++ (LQ)	AUC = 94.2%	[160]
		FF++ (HQ)	AUC = 99.4%	
		FF++ (Raw)	AUC = 99.9%	
Steganalysis + deep	CNN	FF++ (NeuralTextures, LQ)	ACC = 82.1%	[67]
		FF+ (NeuralTextures, HQ)	ACC = 94.5%	
		FF++ (NeuralTextures, Raw)	ACC = 99.36%	
		FF++ (Face2Face, LQ)	ACC = 91.5%	
		FF+ (Face2Face, HQ)	ACC = 98.3%	
		FF++ (Face2Face, Raw)	ACC = 99.6%	
Frequency	CNN	FF++ (Face2Face, LQ)	AUC = 95.8%	[161]
		FF++ (NeuralTextures, LQ)	AUC = 86.1%	
Patch	CNN	FF++ (NeuralTextures, HQ)	ACC = 92%	[162]
		FF++ (Face2Face, HQ)	ACC = 99.6%	
Mesoscopic	CNN	FF++ (Face2Face, LQ)	ACC = 81.3%	[163]
		FF+ (Face2Face, HQ)	ACC = 93.4%	
		FF++ (Face2Face, Raw)	ACC = 96.8%	

Table 7. Cont.

Features	Classifier	Dataset	Performance	Reference
Facial	CNN	FF++ (Face2Face, LQ)	AUC = 91.2%	[164]
		FF+ (Face2Face, HQ)	AUC = 98.1%	
		FF++ (Face2Face, Raw)	AUC = 99.96%	
Facial (source)	CNN	FF++ (Face2Face, -)	AUC = 98.97%	[145]
		FF+ (NeuralTextures, -)	AUC = 97.63%	
Low-level texture	CNN	FF++ (Face2Face, LQ)	AUC = 94.6%	[165]
		FF++ (NeuralTextures, LQ)	AUC = 80.4%	
Temporal	CNN + RNN	FF++ (Face2Face, LQ)	ACC = 94.3%	[147]
Optical flow	CNN	FF++ (Face2Face, -)	AUC = 81.61%	[166]
Temporal + semantic	CNN	DFD (Facial re-enactment, LQ)	AUC = 90%	[167]
		DFD (Facial re-enactment, HQ)	AUC = 87%	

Note: The highest performance on each public dataset is highlighted in bold. AUC indicates area under the curve, ACC is accuracy, and EER represents equal error rate.

Previous research has primarily focused on observing artifacts in the face region of fake videos or specific texture features for facial re-enactment detection. For instance, Kumar et al. trained a CNN model using facial features, achieving state-of-the-art performance in detecting Face2Face manipulations [164] with nearly 100% accuracy on RAW quality. In another study, Liu et al. fed low-level texture features into a CNN model to detect facial re-enactment [165], showing significantly high performance on low-quality images with AUC scores of 94.6% and 80.4% on the FF++ (Face2Face) and FF++ (NeuralTexture) datasets, respectively. To boost facial re-enactment detection performance, Zhang et al. introduced a facial patch diffusion (PD) component that could be easily integrated into existing models [162].

Recently, research considering steganalysis and mesoscopic characteristics has gained attention. In [67], Rossler et al. proposed a framework based on steganalysis and deep features, achieving the highest accuracy of 99.3% and 99.6% on the RAW FF++ (NeuralTexture) and FF++ (Face2Face) datasets, respectively. Additionally, the framework was evaluated at various compression levels to verify the detection performance on different types of videos uploaded to social networks. Another approach by Afchar et al. [163] involved the extraction and use of mesoscopic features, resulting in a high accuracy of 96.8% on the RAW quality FF++ (Face2Face) dataset. These studies demonstrated the effectiveness of steganalysis and mesoscopic characteristics for robust facial re-enactment detection.

Attention mechanisms have gained increasing popularity in enhancing the training process and providing interpretability to a model's outputs. Dang et al. trained a CNN model that integrated attention mechanisms using the FF++ dataset [50], outperforming previous models with an impressive AUC performance of 99.4% on the DFFD database. While stacked convolutions have proven effective in achieving good detection performance by modeling local information, they are constrained in capturing global dependencies due to their limited receptive field. The recent success of transformers in CV tasks [168], which excel at modeling long-term dependencies, has inspired the development of various transformer-based facial re-enactment detection methods. For instance, Wang et al., 2021 proposed a multi-scale transformer that operated on patches of various sizes to capture local inconsistencies at different spatial levels [160]. This model achieved an impressive AUC value of nearly 100% on FF++ RAW and HQ quality datasets.

Another intriguing approach to facial re-enactment detection involves the analysis of both image and temporal information. For instance, Cozzolino et al. introduced a facial re-enactment identification system that leveraged temporal and semantic features [167].

The use of temporal facial features allowed the model to detect facial re-enactment without being explicitly trained on such manipulations, while high-level semantic features ensured the model's robustness against various post-processing techniques.

8. Deepfake Evolution

Figure 7 depicts the evolution of deepfake technology over the past few years. The first notable appearance of a deepfake occurred in September 2017, when a Reddit user named "deepfake" shared a collection of remarkably realistic computer-generated content featuring the faces of well-known celebrities superimposed onto pornographic videos. This incident marked the beginning of public awareness regarding the potential misuse and implications of deepfake technology. Subsequently, the introduction of DeepNude software further fueled public concern as it allowed the creation of highly realistic fake nude images of individuals, contributing to the rapid spread of deepfake-related issues.

Since then, numerous highly realistic deepfake videos have emerged, featuring politicians like Russia's President Vladimir Putin [169] and former US President Barack Obama [45], as well as influencers such as Elon Musk [170] and David Beckham [171]. Even famous celebrities like Tom Cruise [172] have not been immune to this technology's manipulations, underscoring the challenges posed by deepfake and their potential to deceive and mislead viewers.

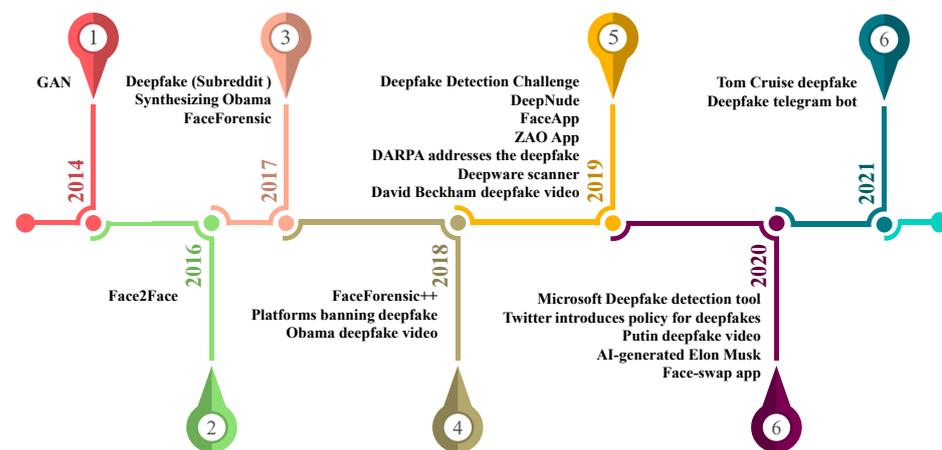


Figure 7. Timeline of deepfake evolution.

Nowadays, deepfake is increasingly carried out on smartphone apps or open-source programs and shared quickly on social networks with just a few simple steps. Table 8 reveals some well-known applications.

Deepfake technology and applications, such as FaceSwap [174], FaceApp [4], and deepfake bot [188], have become more easily accessible, and users with a mobile phone can quickly produce a fake video. In addition, related tutorials and open-source projects on GitHub, such as DeepFaceLab [183], Deepware scanner [184], and Generated Photos [187], are readily available. During the last few years, major tech companies, such as Microsoft [189], Twitter [190,191], and Facebook, have carried out various measures in order to prevent the spread of deepfake.

In response to the growing threat of deepfake, major tech companies like Microsoft [189], Twitter [190,191], and Facebook have taken various measures to curb their dissemination. These efforts aim to prevent the widespread circulation of potentially harmful or deceptive content, as deepfake technology poses significant challenges to the authenticity and trustworthiness of digital media.

Table 8. Overview of deepfake creation software, mobile apps, and open-source projects.

Type	Name	Face Swapping	Re-enactment	Attribute Editing	Face Synthesis	Reference	Note
CS	Adobe Photoshop	✓		✓		[173]	Commercial application
	Faceswap	✓				[174]	For learning and training purposes Faster with GPU
MA	FaceApp			✓		[4]	
	Xpression		✓			[175]	
	Reface			✓		[176]	Simple process Relies on face embeddings
	Impressions	✓				[177]	
	Myheritage		✓			[178]	Animates historical photos
	Wombo		✓			[179]	Ensures users' privacy
WA	Reflect	✓				[180]	
	Deepfake WEB	✓	✓			[181]	Learns various features of the face data Takes hours to render the data
OS	Faceswap-GAN	✓				[182]	
	DeepFaceLab	✓	✓	✓		[183]	Advanced application Needs a powerful PC
	Deepware scanner	✓	✓	✓		[184]	Supports API Hosted website
	Face2Face		✓			[185]	
	Dynamixyz		✓			[186]	
	Generated Photos					✓	[187]

Note: CS stands for commercial applications, MA indicates mobile applications, WA is web applications, and OS refers to open-source applications.

9. Challenges and Future Trends

The following list presents the trending topics for each type of face manipulation that require more attention from the research community:

- Face synthesis. Current face synthesis research typically relies on GAN structures, such as WGAN and PGGAN, to generate highly realistic results [192]. However, it has been demonstrated that face synthesis detectors can easily distinguish between authentic and synthesized images due to GAN-specific fingerprints. The primary challenge in face synthesis is to eliminate GAN-specific artifacts or fingerprints that can be easily detected by face synthesis detectors. The continuous evolution of GAN structures indicates that in the near future, it may become possible to eliminate these GAN fingerprints or add noise patterns to challenge the detectors while simultaneously improving the quality of the generated images [127]. Recent studies have focused on this trend, as it poses a challenge even for state-of-the-art face synthesis detectors. In addition, current face synthesis models struggle to generate facial expressions with high fidelity [193]. Future research should focus on developing methods to improve the realism and naturalness of generated facial expressions in synthesized faces.
- Face attribute editing. Similar to face synthesis, this type of manipulation also relies mainly on GAN architectures, so GAN fingerprint reduction can also be applied here [194]. Another challenge in face attribute editing is disentangling different facial attributes to independently control and manipulate them without affecting other attributes [195]. Finally, it is worth noting that there is a lack of benchmark datasets and standard evaluation protocols for a fair comparison between different studies.

- Face swapping. Even though various face-swapping databases have been described in this review, it may be hard for readers to determine the best one due to many issues. First of all, previous systems have generally shown promising performance because they were explicitly trained on a database with a fixed compression level. However, most of them demonstrated poor generalization ability when tested under unseen conditions. Moreover, each face-swapping generation study has used different experimental protocols and evaluation metrics, such as accuracy, AUC, and EER, and experimental protocols, making it hard to produce a fair comparison between studies. As a result, it is necessary to introduce a uniform evaluation protocol in order to advance the field further. Likewise, it is worth noting that face-swapping detectors have already obtained AUC values near 100% for the first generation of benchmark datasets, such as FaceForensics++ and UADFV [69]. However, most detectors have shown a considerable performance degradation, with AUC values under 60%, for the second-generation face-swapping datasets, particularly DFDC [63] and Celeb-DF [65]. As a result, novel ideas are required in the next generation of fake identification techniques, for example, through large-scale challenges, like the recent DFDC.
- Facial re-enactment. To the best of our knowledge, the most well-known facial re-enactment benchmark dataset is FaceForensics++, though such datasets are relatively scarce compared to those for face-swapping manipulation [67]. The FaceForensics++ dataset contains observable artifacts that can be easily detected using DL models, resulting in the highest AUC value of nearly 100% in various studies. Consequently, there is a pressing need for researchers to develop and introduce a large-scale facial re-enactment public dataset that includes highly realistic and high-quality images/videos. Such a dataset would enable the more comprehensive evaluation and benchmarking of facial re-enactment detection methods, fostering the advancement of this important area of research.

10. Conclusions

The advances in face manipulation techniques and DL have led to the wide spread of fake content, particularly deepfake, on the internet in recent decades. This review provides a comprehensive overview of the topic, covering various aspects of digital face manipulation, including types of manipulation, manipulation methods, benchmark datasets for research, and state-of-the-art results for the detection of each manipulation type.

In general, it was found that existing manipulated images from most current benchmark datasets can be easily detected under controlled scenarios, where fake detection models are tested under the same conditions as the training process. This holds true for the majority of benchmark datasets discussed in this survey, where the detectors achieved significantly low manipulation detection error rates. However, deploying these models in real-life scenarios, where high-quality and realistic fake images and videos can spread rapidly on social networks, poses a challenge. Factors such as noise, compression level, and resizing further complicate the generalization of fake detection models to unseen attacks and real-life circumstances.

To improve face manipulation detection performance and enhance the detectors' adaptability to various conditions, feature fusion has emerged as a trending approach. Recent manipulation detection systems have relied on fusing different feature types to achieve better results. For example, Zhou et al. explored the fusion of pure deep features with steganalysis features [143], while Trinh et al. analyzed spatial and temporal artifacts for effective detection [148]. Exciting fusion methods, such as RGB and near infrared (NIR), have also been studied to enhance manipulation detection.

Moreover, novel approaches are being encouraged to enhance the generalization and robustness of face manipulation detectors. One such example is the study conducted by [196,197], where the authors anticipated the difficulty of differentiating between real and fake content in the future due to advances in deepfake technology. To address this, they proposed a social-based manipulation verification system that used facial geometry across

videos to efficiently identify manipulated videos. Such innovative strategies are essential to stay ahead of the evolving landscape of face manipulation techniques and ensure the security and trustworthiness of digital media in the future.

Author Contributions: Conceptualization, M.D.; validation, T.N.N.; writing—original draft preparation, M.D.; writing—review and editing, T.N.N.; visualization, M.D.; investigation, M.D. and T.N.N.; All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Data Availability Statement: The datasets discussed in this study are readily accessible online: VGGFace2, accessed on 24 May 2022. CelebFacesAttributes (CelebA), accessed on 24 May 2022. Flickr-Faces-HQ dataset (FFHQ), accessed on 24 May 2022. CASIA-WebFace, accessed on 24 May 2022. Labeled Faces in the Wild (LFW), accessed on 24 May 2022. Diverse Fake Face Dataset (DFFD), accessed on 3 June 2022. iFakeFaceDB, accessed on 3 June 2022. 100K-Generated-Images, accessed on 3 June 2022. PGGAN dataset, accessed on 3 June 2022. 100K-Faces, accessed on 3 June 2022. ForgeryNIR, accessed on 5 June 2022. Face Forensics in the Wild (FFIW10K), accessed on 5 June 2022. OpenForensics, accessed on 5 June 2022. ForgeryNet, accessed on 5 June 2022. Korean deepfake detection (KoDF), accessed on 5 June 2022. Deepfake Detection Challenge (DFDC), accessed on 5 June 2022. Deepfake Detection Challenge preview, accessed on 5 June 2022. DeeperForensics-1.0, accessed on 5 June 2022. A large-scale challenging dataset for deepfake (Celeb-DF), accessed on 5 June 2022. WildDeepfake, accessed on 5 June 2022. FaceForensics++, accessed on 5 June 2022. Google DFD, accessed on 5 June 2022. UADFV, accessed on 5 June 2022. Media forensic challenge (MFC), accessed on 5 June 2022. Deepfake-TIMIT, accessed on 5 June 2022.

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

References

1. Radford, A.; Metz, L.; Chintala, S. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv* **2015**, arXiv:1511.06434.
2. Pidhorskyi, S.; Adjeroh, D.A.; Doretto, G. Adversarial latent autoencoders. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 14104–14113.
3. Deepnude. 2021. Available online: <https://www.vice.com/en/article/kzm59x/deepnude-app-creates-fake-nudes-of-any-woman> (accessed on 20 December 2021).
4. FaceApp. 2021. Available online: <https://www.faceapp.com/> (accessed on 24 May 2021).
5. Snapchat. 2021. Available online: <https://www.snapchat.com/> (accessed on 24 May 2021).
6. FaceSwap. 2021. Available online: <https://faceswap.dev/> (accessed on 24 May 2021).
7. Gupta, S.; Mohan, N.; Kaushal, P. Passive image forensics using universal techniques: A review. *Artif. Intell. Rev.* **2021**, *55*, 1629–1679. [[CrossRef](#)]
8. Media Forensics (MediFor). 2019. Available online: <https://www.darpa.mil/program/media-forensics> (accessed on 9 February 2022).
9. Goljan, M.; Fridrich, J.; Kirchner, M. Image manipulation detection using sensor linear pattern. *Electron. Imaging* **2018**, *30*, art00003. [[CrossRef](#)]
10. Vega, E.A.A.; Fernández, E.G.; Orozco, A.L.S.; Villalba, L.J.G. Image tampering detection by estimating interpolation patterns. *Future Gener. Comput. Syst.* **2020**, *107*, 229–237. [[CrossRef](#)]
11. Li, B.; Zhang, H.; Luo, H.; Tan, S. Detecting double JPEG compression and its related anti-forensic operations with CNN. *Multimed. Tools Appl.* **2019**, *78*, 8577–8601. [[CrossRef](#)]
12. Mohammed, T.M.; Bunk, J.; Nataraj, L.; Bappy, J.H.; Flenner, A.; Manjunath, B.; Chandrasekaran, S.; Roy-Chowdhury, A.K.; Peterson, L.A. Boosting image forgery detection using resampling features and copy-move analysis. *arXiv* **2018**, arXiv:1802.03154.
13. Long, C.; Smith, E.; Basharat, A.; Hoogs, A. A c3d-based convolutional neural network for frame dropping detection in a single video shot. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1898–1906.
14. Yu, P.; Xia, Z.; Fei, J.; Lu, Y. A Survey on Deepfake Video Detection. *IET Biom.* **2021**, *10*, 607–624. [[CrossRef](#)]
15. Kwon, M.J.; Nam, S.H.; Yu, I.J.; Lee, H.K.; Kim, C. Learning JPEG compression artifacts for image manipulation detection and localization. *Int. J. Comput. Vis.* **2022**, *130*, 1875–1895. [[CrossRef](#)]
16. Minh, D.; Wang, H.X.; Li, Y.F.; Nguyen, T.N. Explainable artificial intelligence: A comprehensive review. *Artif. Intell. Rev.* **2021**, *55*, 3503–3568. [[CrossRef](#)]
17. Abdolahnejad, M.; Liu, P.X. Deep learning for face image synthesis and semantic manipulations: A review and future perspectives. *Artif. Intell. Rev.* **2020**, *53*, 5847–5880. [[CrossRef](#)]
18. Dang, L.M.; Hassan, S.I.; Im, S.; Moon, H. Face image manipulation detection based on a convolutional neural network. *Expert Syst. Appl.* **2019**, *129*, 156–168. [[CrossRef](#)]

19. Dang, L.M.; Min, K.; Lee, S.; Han, D.; Moon, H. Tampered and computer-generated face images identification based on deep learning. *Appl. Sci.* **2020**, *10*, 505. [CrossRef]
20. Juefei-Xu, F.; Wang, R.; Huang, Y.; Guo, Q.; Ma, L.; Liu, Y. Countering malicious deepfakes: Survey, battleground, and horizon. *Int. J. Comput. Vis.* **2022**, *130*, 1678–1734. [CrossRef] [PubMed]
21. Malik, A.; Kuribayashi, M.; Abdullahi, S.M.; Khan, A.N. DeepFake Detection for Human Face Images and Videos: A Survey. *IEEE Access* **2022**, *10*, 18757–18775. [CrossRef]
22. Deshmukh, A.; Wankhade, S.B. Deepfake Detection Approaches Using Deep Learning: A Systematic Review. *Intell. Comput. Netw.* **2021**, *146*, 293–302.
23. Mirsky, Y.; Lee, W. The creation and detection of deepfakes: A survey. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–41. [CrossRef]
24. Thakur, R.; Rohilla, R. Recent advances in digital image manipulation detection techniques: A brief review. *Forensic Sci. Int.* **2020**, *312*, 110311. [CrossRef]
25. Tolosana, R.; Vera-Rodriguez, R.; Fierrez, J.; Morales, A.; Ortega-Garcia, J. Deepfakes and beyond: A survey of face manipulation and fake detection. *Inf. Fusion* **2020**, *64*, 131–148. [CrossRef]
26. Verdoliva, L. Media forensics and deepfakes: An overview. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 910–932. [CrossRef]
27. Kietzmann, J.; Lee, L.W.; McCarthy, I.P.; Kietzmann, T.C. Deepfakes: Trick or treat? *Bus. Horizons* **2020**, *63*, 135–146. [CrossRef]
28. Zheng, X.; Guo, Y.; Huang, H.; Li, Y.; He, R. A survey of deep facial attribute analysis. *Int. J. Comput. Vis.* **2020**, *128*, 2002–2034. [CrossRef]
29. Walia, S.; Kumar, K. An eagle-eye view of recent digital image forgery detection methods. In Proceedings of the International Conference on Next Generation Computing Technologies, Dehradun, India, 30–31 October 2017; Springer: Singapore, 2017; pp. 469–487.
30. Asghar, K.; Habib, Z.; Hussain, M. Copy-move and splicing image forgery detection and localization techniques: A review. *Aust. J. Forensic Sci.* **2017**, *49*, 281–307. [CrossRef]
31. Barni, M.; Costanzo, A.; Nowroozi, E.; Tondi, B. CNN-based detection of generic contrast adjustment with JPEG post-processing. In Proceedings of the 2018 25th IEEE International Conference on Image Processing (ICIP), Athens, Greece, 7–10 October 2018; pp. 3803–3807.
32. Qian, S.; Lin, K.Y.; Wu, W.; Liu, Y.; Wang, Q.; Shen, F.; Qian, C.; He, R. Make a face: Towards arbitrary high fidelity face manipulation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 10033–10042.
33. Xu, Z.; Yu, X.; Hong, Z.; Zhu, Z.; Han, J.; Liu, J.; Ding, E.; Bai, X. FaceController: Controllable Attribute Editing for Face in the Wild. *Proc. AAAI Conf. Artif. Intell.* **2021**, *35*, 3083–3091. [CrossRef]
34. Liu, Z.; Luo, P.; Wang, X.; Tang, X. Deep Learning Face Attributes in the Wild. In Proceedings of the International Conference on Computer Vision (ICCV), Santiago, Chile, 7–13 December 2015.
35. Westerlund, M. The emergence of deepfake technology: A review. *Technol. Innov. Manag. Rev.* **2019**, *9*, 39–52. [CrossRef]
36. Kwok, A.O.; Koh, S.G. Deepfake: A social construction of technology perspective. *Curr. Issues Tour.* **2021**, *24*, 1798–1802. [CrossRef]
37. Another Fake Video of Pelosi Goes Viral on Facebook. 2020. Available online: <https://www.washingtonpost.com/technology/2020/08/03/nancy-pelosi-fake-video-facebook/> (accessed on 9 February 2022).
38. Paris, B.; Donovan, J. Deepfakes and Cheap Fakes. 2019. Available online: <https://apo.org.au/node/259911> (accessed on 9 February 2022).
39. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
40. Liu, Y.; Chen, W.; Liu, L.; Lew, M.S. Swapgan: A multistage generative approach for person-to-person fashion style transfer. *IEEE Trans. Multimed.* **2019**, *21*, 2209–2222. [CrossRef]
41. Murphy, G.; Flynn, E. Deepfake false memories. *Memory* **2022**, *30*, 480–492. [CrossRef]
42. Kazemi, V.; Sullivan, J. One millisecond face alignment with an ensemble of regression trees. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 1867–1874.
43. Yi, D.; Lei, Z.; Liao, S.; Li, S.Z. Learning face representation from scratch. *arXiv* **2014**, arXiv:1411.7923.
44. Cao, Q.; Shen, L.; Xie, W.; Parkhi, O.M.; Zisserman, A. Vggface2: A dataset for recognising faces across pose and age. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi’an, China, 15–19 May 2018; pp. 67–74.
45. Huang, G.B.; Mattar, M.; Berg, T.; Learned-Miller, E. Labeled faces in the wild: A database for studying face recognition in unconstrained environments. In Proceedings of the Workshop on Faces in ‘Real-Life’ Images: Detection, Alignment, and Recognition, Marseille, France, 12–18 October 2008.
46. Zhang, G.; Kan, M.; Shan, S.; Chen, X. Generative adversarial network with spatial attention for face attribute editing. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 417–432.
47. Wang, L.; Chen, W.; Yang, W.; Bi, F.; Yu, F.R. A state-of-the-art review on image synthesis with generative adversarial networks. *IEEE Access* **2020**, *8*, 63514–63537. [CrossRef]
48. Karras, T.; Laine, S.; Aila, T. A style-based generator architecture for generative adversarial networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 4401–4410.

49. 100KGenerated. 100,000 Faces Generated by AI. 2018. Available online: <https://mymodernmet.com/free-ai-generated-faces/> (accessed on 24 May 2021).
50. Dang, H.; Liu, F.; Stehouwer, J.; Liu, X.; Jain, A.K. On the detection of digital face manipulation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5781–5790.
51. Neves, J.C.; Tolosana, R.; Vera-Rodriguez, R.; Lopes, V.; Proença, H.; Fierrez, J. Ganprintr: Improved fakes and evaluation of the state of the art in face manipulation detection. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 1038–1048. [[CrossRef](#)]
52. Li, S.; Yi, D.; Lei, Z.; Liao, S. The casia nir-vis 2.0 face database. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Portland, OR, USA, 23–28 June 2013; pp. 348–353.
53. Wang, Y.; Peng, C.; Liu, D.; Wang, N.; Gao, X. ForgeryNIR: Deep Face Forgery and Detection in Near-Infrared Scenario. *IEEE Trans. Inf. Forensics Secur.* **2022**, *17*, 500–515. [[CrossRef](#)]
54. Zhou, T.; Wang, W.; Liang, Z.; Shen, J. Face Forensics in the Wild. *arXiv* **2021**, arXiv:2103.16076.
55. Le, T.N.; Nguyen, H.H.; Yamagishi, J.; Echizen, I. OpenForensics: Large-Scale Challenging Dataset For Multi-Face Forgery Detection And Segmentation In-The-Wild. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 10117–10127.
56. Cao, H.; Cooper, D.G.; Keutmann, M.K.; Gur, R.C.; Nenkova, A.; Verma, R. Crema-d: Crowd-sourced emotional multimodal actors dataset. *IEEE Trans. Affect. Comput.* **2014**, *5*, 377–390. [[CrossRef](#)] [[PubMed](#)]
57. Livingstone, S.R.; Russo, F.A. The Ryerson Audio-Visual Database of Emotional Speech and Song (RAVD ESS): A dynamic, multimodal set of facial and vocal expressions in North American English. *PLoS ONE* **2018**, *13*, e0196391. [[CrossRef](#)]
58. Chung, J.S.; Nagrani, A.; Zisserman, A. Voxceleb2: Deep speaker recognition. *arXiv* **2018**, arXiv:1806.05622.
59. Ephrat, A.; Mosseri, I.; Lang, O.; Dekel, T.; Wilson, K.; Hassidim, A.; Freeman, W.T.; Rubinstein, M. Looking to listen at the cocktail party: A speaker-independent audio-visual model for speech separation. *ACM Trans. Graph. (TOG)* **2018**, *37*, 1–11. [[CrossRef](#)]
60. He, Y.; Gan, B.; Chen, S.; Zhou, Y.; Yin, G.; Song, L.; Sheng, L.; Shao, J.; Liu, Z. Forgerynet: A versatile benchmark for comprehensive forgery analysis. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 4360–4369.
61. Kwon, P.; You, J.; Nam, G.; Park, S.; Chae, G. KoDF: A Large-scale Korean DeepFake Detection Dataset. *arXiv* **2021**, arXiv:2103.10094.
62. Dolhansky, B.; Bitton, J.; Pflaum, B.; Lu, J.; Howes, R.; Wang, M.; Canton Ferrer, C. The deepfake detection challenge dataset. *arXiv* **2020**, arXiv:2006.07397.
63. Dolhansky, B.; Howes, R.; Pflaum, B.; Baram, N.; Ferrer, C.C. The deepfake detection challenge (dfdc) preview dataset. *arXiv* **2019**, arXiv:1910.08854.
64. Jiang, L.; Li, R.; Wu, W.; Qian, C.; Loy, C.C. Deeperforensics-1.0: A large-scale dataset for real-world face forgery detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 2889–2898.
65. Li, Y.; Sun, P.; Qi, H.; Lyu, S. Celeb-DF: A Large-scale Challenging Dataset for DeepFake Forensics. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Seattle, WA, USA, 13–19 June 2020.
66. Zi, B.; Chang, M.; Chen, J.; Ma, X.; Jiang, Y.G. Wilddeepfake: A challenging real-world dataset for deepfake detection. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 2382–2390.
67. Rossler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C.; Thies, J.; Nießner, M. Faceforensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 1–11.
68. Google DFD. 2020. Available online: <https://ai.googleblog.com/2019/09/contributing-data-to-deepfake-detection.html> (accessed on 17 December 2021).
69. Yang, X.; Li, Y.; Lyu, S. Exposing deep fakes using inconsistent head poses. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 8261–8265.
70. Guan, H.; Kozak, M.; Robertson, E.; Lee, Y.; Yates, A.N.; Delgado, A.; Zhou, D.; Kheyrkhan, T.; Smith, J.; Fiscus, J. MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation. In Proceedings of the 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), Waikoloa Village, HI, USA, 7–11 January 2019; pp. 63–72.
71. Korshunov, P.; Marcel, S. Deepfakes: A new threat to face recognition? assessment and detection. *arXiv* **2018**, arXiv:1812.08685.
72. Deepfakes. 2021. Available online: <https://github.com/deepfakes/faceswap> (accessed on 18 December 2021).
73. Thies, J.; Zollhofer, M.; Stamminger, M.; Theobalt, C.; Nießner, M. Face2face: Real-time face capture and reenactment of rgb videos. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2387–2395.
74. Faceswap. 2016. Available online: <https://github.com/MarekKowalski/FaceSwap/> (accessed on 18 December 2021).
75. Thies, J.; Zollhofer, M.; Nießner, M. Deferred neural rendering: Image synthesis using neural textures. *ACM Trans. Graph. (TOG)* **2019**, *38*, 1–12. [[CrossRef](#)]
76. Lahasan, B.; Lutfi, S.L.; San-Segundo, R. A survey on techniques to handle face recognition challenges: Occlusion, single sample per subject and expression. *Artif. Intell. Rev.* **2019**, *52*, 949–979. [[CrossRef](#)]

77. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A. Improved training of wasserstein GANs. In Proceedings of the 31st International Conference on Neural Information Processing Systems, Long Beach, CA, USA, 4–9 December 2017; pp. 5769–5779.
78. Mirza, M.; Osindero, S. Conditional generative adversarial nets. *arXiv* **2014**, arXiv:1411.1784.
79. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 2223–2232.
80. Gauthier, J. Conditional generative adversarial nets for convolutional face generation. *Cl. Proj. Stanf. CS231N Convolutional Neural Netw. Vis. Recognit. Winter Semester* **2014**, 2014, 2.
81. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 1–9.
82. Luo, G.; Xiong, G.; Huang, X.; Zhao, X.; Tong, Y.; Chen, Q.; Zhu, Z.; Lei, H.; Lin, J. Geometry Sampling-Based Adaption to DCGAN for 3D Face Generation. *Sensors* **2023**, *23*, 1937. [[CrossRef](#)]
83. Wang, Y.; Dantcheva, A.; Bremond, F. From attribute-labels to faces: Face generation using a conditional generative adversarial network. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018.
84. Bau, D.; Zhu, J.Y.; Wulff, J.; Peebles, W.; Strobel, H.; Zhou, B.; Torralba, A. Seeing what a gan cannot generate. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 4502–4511.
85. Liu, B.; Zhu, Y.; Song, K.; Elgammal, A. Towards faster and stabilized gan training for high-fidelity few-shot image synthesis. In Proceedings of the International Conference on Learning Representations, Addis Ababa, Ethiopia, 26–30 April 2020.
86. Neyshabur, B.; Bhojanapalli, S.; Chakrabarti, A. Stabilizing GAN training with multiple random projections. *arXiv* **2017**, arXiv:1705.07831.
87. Shahriar, S. GAN computers generate arts? A survey on visual arts, music, and literary text generation using generative adversarial network. *Displays* **2022**, *73*, 102237. [[CrossRef](#)]
88. Brock, A.; Donahue, J.; Simonyan, K. Large Scale GAN Training for High Fidelity Natural Image Synthesis. In Proceedings of the International Conference on Learning Representations, Vancouver, BC, Canada, 30 April–3 May 2018.
89. Jeong, Y.; Kim, D.; Min, S.; Joe, S.; Gwon, Y.; Choi, J. BiHPF: Bilateral High-Pass Filters for Robust Deepfake Detection. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikoloa, HI, USA, 3–8 January 2022; pp. 48–57.
90. Dzanic, T.; Shah, K.; Witherden, F. Fourier spectrum discrepancies in deep network generated images. *Adv. Neural Inf. Process. Syst.* **2020**, *33*, 3022–3032.
91. Dang, L.M.; Hassan, S.I.; Im, S.; Lee, J.; Lee, S.; Moon, H. Deep learning based computer generated face identification using convolutional neural network. *Appl. Sci.* **2018**, *8*, 2610. [[CrossRef](#)]
92. Chen, B.; Ju, X.; Xiao, B.; Ding, W.; Zheng, Y.; de Albuquerque, V.H.C. Locally GAN-generated face detection based on an improved Xception. *Inf. Sci.* **2021**, *572*, 16–28. [[CrossRef](#)]
93. Liu, Z.; Qi, X.; Torr, P.H. Global texture enhancement for fake face detection in the wild. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 8060–8069.
94. Wang, R.; Juefei-Xu, F.; Ma, L.; Xie, X.; Huang, Y.; Wang, J.; Liu, Y. Fakespotter: A simple yet robust baseline for spotting ai-synthesized fake faces. *arXiv* **2019**, arXiv:1909.06122.
95. Nataraj, L.; Mohammed, T.M.; Manjunath, B.; Chandrasekaran, S.; Flenner, A.; Bappy, J.H.; Roy-Chowdhury, A.K. Detecting GAN generated fake images using co-occurrence matrices. *arXiv* **2019**, arXiv:1903.06836.
96. Yang, X.; Li, Y.; Qi, H.; Lyu, S. Exposing gan-synthesized faces using landmark locations. In Proceedings of the ACM Workshop on Information Hiding and Multimedia Security, Paris, France, 3–5 July 2019; pp. 113–118.
97. McCloskey, S.; Albright, M. Detecting gan-generated imagery using color cues. *arXiv* **2018**, arXiv:1812.08247.
98. Marra, F.; Gagnaniello, D.; Verdoliva, L.; Poggi, G. Do gans leave artificial fingerprints? In Proceedings of the 2019 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), San Jose, CA, USA, 29–30 March 2019; pp. 506–511.
99. Mi, Z.; Jiang, X.; Sun, T.; Xu, K. GAN-Generated Image Detection With Self-Attention Mechanism Against GAN Generator Defect. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 969–981. [[CrossRef](#)]
100. Marra, F.; Gagnaniello, D.; Cozzolino, D.; Verdoliva, L. Detection of gan-generated fake images over social networks. In Proceedings of the 2018 IEEE Conference on Multimedia Information Processing and Retrieval (MIPR), Miami, FL, USA, 10–12 April 2018; pp. 384–389.
101. Wang, M.; Deng, W. Deep face recognition: A survey. *Neurocomputing* **2021**, *429*, 215–244. [[CrossRef](#)]
102. Karkkainen, K.; Joo, J. FairFace: Face Attribute Dataset for Balanced Race, Gender, and Age for Bias Measurement and Mitigation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Virtual, 5–9 January 2021; pp. 1548–1558.
103. Lu, Z.; Hu, T.; Song, L.; Zhang, Z.; He, R. Conditional expression synthesis with face parsing transformation. In Proceedings of the 26th ACM International Conference on Multimedia, Seoul, Republic of Korea 22–26 October 2018; pp. 1083–1091.
104. Liu, S.; Li, D.; Cao, T.; Sun, Y.; Hu, Y.; Ji, J. GAN-based face attribute editing. *IEEE Access* **2020**, *8*, 34854–34867. [[CrossRef](#)]
105. Choi, Y.; Choi, M.; Kim, M.; Ha, J.W.; Kim, S.; Choo, J. Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8789–8797.

106. Liu, M.; Ding, Y.; Xia, M.; Liu, X.; Ding, E.; Zuo, W.; Wen, S. STGAN: A unified selective transfer network for arbitrary image attribute editing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3673–3682.
107. Tripathy, S.; Kannala, J.; Rahtu, E. Facegan: Facial attribute controllable reenactment gan. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Virtual, 5–9 January 2021; pp. 1329–1338.
108. Muhammad, S.; Dailey, M.N.; Farooq, M.; Majeed, M.F.; Ekpanyapong, M. Spec-Net and Spec-CGAN: Deep learning models for specular removal from faces. *Image Vis. Comput.* **2020**, *93*, 103823. [[CrossRef](#)]
109. Antipov, G.; Baccouche, M.; Dugelay, J.L. Face aging with conditional generative adversarial networks. In Proceedings of the 2017 IEEE International Conference on Image Processing (ICIP), Beijing, China, 17–20 September 2017; pp. 2089–2093.
110. Perarnau, G.; Van De Weijer, J.; Raducanu, B.; Álvarez, J.M. Invertible conditional gans for image editing. *arXiv* **2016**, arXiv:1611.06355.
111. Ardizzzone, L.; Lüth, C.; Kruse, J.; Rother, C.; Köthe, U. Guided image generation with conditional invertible neural networks. *arXiv* **2019**, arXiv:1907.02392.
112. Pumarola, A.; Agudo, A.; Martinez, A.M.; Sanfeliu, A.; Moreno-Noguer, F. Ganimation: Anatomically-aware facial animation from a single image. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 818–833.
113. Thomas, C.; Kovashka, A. Persuasive faces: Generating faces in advertisements. *arXiv* **2018**, arXiv:1807.09882.
114. Mobini, M.; Ghaderi, F. StarGAN Based Facial Expression Transfer for Anime Characters. In Proceedings of the 2020 25th International Computer Conference, Computer Society of Iran (CSICC), Tehran, Iran, 1–2 January 2020; pp. 1–5.
115. Lee, C.H.; Liu, Z.; Wu, L.; Luo, P. Maskgan: Towards diverse and interactive facial image manipulation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5549–5558.
116. Kim, H.; Choi, Y.; Kim, J.; Yoo, S.; Uh, Y. Exploiting spatial dimensions of latent in gan for real-time image editing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 852–861.
117. Xiao, T.; Hong, J.; Ma, J. Dna-gan: Learning disentangled representations from multi-attribute images. *arXiv* **2017**, arXiv:1711.05415.
118. He, Z.; Zuo, W.; Kan, M.; Shan, S.; Chen, X. Attgan: Facial attribute editing by only changing what you want. *IEEE Trans. Image Process.* **2019**, *28*, 5464–5478. [[CrossRef](#)] [[PubMed](#)]
119. Collins, E.; Bala, R.; Price, B.; Susstrunk, S. Editing in style: Uncovering the local semantics of gans. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5771–5780.
120. Xu, Y.; Yin, Y.; Jiang, L.; Wu, Q.; Zheng, C.; Loy, C.C.; Dai, B.; Wu, W. TransEditor: Transformer-Based Dual-Space GAN for Highly Controllable Facial Editing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 7683–7692.
121. Guarnera, L.; Giudice, O.; Battiato, S. Deepfake detection by analyzing convolutional traces. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 666–667.
122. Tariq, S.; Lee, S.; Kim, H.; Shin, Y.; Woo, S.S. Detecting both machine and human created fake face images in the wild. In Proceedings of the 2nd International Workshop on Multimedia Privacy and Security, Toronto, ON, Canada, 15–19 October 2018; pp. 81–87.
123. Jain, A.; Singh, R.; Vatsa, M. On detecting gans and retouching based synthetic alterations. In Proceedings of the 2018 IEEE 9th International Conference on Biometrics Theory, Applications and Systems (BTAS), Redondo Beach, CA, USA, 22–25 October 2018; pp. 1–7.
124. Bharati, A.; Singh, R.; Vatsa, M.; Bowyer, K.W. Detecting facial retouching using supervised deep learning. *IEEE Trans. Inf. Forensics Secur.* **2016**, *11*, 1903–1913. [[CrossRef](#)]
125. Jain, A.; Majumdar, P.; Singh, R.; Vatsa, M. Detecting GANs and retouching based digital alterations via DAD-HCNN. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 672–673.
126. Karras, T.; Laine, S.; Aittala, M.; Hellsten, J.; Lehtinen, J.; Aila, T. Analyzing and improving the image quality of stylegan. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 8110–8119.
127. Yu, N.; Davis, L.S.; Fritz, M. Attributing fake images to gans: Learning and analyzing gan fingerprints. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 7556–7566.
128. Bitouk, D.; Kumar, N.; Dhillon, S.; Belhumeur, P.; Nayar, S.K. Face swapping: Automatically replacing faces in photographs. *ACM Trans. Graph. (TOG)* **2008**, *27*, 1–8. [[CrossRef](#)]
129. Li, L.; Bao, J.; Yang, H.; Chen, D.; Wen, F. Faceshifter: Towards high fidelity and occlusion aware face swapping. *arXiv* **2019**, arXiv:1912.13457.
130. Yan, S.; He, S.; Lei, X.; Ye, G.; Xie, Z. Video face swap based on autoencoder generation network. In Proceedings of the 2018 International Conference on Audio, Language and Image Processing (ICALIP), Shanghai, China, 16–17 July 2018; pp. 103–108.
131. Xingjie, Z.; Song, J.; Park, J.I. The image blending method for face swapping. In Proceedings of the 2014 4th IEEE International Conference on Network Infrastructure and Digital Content, Beijing, China, 19–21 September 2014; pp. 95–98.

132. Chen, D.; Chen, Q.; Wu, J.; Yu, X.; Jia, T. Face swapping: Realistic image synthesis based on facial landmarks alignment. *Math. Probl. Eng.* **2019**, *2019*, 8902701. [[CrossRef](#)]
133. Dale, K.; Sunkavalli, K.; Johnson, M.K.; Vlasic, D.; Matusik, W.; Pfister, H. Video face replacement. In Proceedings of the 2011 SIGGRAPH Asia Conference, Hong Kong, China, 12–15 December 2011; pp. 1–10.
134. Perov, I.; Gao, D.; Chervoniy, N.; Liu, K.; Marangonda, S.; Umé, C.; Dpfks, M.; Facenheim, C.S.; RP, L.; Jiang, J.; et al. Deepfacelab: A simple, flexible and extensible face swapping framework. *arXiv* **2020**, arXiv:2005.05535.
135. Nirkin, Y.; Keller, Y.; Hassner, T. Fsgan: Subject agnostic face swapping and reenactment. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 7184–7193.
136. Liu, K.; Wang, P.; Zhou, W.; Zhang, Z.; Ge, Y.; Liu, H.; Zhang, W.; Yu, N. Face Swapping Consistency Transfer with Neural Identity Carrier. *Future Internet* **2021**, *13*, 298. [[CrossRef](#)]
137. Zhu, Y.; Li, Q.; Wang, J.; Xu, C.Z.; Sun, Z. One Shot Face Swapping on Megapixels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 4834–4844.
138. Naruniec, J.; Helminger, L.; Schroers, C.; Weber, R.M. High-resolution neural face swapping for visual effects. In *Computer Graphics Forum*; Wiley Online Library: Hoboken, NJ, USA, 2020; Volume 39, pp. 173–184.
139. Xu, Y.; Deng, B.; Wang, J.; Jing, Y.; Pan, J.; He, S. High-resolution face swapping via latent semantics disentanglement. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 7642–7651.
140. Chen, R.; Chen, X.; Ni, B.; Ge, Y. Simswap: An efficient framework for high fidelity face swapping. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 2003–2011.
141. Xu, Z.; Hong, Z.; Ding, C.; Zhu, Z.; Han, J.; Liu, J.; Ding, E. MobileFaceSwap: A Lightweight Framework for Video Face Swapping. *arXiv* **2022**, arXiv:2201.03808.
142. Nguyen, H.H.; Fang, F.; Yamagishi, J.; Echizen, I. Multi-task Learning For Detecting and Segmenting Manipulated Facial Images and Videos. In Proceedings of the 10th IEEE International Conference on Biometrics: Theory, Applications, and Systems (BTAS 2019), Tampa, FL, USA, 23–16 September 2020.
143. Zhou, P.; Han, X.; Morariu, V.I.; Davis, L.S. Two-stream neural networks for tampered face detection. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 21–26 July 2017; pp. 1831–1839.
144. Dong, X.; Bao, J.; Chen, D.; Zhang, T.; Zhang, W.; Yu, N.; Chen, D.; Wen, F.; Guo, B. Protecting Celebrities from DeepFake with Identity Consistency Transformer. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 9468–9478.
145. Zhao, T.; Xu, X.; Xu, M.; Ding, H.; Xiong, Y.; Xia, W. Learning Self-Consistency for Deepfake Detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 15023–15033.
146. Das, S.; Seferbekov, S.; Datta, A.; Islam, M.; Amin, M. Towards Solving the DeepFake Problem: An Analysis on Improving DeepFake Detection using Dynamic Face Augmentation. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 3776–3785.
147. Sabir, E.; Cheng, J.; Jaiswal, A.; AbdAlmageed, W.; Masi, I.; Natarajan, P. Recurrent convolutional strategies for face manipulation detection in videos. *Interfaces (GUI)* **2019**, *3*, 80–87.
148. Trinh, L.; Tsang, M.; Rambhatla, S.; Liu, Y. Interpretable and trustworthy deepfake detection via dynamic prototypes. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Virtual, 5–9 January 2021; pp. 1973–1983.
149. Liu, J.; Zhu, K.; Lu, W.; Luo, X.; Zhao, X. A lightweight 3D convolutional neural network for deepfake detection. *Int. J. Intell. Syst.* **2021**, *36*, 4990–5004. [[CrossRef](#)]
150. Dong, S.; Wang, J.; Liang, J.; Fan, H.; Ji, R. Explaining Deepfake Detection by Analysing Image Matching. *arXiv* **2022**, arXiv:2207.09679.
151. Li, Y.; Lyu, S. Exposing DeepFake Videos By Detecting Face Warping Artifacts. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Long Beach, CA, USA, 16–17 June 2019.
152. Matern, F.; Riess, C.; Stamminger, M. Exploiting visual artifacts to expose deepfakes and face manipulations. In Proceedings of the 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW), Waikoloa Village, HI, USA, 7–11 January 2019; pp. 83–92.
153. Zhao, H.; Zhou, W.; Chen, D.; Wei, T.; Zhang, W.; Yu, N. Multi-attentional deepfake detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 2185–2194.
154. Zhang, J.; Zeng, X.; Wang, M.; Pan, Y.; Liu, L.; Liu, Y.; Ding, Y.; Fan, C. Freenet: Multi-identity face reenactment. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5326–5335.
155. Rössler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C.; Thies, J.; Nießner, M. Faceforensics: A large-scale video dataset for forgery detection in human faces. *arXiv* **2018**, arXiv:1803.09179.
156. Averbuch-Elor, H.; Cohen-Or, D.; Kopf, J.; Cohen, M.F. Bringing portraits to life. *ACM Trans. Graph. (TOG)* **2017**, *36*, 1–13. [[CrossRef](#)]
157. Wang, T.C.; Mallya, A.; Liu, M.Y. One-shot free-view neural talking-head synthesis for video conferencing. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 10039–10049.

158. Zakharov, E.; Shysheya, A.; Burkov, E.; Lempitsky, V. Few-shot adversarial learning of realistic neural talking head models. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 9459–9468.
159. Gu, Z.; Chen, Y.; Yao, T.; Ding, S.; Li, J.; Ma, L. Delving into the local: Dynamic inconsistency learning for deepfake video detection. *Proc. AAAI Conf. Artif. Intell.* **2022**, *36*, 744–752. [CrossRef]
160. Wang, J.; Wu, Z.; Chen, J.; Jiang, Y.G. M2TR: Multi-modal Multi-scale Transformers for Deepfake Detection. *arXiv* **2021**, arXiv:2104.09770.
161. Qian, Y.; Yin, G.; Sheng, L.; Chen, Z.; Shao, J. Thinking in frequency: Face forgery detection by mining frequency-aware clues. In *European Conference on Computer Vision*; Springer: Cham, Switzerland, 2020; pp. 86–103.
162. Zhang, B.; Li, S.; Feng, G.; Qian, Z.; Zhang, X. Patch Diffusion: A General Module for Face Manipulation Detection. *Proc. Assoc. Adv. Artif. Intell. (AAAI)* **2022**, *36*, 3243–3251. [CrossRef]
163. Afchar, D.; Nozick, V.; Yamagishi, J.; Echizen, I. Mesonet: A compact facial video forgery detection network. In Proceedings of the 2018 IEEE International Workshop on Information Forensics and Security (WIFS), Hong Kong, China, 11–13 December 2018; pp. 1–7.
164. Kumar, P.; Vatsa, M.; Singh, R. Detecting face2face facial reenactment in videos. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 2589–2597.
165. Liu, H.; Li, X.; Zhou, W.; Chen, Y.; He, Y.; Xue, H.; Zhang, W.; Yu, N. Spatial-phase shallow learning: Rethinking face forgery detection in frequency domain. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 772–781.
166. Amerini, I.; Galteri, L.; Caldelli, R.; Del Bimbo, A. Deepfake video detection through optical flow based cnn. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27 October–2 November 2019.
167. Cozzolino, D.; Rossler, A.; Thies, J.; Nießner, M.; Verdoliva, L. Id-reveal: Identity-aware deepfake video detection. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 15108–15117.
168. Khan, S.; Naseer, M.; Hayat, M.; Zamir, S.W.; Khan, F.S.; Shah, M. Transformers in vision: A survey. *ACM Comput. Surv. (CSUR)* **2021**, *54*, 1–41. [CrossRef]
169. These Deepfake Videos of Putin and Kim Have Gone Viral. 2020. Available online: <https://fortune.com/2020/10/02/deepfakes-putin-kim-jong-un-democracy-disinformation/> (accessed on 9 February 2022).
170. This Disturbingly Realistic Deepfake Puts Jeff Bezos and Elon Musk in a Star Trek Episode. 2020. Available online: <https://www.theverge.com/tldr/2020/2/20/21145826/deepfake-jeff-bezos-elon-musk-alien-star-trek-the-cage-amazon-tesla> (accessed on 9 February 2022).
171. Deepfake' Voice Tech Used for Good in David Beckham Malaria Campaign. 2019. Available online: <https://www.prweek.com/article/1581457/deepfake-voice-tech-used-good-david-beckham-malaria-campaign> (accessed on 9 February 2022).
172. How a Deepfake Tom Cruise on TikTok Turned into a Very Real AI Company. 2021. Available online: <https://edition.cnn.com/2021/08/06/tech/tom-cruise-deepfake-tiktok-company/index.html> (accessed on 9 February 2022).
173. Adobe. Adobe Photoshop. 2021. Available online: <https://www.adobe.com/products/photoshop.html> (accessed on 24 May 2021).
174. Faceswap. 2021. Available online: <https://faceswap.dev/download/> (accessed on 24 May 2021).
175. Xpression. 2021. Available online: <https://xpression.jp/> (accessed on 24 May 2021).
176. REFACE. 2021. Available online: <https://hey.reface.ai/> (accessed on 24 May 2021).
177. Impressions. 2021. Available online: <https://appadvice.com/app/impressions-face-swap-videos/1489186216> (accessed on 24 May 2021).
178. Myheritage. 2021. Available online: <https://www.myheritage.com/> (accessed on 24 May 2021).
179. Wombo. 2021. Available online: <https://www.wombo.ai/> (accessed on 24 May 2021).
180. Reflect. 2021. Available online: <https://oncreate.com/en/portfolio/reflect#:~:text=A%20first%2Dever%20artificial%20intelligence,picture%20in%20a%20split%20second.> (accessed on 24 May 2021).
181. DEEPFAKES WEB. 2021. Available online: <https://deepfakesweb.com/> (accessed on 24 May 2021).
182. FaceswapGAN. 2021. Available online: <https://github.com/shaoanlu/faceswap-GAN> (accessed on 24 May 2021).
183. DeepFaceLab. 2021. Available online: <https://github.com/iperov/DeepFaceLab> (accessed on 24 May 2021).
184. Deepware Scanner. 2019. Available online: <https://scanner.deepware.ai/> (accessed on 9 February 2022).
185. Face2face. 2021. Available online: <https://github.com/datitran/face2face-demo> (accessed on 24 May 2021).
186. Dynamixyz. 2021. Available online: <https://www.dynamixyz.com/> (accessed on 24 May 2021).
187. GeneratedPhotos. 2021. Available online: <https://generated.photos/> (accessed on 24 May 2021).
188. Deepfake Bots on Telegram Make the Work of Creating Fake Nudes Dangerously Easy. 2021. Available online: <https://www.theverge.com/2020/10/20/21519322/deepfake-fake-nudes-telegram-bot-deepnude-sensity-report> (accessed on 9 February 2022).
189. Microsoft Launches a Deepfake Detector Tool ahead of US Election. 2020. Available online: <https://techcrunch.com/2020/09/02/microsoft-launches-a-deepfake-detector-tool-ahead-of-us-election/> (accessed on 9 February 2022).
190. Synthetic and Manipulated Media Policy. 2020. Available online: <https://help.twitter.com/en/rules-and-policies/manipulated-media> (accessed on 9 February 2022).
191. Reddit, Twitter Ban Deepfake Celebrity Porn Videos. 2018. Available online: <https://www.complex.com/life/a/julia-pimentel/twitter-reddit-and-more-ban-deepfake-celebrity-videos> (accessed on 9 February 2022).

192. Mokhayeri, F.; Kamali, K.; Granger, E. Cross-domain face synthesis using a controllable GAN. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Snowmass Village, CO, USA, 1–5 March 2020; pp. 252–260.
193. Fu, C.; Hu, Y.; Wu, X.; Wang, G.; Zhang, Q.; He, R. High-fidelity face manipulation with extreme poses and expressions. *IEEE Trans. Inf. Forensics Secur.* **2021**, *16*, 2218–2231. [[CrossRef](#)]
194. Wang, J.; Alamayreh, O.; Tondi, B.; Barni, M. Open Set Classification of GAN-based Image Manipulations via a ViT-based Hybrid Architecture. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Vancouver, BC, Canada, 18–22 June 2023; pp. 953–962.
195. Shen, Y.; Yang, C.; Tang, X.; Zhou, B. Interfacegan: Interpreting the disentangled face representation learned by gans. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, *44*, 2004–2018. [[CrossRef](#)] [[PubMed](#)]
196. Tursman, E.; George, M.; Kamara, S.; Tompkin, J. Towards untrusted social video verification to combat deepfakes via face geometry consistency. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 654–655.
197. Tursman, E. Detecting deepfakes using crowd consensus. *XRDS Crossroads ACM Mag. Stud.* **2020**, *27*, 22–25. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.