Journal of
Sensor and Actuator Networks

MDPI

Article

An Investigation of the Effectiveness of Deepfake Models and Tools

Md. Saddam Hossain Mukta [1,*], Jubaer Ahmad [1], Mohaimenul Azam Khan Raiaan [1], Salekul Islam [1], Sami Azam [2], Mohammed Eunus Ali [3] and Mirjam Jonkman [2]

[1] Department of Computer Science and Engineering, United International University, Madani Avenue, Dhaka 1212, Bangladesh; jahmad181023@bscse.uiu.ac.bd (J.A.); mraiaan191228@bscse.uiu.ac.bd (M.A.K.R.); salekul@cse.uiu.ac.bd (S.I.)
[2] Faculty of Science and Technology, Charles Darwin University, Casuarina, NT 0909, Australia; sami.azam@cdu.edu.au (S.A.); mirjam.jonkman@cdu.edu.au (M.J.)
[3] Department of Computer Science and Engineering, Bangladesh University of Engineering and Technology (BUET), West Palasi, Dhaka 1000, Bangladesh; eunus@cse.buet.ac.bd
* Correspondence: saddam@cse.uiu.ac.bd; Tel.: +880-1712-095216

Abstract: With the development of computer vision and deep learning technologies, rapidly expanding approaches have been introduced that allow anyone to create videos and pictures that are both phony and incredibly lifelike. The term *deepfake methodology* is used to describe such technologies. Face alteration can be performed both in videos and pictures with extreme realism using deepfake innovation. Deepfake recordings, the majority of them targeting politicians or celebrity personalities, have been widely disseminated online. On the other hand, different strategies have been outlined in the research to combat the issues brought up by deepfake. In this paper, we carry out a review by analyzing and comparing (1) the notable research contributions in the field of deepfake models and (2) widely used deepfake tools. We have also built two separate taxonomies for deepfake models and tools. These models and tools are also compared in terms of underlying algorithms, datasets they have used and their accuracy. A number of challenges and open issues have also been identified.

Keywords: deepfake; deep learning; autoencoder; GANs; CNN; RNN; transformer

Citation: Mukta, M.S.H.; Ahmad, J.; Raiaan, M.A.K.; Islam, S.; Azam, S.; Ali, M.E.; Jonkman, M. An Investigation of the Effectiveness of Deepfake Models and Tools. *J. Sens. Actuator Netw.* **2023**, *12*, 61. https://doi.org/10.3390/jsan12040061

Academic Editor: Lei Shu

Received: 6 May 2023
Revised: 14 June 2023
Accepted: 16 June 2023
Published: 4 August 2023

1. Introduction

With the advent of deepfake [1], it is now possible to produce very realistic videos to show humans talking and performing actions that have never happened. Deepfake is one of the most recent innovations making a severe social impact. Combining the velocity and accessibility of social media with convincing deepfake content can quickly reach a large number of audiences. Therefore, many users experience anxiety as a result of the spread of these phony videos, which have made deepfake scary as well as infamous. Distinguishing between legitimate and fake media (i.e., video, image, etc.) is becoming impossible with the naked eye due to the enhancement of deepfake technologies; it is even difficult to differentiate with the help of available tools. For facial expression swapping in photos and videos, artificial intelligence (AI)-based applications including *Face2Face* [2] and *FaceSwap* [3] have been widely utilized. Using such face-altering techniques, it is possible to change someone's appearance, haircut, age, movement of the lips and eyes, as well as other physical properties. In this paper, we first conduct a review that investigates different deepfake models and tools and their respective performance and effectiveness.

The advancement of computer vision and deep learning technologies has significantly contributed to the proliferation of deepfake techniques. Deepfake creators alter facial expressions, gestures, and even entire identities, which is viable due to the capability of manipulating and generating content by learning from vast datasets [4]. The emergence of deep fake videos and images raises grave concerns in numerous fields. Deepfakes can

undermine public trust and manipulate public opinion in politics [5]. The technology can be used to create counterfeit interviews and speeches that misrepresent politicians, resulting in misinformation and the propagation of deceptive narratives. Deepfakes can be used in the entertainment industry to superimpose the images of celebrities onto explicit content, violating their privacy and harming their reputations [5]. Moreover, in journalism, deepfakes can blur the line between reality and fiction, making it harder to verify news integrity and undermining media outlets' credibility. Beyond these domains, deep fake technology threatens security, privacy, and trust in numerous societal contexts. Therefore, understanding the effectiveness of available deepfake-detection tools [6–8] is a pressing topic to combat the social, political, and personal issues raised by deepfake images and videos.

In recent times, a significant amount of research [9–11] has been performed in the area of deepfake. As a result, various review papers have compiled summaries in various disciplines. In order to conduct this research, we first looked at papers for deepfake-detection tools. Various fundamental issues include *privacy* [12,13], *face-swapping precision* [13], *model design*, *social effects* [14], and AI threat [15] depending on the area and particular use cases. The purpose of this paper is to review the existing research literature and to summarize the cutting-edge strategies that have recently been invented to address these issues. We also determine the accuracy of deepfake detection in a variety of well-known online tools using different technologies and datasets they have employed. Additionally, by reviewing the most recent advancements in each of the mentioned deepfake fields of study, we pinpoint the holes in the studied deepfake surveys. We discuss the issues, usages, and design considerations comprehensively and suggest possible future research possibilities.

While some studies [16–18] have evaluated the quality of specific deepfake movies, these assessments are sometimes subjective and lack uniformity, making it difficult to compare findings across research or to identify areas for improvement. The specific objectives of this study are to review and evaluate deepfake generative models and detection tools. Instead of focusing solely on the technical details of deepfake tools and technologies, we intend to determine how effectively these tools can generate convincing and realistic synthetic media. This analysis will provide insight into the strengths and shortcomings of current deepfake technology, which may guide future development. Our goal is to determine how successfully these tools can produce authentic and convincing synthetic media that are indistinguishable from real footage or audio. These data will enable us to identify any shortcomings in current deepfake technology and direct future development.

In summary, we make the following contributions:

- An in-depth, up-to-date review is carried out in the field of deepfake models and deepfake tools.
- Two separate taxonomies are proposed that categorize the existing deepfake models and tools.
- The effectiveness of existing deepfake models and detection tools are compared in terms of underlying algorithms, datasets used and accuracy.

The organization of the paper is graphically presented in Figure 1. Section 2 describes deepfake and its evolution, and Section 3 illustrates the materials and methods of the paper. Section 4 presents the positive and negative impacts of deepfake tools and techniques. Section 5 initializes deepfake classification. Section 6 describes different deepfake models, Section 7 describes different deepfake tools, and Section 8 presents deep-learning-based deepfake tools. Sections 9 and 10 discuss the open issues and challenges in deepfake, and Section 11 concludes the paper.

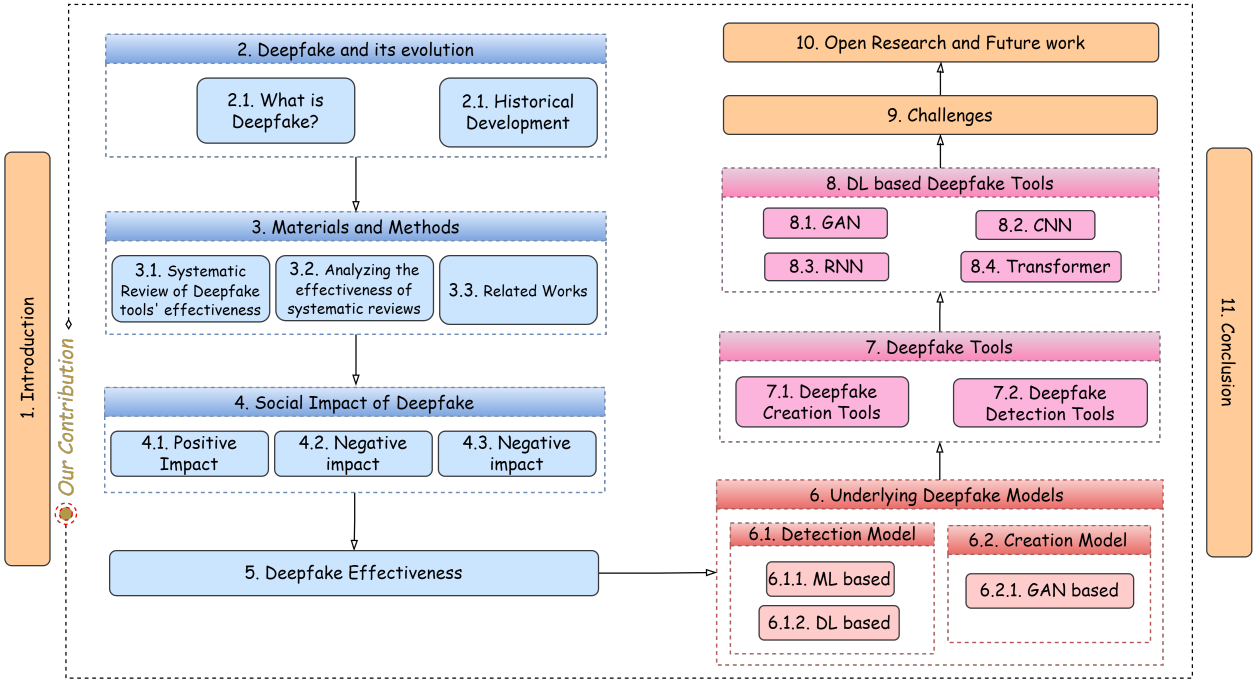

Figure 1. Structure of the paper.

2. Deepfake and Its Evolution

In this section, we first explore the concept of deepfake technology and its generation process. Additionally, we examine the historical evolution of deepfake technologies.

2.1. What Is Deepfake?

Deepfakes [1] are pragmatic media (i.e., videos and images) that have been digitized to represent individuals acting and performing actions that have never actually occurred. They integrate "*deep learning*" and "*fake*". Deepfakes use neural networks that learn to replicate a person's facial gestures, characteristics, speech, and intonations by analyzing enormous quantities of data samples. In order to train a deep learning system to interchange facial expressions, two people's video contents are fed into the system. Deepfakes employ a facial positioning system and AI to replace one person's face into a video with the facial appearance of another individual.

Figure 2 depicts the face swapping of two persons with their respective facial characteristics and expressions where the *FaceDancer* [19,20] tool has been used for this. We collect these human photos from a public repository, Pexels (https://www.pexels.com/search/human/ (accessed on 3 February 2023)). To accomplish the face swapping, we provide the source and target photos. Prior to swapping in the target image, the tool first recognizes the source face's facial region and emotion. We see the facial region surrounded by a shaded circle in the figure. The facial region consists of the eyes, eyebrows, nose, and mouth region. The blue-shaded circle represents the source image's facial area that will be swapped with the red-shaded circle area of the target image. The facial attributes of the source image (e.g., eyes, nose, mouth, etc.) will be swapped with the target facial attributes, but the expression of the target image will remain the same. Finally, the output face-swapped image as a green-shaded circle area represents the difference.

The majority of successful deepfake video-generation and detection algorithms use two subgroups of neural networks: (1) *autoencoders* and (2) *generative adversarial networks (GANs)*.

An *autoencoder* [21–23] is a form of neural network that is used in deepfakes. These are made up of an encoder, which shrinks an image to a hidden space with fewer dimensions, as well as a decoder, which builds the image back up from the series of hidden layers.

Deepfakes make use of this approach by encoding a person through to the latent space using a generic encoder. Important details about their face characteristics and body movements are contained in the series of hidden layers. Then, a model is trained particularly so that the objective may be decoded, which is presented in Figure 3. In other words, the latent space's representation of the source video's body characteristics and facial traits will be overlaid with the target's specific information. After training is finished, a latent face created from latent picture A can be sent to decoder A. Figure 4 shows how the decoder will attempt to recreate latent image B using data related to latent image A. On the other hand, there are some drawbacks of the autoencoder: the average of the input set from an autoencoder might always be received, it might always rebuild the input set precisely, it might trickily combine the two flaws, there may be an incorrect use case for training, decoding errors, failing to recognize key elements, etc.

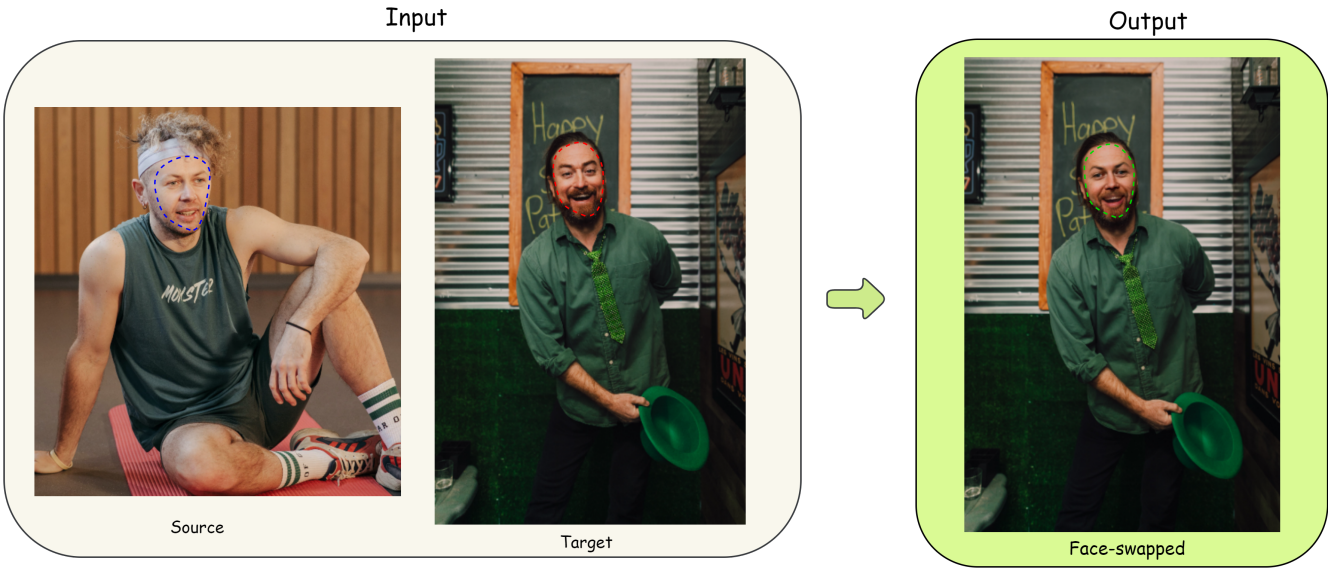

Figure 2. Example of face swapping in deepfake.

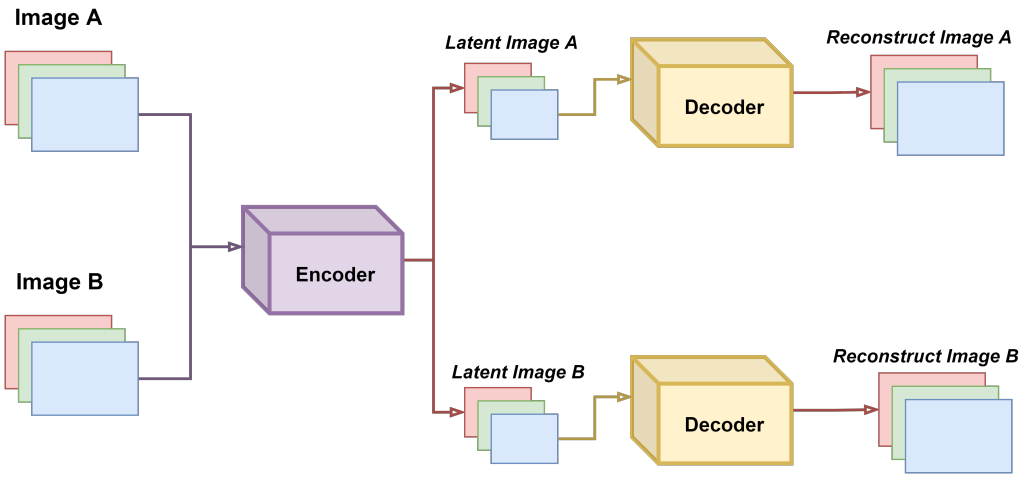

Figure 3. Training phase in autoencoder for deepfake-generation model.

To fix any potential weaknesses in the autoencoders, there are *generative adversarial network (GAN)* [24] models. GAN employs the generator and discriminator as unsupervised polarized sub-models. The generator alters the source that it was trained on to produce phony visual or aural outputs. The discriminator tries to detect whether the image is generated, while the generator generates new images using the latent representation of the original material. As a result, the generator produces incredibly realistic images because

any flaws would be detected by the discriminator. The fundamental GAN architecture procedure is depicted in Figure 5.

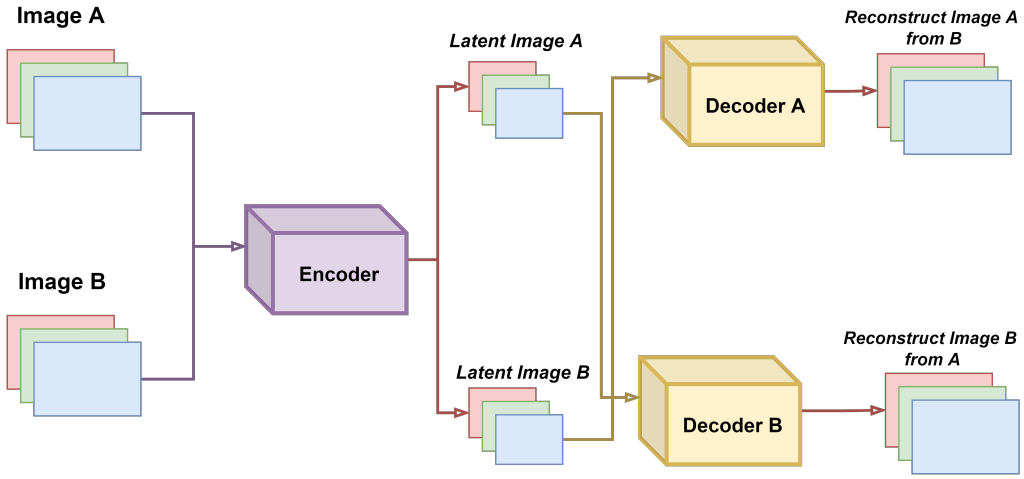

Figure 4. Generating deepfake images using autoencoder.

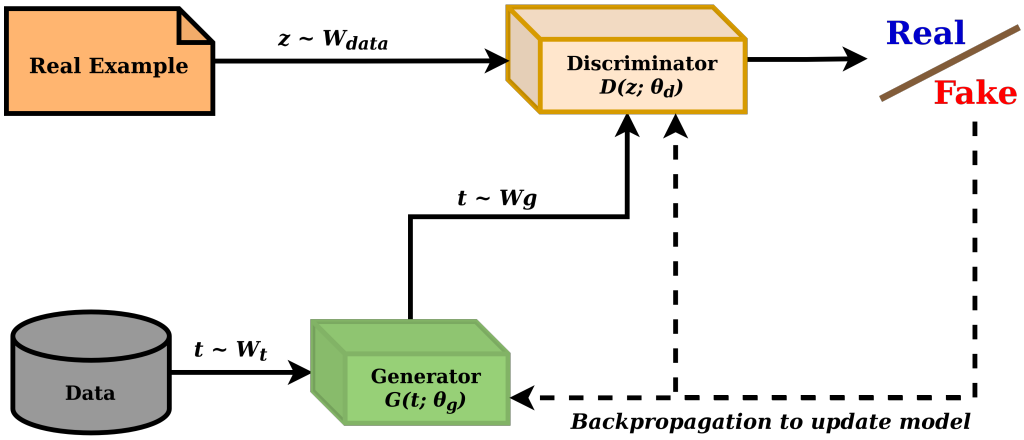

Figure 5. Generate deepfake using GAN.

The mentioned deepfake architectures are baseline models used for generating synthetic data. Based on these architectures, several optimized models have been proposed for the creation or detection of deepfakes. There are many areas of research that are being investigated to generate and detect cutting-edge models, novel datasets, and deepfake-detection methods. Additionally, the present research initiatives seek to lessen the main issues with deepfake, including its negative social impacts and privacy and security concerns.

2.2. Historical Development of Deepfake

The historical developments in deepfake technology by year from 2014 to 2022:

2014: First mention of deep learning in deepfake technology [11] and research on deep-learning-based face recognition systems.

2016: Face2Face: Real-time face capture and reenactment of RGB videos and the first significant deepfake viral video (President Obama) [2].

2017: Development of advanced algorithms for automatic face swap technology and the emergence of advanced tools such as DeepFaceLab [25].

2018: Deepfake pornographic videos [26] start to appear on the internet. Researchers develop detection methods for deepfakes and manipulated videos that impact political campaigns.

2019: Social media companies begin to take action against deepfake videos, and there is development of deepfake-detection software by AI companies.

2020: Deepfake technology continues to evolve with the creation of deepfake voice technology [27], and AI companies develop advanced detection software to counter deepfakes.

2021: The emergence of deepfake [1] text technology, allowing for the creation of fake news articles and other written content, increases the use of deepfakes for fraud and scams and continued development of deepfake-detection methods and software.

2022: The emergence of deepfake technologies security [10], solving the creation of fake news articles and other written content.

Overall, the historical development of deepfake technology has seen significant advances in both the creation and detection of deepfakes. While the technology has many potential uses, such as in the entertainment industry, it also poses significant risks, particularly in regard to its potential for spreading false information and manipulating public opinion. Figure 6 depicts the historical development of deepfake technologies from 2014 to 2022.

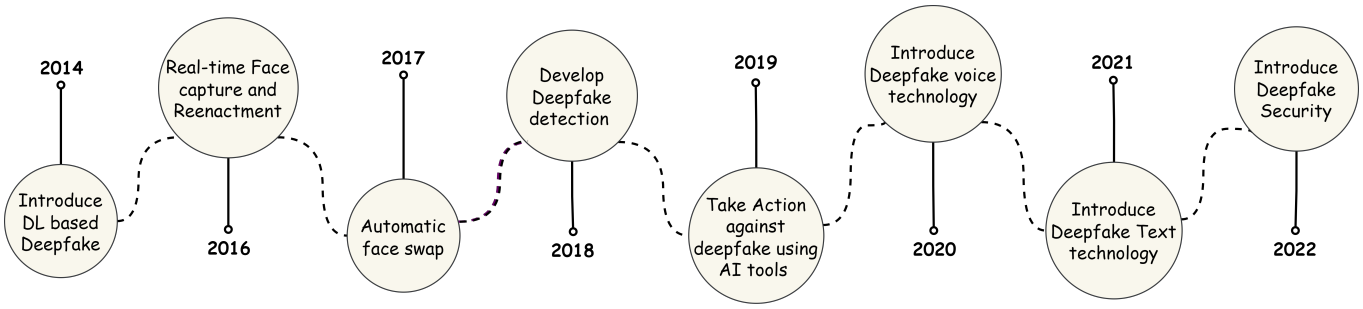

Figure 6. Historical development of deepfake technology.

3. Materials and Methods

In this section, first, we conduct a systematic review of deepfake tools, then we analyze the quality of the papers, and finally, we discussed the related works.

3.1. Systematic Review of Deepfake Tools' Effectiveness

According to Pilares [28], a systematic review is a suitable method for compiling existing studies and identifying the gaps that can suggest a new area of research. A comprehensive review was performed to compile a summary of the effectiveness of the deepfake tools now available and their underlying models' attempts to address the problems with synthetic media. The search phase and the definition of the inclusion and exclusion criteria are the two main sections of the review methodology.

The search step entails specifying the academic resources, digital databases, and search engines that may be used to look for appropriate research, as well as the questions that are going to use AND/OR Boolean operators to identify all connected results. The databases used for this systematic review are displayed in Table 1. All of the searches are included in Table 2.

Table 1. Electronic database search.

Electronic Database	Type	URL
IEEE Xplore	Digital Library	https://ieeexplore.ieee.org/Xplore/home.jsp (accessed on 12 January 2023)
Springer	Digital Library	https://www.springer.com/gp (accessed on 12 January 2023)
Google Scholar	Search Engine	https://scholar.google.com.au (accessed on 12 January 2023)
Science Direct—Elsevier	Digital Library	https://www.sciencedirect.com (accessed on 12 January 2023)
MDPI	Digital Library	https://www.mdpi.com (accessed on 12 January 2023)
Researchgate	Social networking site	https://www.researchgate.net (accessed on 12 January 2023)

Table 2. Search queries used for the systematic review.

	Search Queries (SQ)
SQ1	"deepfake generation" AND machine learning OR deep learning OR models
SQ2	"deepfake generation" AND machine learning OR deep learning OR tools
SQ3	"deepfake detection" AND machine learning OR deep learning OR models
SQ4	"deepfake detection" AND machine learning OR deep learning OR tools

The following stage involves defining the inclusion criteria (IC) and exclusion criteria (EC). To improve query results, these were carefully laid out. The breakdown of IC and EC for this investigation is shown in Table 3. All EC-related studies were immediately disqualified. To obtain a more targeted search outcome, the titles, abstracts, and full texts of findings can be further filtered.

Title: The keywords given in Table 3 were used to filter out studies that failed to include at least one of them.

Abstract: Papers that fulfill no less than 40% of IC were kept for review.

Full text: Papers should describe diverse models and techniques as well as approaches that address deepfake issues.

Table 3. Inclusion and exclusion criteria.

	List of Inclusion and Exclusion Criteria
Inclusion Criteria (IC)	
IC1	Should contain at least one of the keywords
IC2	Must be included in one of the selected databases
IC3	Published within the last ten years (2014–2023)
IC4	Publication in a journal, conference is required
IC5	The research being examined should have a matching title, abstract, and full text
Exclusion Criteria (EC)	
EC1	Redundant items
EC2	Whole text of paper cannot be taken
EC3	Purpose of the paper is not related to deepfake
EC4	Non-english documents

At the start of the study, a total of 1365 records were obtained, where 805 were gathered through database searches and 560 from other resources. After screening, 214 records from database searches and 128 from other resources were retained. Upon applying eligibility criteria, 31 papers from other sources and 38 studies from the database search were deemed suitable. Eventually, a total of 69 studies were considered relevant for this paper. Figure 7 provides a more detailed illustration of this process.

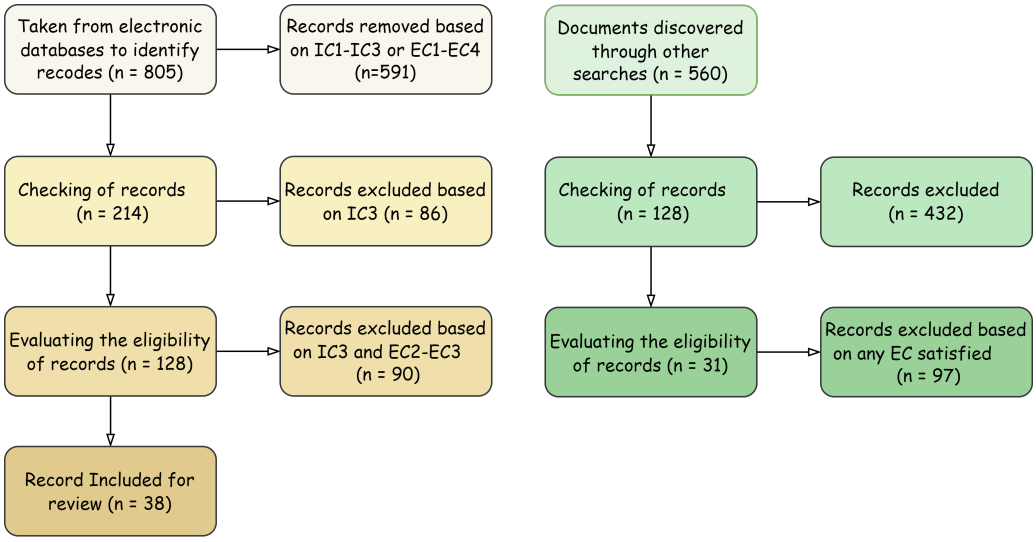

Figure 7. Flow diagram of systematic review.

3.2. Analyzing the Effectiveness of Systematic Reviews

The quality of papers included in the systematic review is evaluated using two methods. The first method involves quality evaluation (QE) questions, where each question can be marked 0, 1, or 2 per reviewer ("no" (0), "partially" (1), or "yes" (2)). The score for each question is a cumulative score of two markers, with a maximum score of 4 for each question (i.e., if marker1 gives 2 and marker2 gives 1, then the quality evaluation of this question is 3). The second method involves a pass/fail criterion based on a cumulative score between 0 and 15 as a failure and between 16 and 28 as a pass. The reviewers decided on these criteria for inclusion/exclusion of papers in the review. Table 4 provides a summary of the quality assessment conducted for the included papers in the review. All 65 papers reviewed met the threshold criteria and were included in the review. A visual representation of these results can be seen in Figure 8. Table 5 illustrates related survey papers and Table 6 depicts the accuracy of deepfake-detection techniques based on underlying model and dataset.

The second approach involves making sure that the papers selected for the review are of high quality, with a conference ranking of A or B (conference rank obtained from (http://www.conferenceranks.com/ (accessed on 11 January 2023)) or a journal ranking of Q1 or Q2. Out of the 38 papers that were included in this systematic review, 14 papers were from conferences ranked as A, 4 papers were from conferences ranked as B, 7 papers were from arXiv, and the remaining 13 papers were from Q1- and Q2-ranked journals. Figure 9 illustrates the distribution of these papers.

Table 4. Summary of the results for rating the quality of the papers.

Paper	QE1: Is the Publication Associated with Deepfake?	QE2: Is the Suggested Solution Completely Obvious?	QE3. Is There a Deepfake Model Proposed in the Publication?	QE4. Is a Deepfake Tool Implemented in the Suggested Solution?	QE5. Are Challenges Addressed in the Proposed Solution?	QE6. Is the Proposed Solution Ready for Implementation?	QE7. Did the Publication Define the Limitations of the Proposed Solutions?	Score
[29]	4	3	4	1	4	2	4	22
[30]	4	2	4	1	4	3	4	22
[31]	4	3	2	1	4	3	3	20
[32]	4	2	3	1	4	3	4	21
[33]	4	2	4	3	3	1	3	20

Table 4. *Cont.*

Paper	QE1: Is the Publication Associated with Deepfake?	QE2: Is the Suggested Solution Completely Obvious?	QE3. Is There a Deepfake Model Proposed in the Publication?	QE4. Is a Deepfake Tool Implemented in the Suggested Solution?	QE5. Are Challenges Addressed in the Proposed Solution?	QE6. Is the Proposed Solution Ready for Implementation?	QE7. Did the Publication Define the Limitations of the Proposed Solutions?	Score
[34]	4	2	3	2	2	3	3	19
[35]	4	3	4	4	2	4	2	23
[36]	4	3	4	4	3	4	4	26
[37]	4	3	4	3	4	4	3	25
[38]	4	3	4	3	2	4	4	24
[39]	4	4	2	3	3	4	2	22
[31]	4	3	4	2	3	4	4	24
[40]	4	4	3	4	3	2	4	24
[41]	4	3	3	4	3	3	3	23
[42]	4	2	3	4	4	3	4	24
[43]	4	4	3	3	3	2	4	23
[44]	4	3	2	3	2	2	4	20
[45]	4	4	1	4	4	3	2	22
[46]	4	4	2	4	4	3	3	24
[47]	4	3	4	4	4	4	2	25
[48]	4	4	4	4	4	2	4	26
[49]	4	3	4	3	4	1	4	23
[50]	4	2	4	4	3	1	2	20
[51]	4	4	3	4	3	4	3	25
[52]	4	4	3	3	2	3	2	21
[53]	4	4	2	4	3	2	2	21
[54]	4	3	2	4	3	4	4	26
[55]	4	2	3	3	2	2	4	20
[56]	4	4	3	4	1	3	2	21
[57]	4	4	1	4	4	4	2	23
[58]	4	3	1	4	4	3	2	21
[59]	4	4	1	4	3	4	4	24
[60]	4	4	3	4	4	3	4	26
[61]	4	4	3	3	4	4	2	24
[62]	4	4	2	3	3	2	3	21
[63]	4	4	4	3	4	2	2	23
[64]	4	4	4	4	4	4	1	25
[65]	4	4	2	4	4	4	3	25

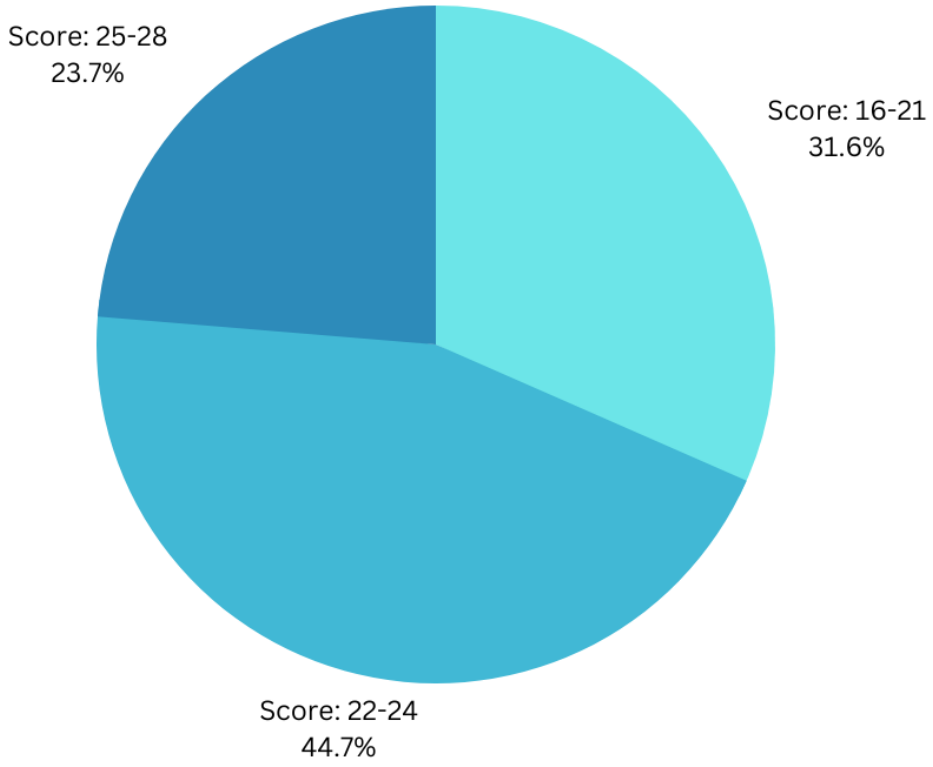

Figure 8. Outcome based on the quality assessment of the 38 papers.

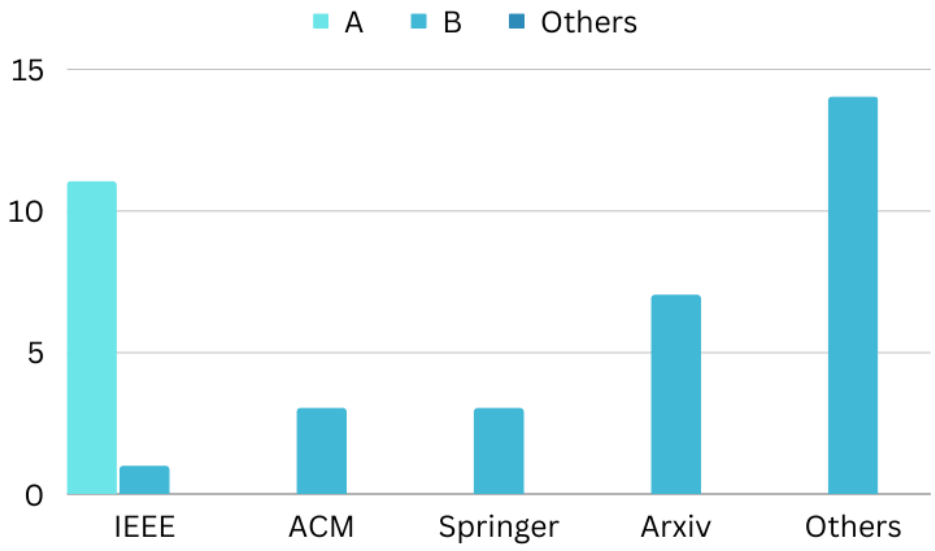

Figure 9. Distribution of papers based on their ranking (according to: http://conferenceranks.com (accessed on 11 January 2023)).

3.3. Related Works

The study of deepfake technology and its potential impact has gained significant attention in recent years. Previous research has explored the technical aspects of deepfake creation, including the use of machine learning algorithms and neural networks. However, there is a need for more comprehensive surveys that examine the effectiveness of deepfake tools in creating convincing synthetic media. Several studies [66] have evaluated the quality of specific deepfake videos, but these evaluations are often subjective and lack standardization. To address this gap in the literature, this paper presents a survey of the techniques used to measure the effectiveness of deepfake tools.

One of the most important areas of research regarding deepfake is creation and detection of deepfake. Several studies have illustrated various techniques for deepfake creation as well as detection. In [9], a survey of thorough details of deepfake creation and detection using various machine learning techniques was performed. The authors of the study [67] categorized deepfake methodologies into four groups, including deep-learning-based, classical machine-learning-based, statistical, and blockchain-based techniques. They conducted an evaluation of the performance of these methods in terms of their detection capability, using various datasets. Their findings indicate that deep-learning-based methods perform better than the other three categories in detecting deepfakes. On the other hand, in [68], the authors provide a summary of the deepfake-detection methods applied to both face images and videos, based on their performance, results, detection type, and methodology. In addition, they classify the existing deepfake creation techniques into five major categories, which will be reviewed in this study. Similarly, the focus of the paper [11] is on providing a survey of the algorithms utilized in the creation of deepfakes, as well as the proposed methods to detect such deepfakes in the current literature. The paper includes in-depth discussions regarding the challenges and research trends that are relevant to the domain of deepfake technologies and offers insights on possible future directions for research in this area. The paper [9] conducts a survey on deepfake creation and detection techniques, emphasizing the network architectures used for this purpose. It includes thorough discussions on the effectiveness of different deep learning networks and their corresponding architectures utilized in various studies. The paper by [69] surveys the tools and algorithms employed for creating and detecting deepfakes. Although it briefly touches on the challenges, advances, and strategies related to deepfakes, the number of studies discussed is limited. The paper [70] conducts a systematic literature review (SLR) of deepfake creation and detection techniques, covering images and videos, and studies related to deepfake tweets, which are not commonly included in similar surveys.

Finally, there is a growing body of research that has examined the ethical implications of deepfake technology. This includes studies on the potential risks and harms associated with deepfakes, as well as discussions of the ethical considerations involved in their creation and use.

Overall, the literature suggests that there is a need for more standardized and objective techniques for measuring the effectiveness of deepfake tools. This survey aims to provide a comprehensive overview of the existing evaluation methods, to identify areas for improvement, and to inform the development of more robust and effective techniques for evaluating deepfake technology.

4. Social Impact of Deepfake

One of the most well-known innovations, deepfake, has many ethical concerns with both positive and negative effects on society. We first briefly explain the merits and limitations of this technology as follows.

4.1. Positive Impact

Films, curriculum content, electronic communications, videogames, entertainment, social platforms, medical, nanotechnology, and numerous industry sectors along with clothing and e-commerce are just a few of the industries that benefit from deepfake innovation [1,71].

- Deepfake technology has many advantages for the movie business. For instance, it can be used to update film footage rather than reshoot it or to create artificial voices for performers who lost theirs due to illness. The ability for filmmakers to reproduce iconic movie moments and produce new films starring long-dead performers can be brought back to life in post-production with the use of cutting-edge facial editing and visual effects. Deepfake technology also enables automatic and lifelike voice dubbing for films in any language, enhancing the viewing experience for different audiences of movies and instructional media.

- Deepfake technology enables digital doubles of individuals, realistic-sounding and smart-looking assistants [72], and enhanced telepresence in online games and virtual chat environments [73]. This promotes improved online communication and interpersonal relationships [74,75]. In the social and medical spheres, technology can also be beneficial. By digitally bringing a deceased friend "back to life," deepfakes can assist a grieving loved one in saying goodbye to her. This can help people deal with the death of a loved one [76,77]. Additionally, it can be used to digitally replicate an amputee's leg or assist transgender people in better visualizing their preferred gender. It is even possible to engage with a younger face that the used may remember thanks to deep-fake technology [76]. In order to accelerate the development of new materials and medical treatments [78], researchers are also investigating the use of GANs to detect anomalies in X-rays [79].
- Businesses are intrigued by the possibility of brand-applicable deepfake technology since it has the chance to significantly change e-commerce and marketing [75]. For instance, businesses can hire phony models and actresses to display fashionable attire on a wide range of models with various heights, weights, and skin colors [80]. Additionally, deepfakes facilitate super-personal information that transforms customers into models; the technology allows for virtual modification to help customers see how an outfit would appear on themselves before buying it and can produce specifically aimed fashion advertisements that change based on the period, climate, and viewing public [75,80].

4.2. Negative Impact

Deepfakes create overwhelming risks, nevertheless. They have been employed as a famous weapon for political propaganda efforts, to tarnish the image of journalists by including them in pornographic films, and to make totally new types of interpersonal deceiving. Because of this, decision makers and experts are alerting the public to the threats that deepfakes may pose. We go over a few of them and describe their effects on society, politics, and the economy.

Recrimination Porn: Without bringing up its pornographic use, the list of harms that deepfakes cause can never be fully expressed. As we explained before in this essay, the term "deepfakes" derives its name from being used to create pornographic material on purpose. To put this into context, VOX research [81–84] indicates that 96% of the deepfake movies in online pornography in 2020 were produced to harm the reputations of their victims. In particular, this puts the prestige of female celebrities at risk, which may cause severe harm if not regulated.

Propagation of Politicians: Deepfakes can be used by political enemies to sway the people and to foster mistrust. Barack Obama was caught on camera in 2008, 2012, and 2016 stating that individuals in hard-hit areas frequently turned to religion and guns, Mitt Romney said that 47% of Americans were content to rely on the government for basic necessities, and Hillary Clinton was caught on camera rejecting a community of Trump fans and calling them "deplorable", all of which were later revealed to be forgeries. Deepfakes' political complexities and effects have the potential to harm global democracy if they are not well regulated.

Scamming and phishing: Deepfake technology might potentially lead to a rise in internet scams, fraudulent allegations, or complaints against businesses. A deepfake such as this is created by capturing an actual incident and then editing the audio to add new conversation in order to deceive viewers.

4.3. Ethics of Deepfake

Even though deepfakes have become beneficial in areas such as the film industry, expression, and the arts, they have primarily been weaponized for malicious ends. Deepfakes have the potential to weaken democracy, do harm to people and businesses, and further diminish public confidence in the media. The use of deepfakes to create a false story is risky and may hurt people and the wider community, whether intentionally or accidentally. Deepfakes, which are not just fake but which are also incredibly lifelike, could exacerbate the post-truth dilemma since they deceive our most basic auditory and visual senses. Deepfakes made with the intention of intimidating, humiliating or blackmailing a person are categorically unethical, and their effects on the democratic system need to be considered. There are ethical problems with different deepfake domains.

- A person may be threatened, intimidated, or suffer mental damage as a result of pornographic deepfakes. Women are treated with harshness and discrimination, which results in psychological pain, injury to one's reputation, bullying, and in certain situations even loss of money or career. When it comes to consenting to artificial pornography, the ethical problem is considerably more complicated. Mutually acceptable deepfakes could normalize the concept of synthetic pornography, which might increase worries about the harmful effects of pornography on emotional and sexual development. Some may claim that this is similar to the ethically right activity of sexual fantasizing.
- Synthetic resurrection is one more field of worry. People have the ability to determine how their likenesses are used for commercial purposes. The biggest issue with public figures is who will control their voice and appearance once they pass away. Most of the time, they are used primarily for marketing, propaganda, and financial benefit. Deepfakes can be employed to falsely depict political leaders' reputations after their deaths in order to further political and legal objectives, which raises issues of morality and ethics. Despite the fact that there are valid barriers against using a dead person's voice or image for profit, relatives who are granted the right to utilize these attributes may do so for their own business advantage.
- Extending the truth, emphasizing a political platform, and offering other facts are common strategies in politics. They aid in organizing, influencing, and persuading individuals to collect funds and votes. Although unethical, political opportunism has become the standard. If politicians decide to employ deepfakes and artificial media, the results of the election could be significantly affected. People who are deceived cannot make judgments that are in their individual greatest advantage because deception prevents them from doing so. Voters are manipulated into supporting the deceiver's agenda when misleading information about the opposing party is purposefully spread or a candidate is presented with a different version of events [85]. There is a little legal remedy for these immoral activities. A deepfake that is employed to frighten people into not casting their ballots is also unethical.

Deepfake suppliers and manufacturers need to be certain that they use and apply artificial media in an ethical manner. Large technological companies such as Microsoft, Google, and Amazon have an ethical responsibility [86] since they offer the cloud infrastructure and tools needed to quickly and efficiently produce deepfakes. The use of deepfakes requires that social media such as Facebook, Twitter, LinkedIn, and TikTok, as well as news media organizations, news reporters, politicians and policymakers, and civil society, demonstrate an ethical and moral obligation. These platforms allow for the mass distribution of deepfakes. To overcome ethical concerns, paper [70] proposes how to counter the threats from deepfake technology and to alleviate its impact. Figure 10 depicts the ethics of deepfake.

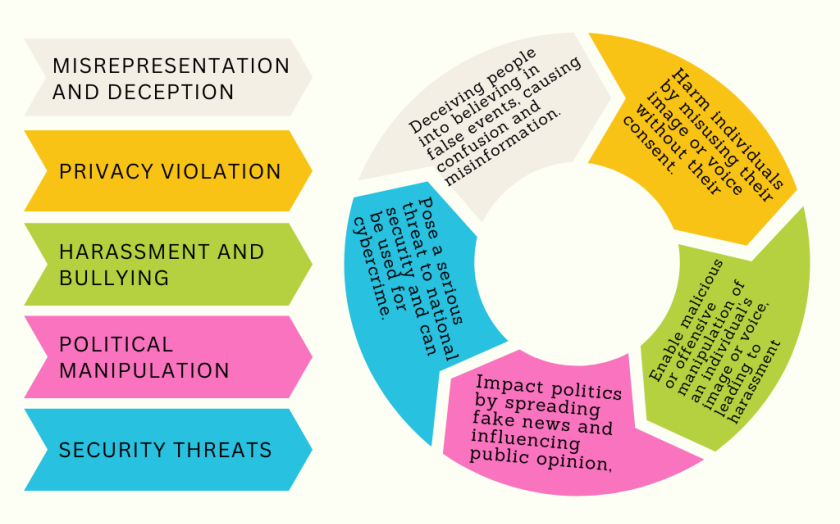

Figure 10. Ethics of deepfake.

5. Deepfake Effectiveness

This study has two taxonomies: (i) underlying deepfake models and (ii) deepfake tools. Both of these taxonomies are divided into two sections: detection and deepfake creation. For the first section of the first taxonomy (the deepfake-detection model), ML-based and DL-based models are analyzed on benchmark datasets such as FF, FF++, VidTIMIT, Celeb-DF, DFDC, and COHFACE based on accuracy. The second section (creation model) of the first taxonomy analyzes the GAN-based model, which is studied with a brief inclusion of the process of the models, application area, and limitations. The first section of the second taxonomy (deepfake tools analysis) is introduced for the creation tools and for analyzing the comprehensive details of the underlying model, its main focus, accuracy, speed, usability, and security. For the second section, the detection tools, along with the accuracy and speed of user-friendliness and scalability, are also introduced. Section 6 analyzes the deepfake models, and Section 7 explores the tools.

6. Underlying Deepfake Models

This section covers different techniques for creating and detecting deepfake content such as videos, images, audio, and text. Additionally, deepfake models are divided into creation and detection models. In Figure 11, we present the taxonomy of deepfake models.

6.1. Detection Models

The existing models in the literature that are being developed to detect deepfake are studied in this section. Based on the underlying detection algorithm, we can broadly categorize these models into two types: (1) conventional machine-learning-based models and (2) deep-learning-based models. Deep-learning-based models can be further classified into CNN, RNN, and transformer-based models.

6.1.1. Conventional ML-Based Deepfake

Modern machine learning techniques are crucial for comprehending the reasoning behind any choice that may be justified from a human perspective. These techniques provide more control over data and procedures, making them appropriate for the deepfake area. Additionally, changing the model's design and hyperparameters is significantly simpler. Decision trees [40,87], random forests [41,43,88], and other tree-based machine learning techniques use a tree to symbolize the decision-making process. As a result, there are no interpretability issues with the tree-based approach. On the other hand, there are several studies that utilize support vector machine [42–44,89,90], logistic regression [43,44,46], and

KNN [44] classifiers and some boosting models (e.g., XGBoost [40,45,91], ADABoost [46]) to identify deepfake.

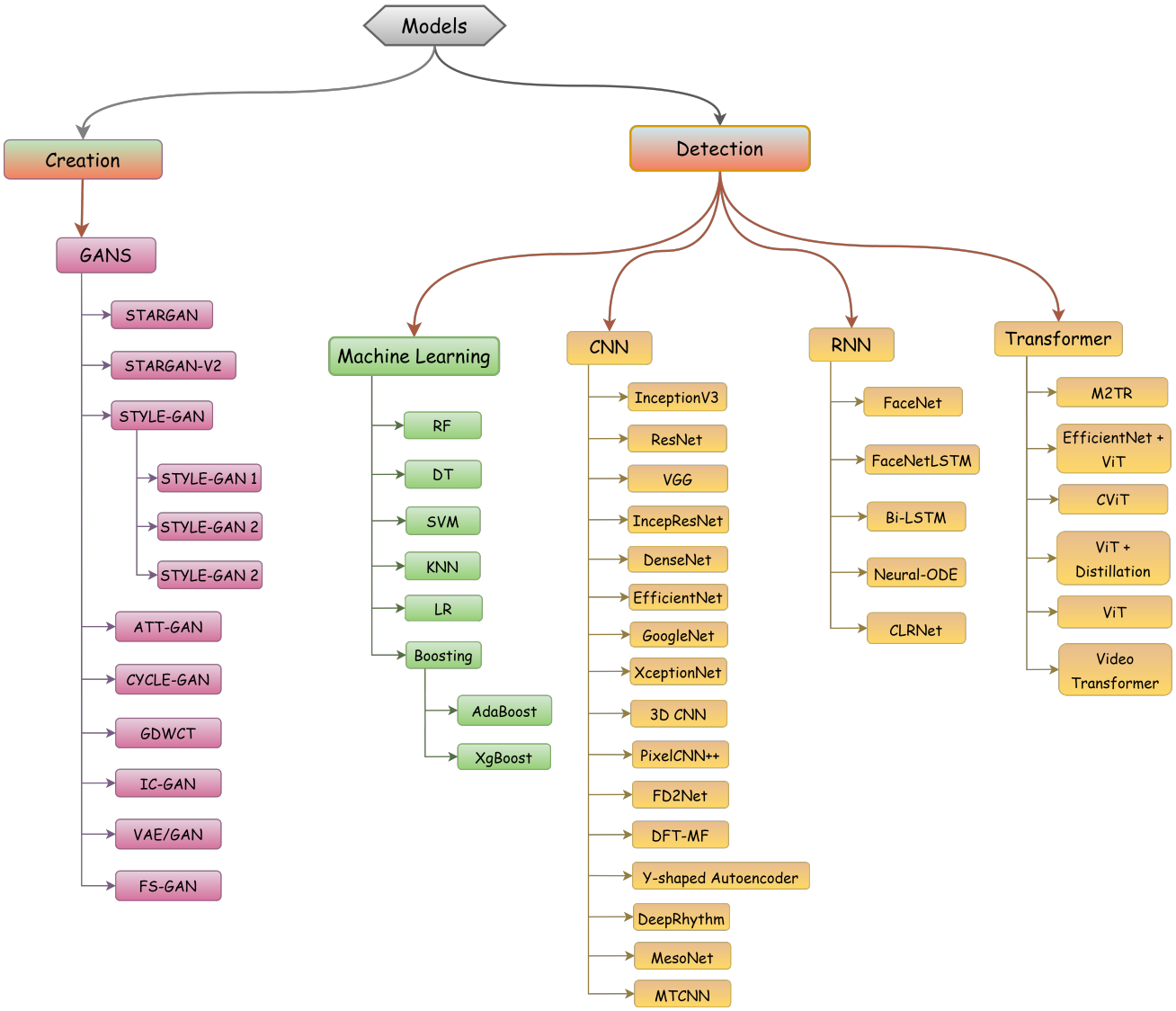

Figure 11. Taxonomy of deepfake models.

6.1.2. DL-Based Deepfake

Deep learning models have been extensively utilized in computer vision because of their method for the selection and extraction of features, which allows them to immediately retrieve or learn features from input.

CNN-based Models

One of the best deep learning models is CNN. It is well known and directly utilizes pretrained CNN methodologies to learn distinguishing features from each individual frame of the sequences. We discovered that the aforementioned CNN-based approaches have been employed in deepfake detection studies: Xception [47,54], GoogleNet [48], VGG16 [50,92], VGG19 [49], ResNet50/101/152 [50], Y-shaped Autoencoder [51], MesoNet [52], DeepRhythm [53], InceptionNet/InceptionResNet/InceptionV4 [54,93], Convolutional Attention Network (CAN) [55], PixelCNN++ [56], EfficentNet [57], 3D CNN [58], FD2Net [59], DFT-MF [60], MTCNN [61].

RNN-based Models

One of the most used models for sequential data in deep learning is RNN. Numerous RNN models have been found to be employed in the creation and detection of deepfake images and videos. Here are several RNN models that can be used to create and detect deepfakes: BiLSTM [94], FaceNet, FacenetLSTM [62], Neural-ODE [95], CLRNet [96,97], CNN+(Bidirectional+entropy RNN) [98], CNN+RNN [99].

Transformer-based Models

Transformers are highly well-known deep learning models, and they are now used in the field of deepfake generation and detection. The following transformer models can be used to generate and recognize deepfakes: EfficentNet+ViT [57], M2TR [63], CViT [65], ViT [100], ViT+Distillation [64], Video Transformer [101].

We looked at different deep-learning-based models to determine which ones used which datasets and had the best accuracy for detecting and generating fake news, as shown in Table 6. Our taxonomy of these models is based on the technologies illustrated in Figure 11, namely traditional machine learning and deep learning (CNN, GAN, RNN, and transformer).

Accuracy of deepfake-detection models

Table 6 presents different deepfake-detection model-based algorithms, their types of models, popular datasets, and the respective accuracy of those models in detecting deepfake with very high accuracy. The table presents both independent (i.e., CNN, RNN, and Transformer) and hybrid (MTCNN+RNN) models. We find that CNN-based deepfake models have a greater diversity with many transfer learning and attention-based techniques. The majority of the models have been tested over FaceForensics++ (FF++), VidTIMIT, and other datasets. The accuracy of the models is in the range from 64.10% (DeepRhythm model and DFDC dataset) to 99.90% (CNN+attention hybrid model and Celeb-DF dataset).

6.2. Creation Models

This section examines the models that are currently being used in the literature to generate deepfake. We concentrate on GAN-based creation methods based on the underlying creation methodology.

6.2.1. GANs-Based Deepfake

Deepfakes are typically produced using methods that rely on generative adversarial networks (GANs), which Goodfellow et al. [24] initially introduced. In an adversarial mode, the authors devised a new method for determining generative models that involve training two models at once: (1) *Discriminative model D*, and (2) *Generative model G*. We may observe from Figure 5 that a sample's probability of coming from the training data as opposed to the *Generative model G* can be determined by the *Discriminative model D*. To make it more likely that D will make a mistake and start a min–max two-player game, the training procedure for *G* aims to enhance the likelihood of this happening. The generator mathematically accepts a random input *t* with a density *Wt* and then generates an output $x = G(t; \theta g)$ with a particular probability distribution *Wg* (θg: parameters of the generative model).

The discriminator, $D(z; \theta d)$, determines the likelihood that x originates from the real example data *Wdata* (*d* denotes the discriminative model's parameters). After the training phase, the main goal is to obtain a generator that is a factor that gives rise to *Wdata*. As a result, the desired probability distribution, i.e., *Wdata*, will be followed by pg since the discriminator is "deluded" and can no longer tell the samples apart from *Wdata* and *Pg*.

By treating the unsupervised problem as supervised and producing photo-realistic imitation faces in photos or videos, GANs are utilized to autonomously train a generative model. Many different types of GAN-based techniques are utilized to identify deepfake. Table 7 provides an overview of the GAN-based deepfake detection and generation model.

Table 5. Related survey papers.

Paper	Type	Published	DF Tools	DF Detection Model	DF Creation Model	Tool's Effectiveness			DF Detection Tools	DF Creation Tools	Comparison of DF Tools	Scope
						Accuracy	Speed	Usabilty				
[67]	Survey	IEEE	×	✓	×	×	×	×	×	×	×	Groups 112 articles into 4 categories: deep learning, classical machine learning, statistical, and blockchain techniques. Evaluates detection performance on various datasets.
[68]	Survey	IEEE	✓	✓	✓	×	×	×	×	×	×	Explores trends and challenges in deepfake datasets and detection models, as well as challenges in creating and detecting deepfakes.
[11]	Survey	Elsevier	✓	✓	✓	×	×	×	×	×	×	Provides an overview of deepfake creation algorithms and detection methods. It also covers the challenges and future directions of deepfake technology.
[9]	Survey	ACM	×	✓	✓	×	×	×	×	×	×	Improves understanding of deepfakes by discussing their creation, detection, trends, limitations of current defenses, and areas requiring further research.
[69]	Survey	Springer	✓	✓	✓	×	×	×	×	×	×	Offers survey of deepfake algorithms and tools, along with discussions on challenges and research trends.
[70]	SLR	MDPI	×	✓	✓	×	×	×	×	×	×	Focuses on recent research on deepfake creation and detection methods, covering tweets, pictures, and videos. It also discusses popular deepfake apps and research in the field.
Our	Systematic review	MDPI	✓	✓	✓	✓	✓	✓	✓	✓	✓	Provides a thorough and current evaluation of deepfake models and tools. It proposes two classification systems to categorize these models and tools and compares their effectiveness based on factors such as algorithms, datasets, and accuracy.

✓→ symbol used to mark that the feature is present in this paper. × → symbol used to mark that the feature is not present in this paper.

Table 6. Accuracy of deepfake-detection techniques based on underlying model and dataset.

Category	Model	Dataset	Accuracy
CNN	Xception [54]	FF++	99.26%
	GoogleNet [48]	Own Dataset	92.70%
	VGG16 [50]	VidTIMIT	84.60%
	VGG19 [49]	FF++	92.02%
	ResNet50 [50]	VidTIMIT	99.90%
	ResNet101 [50]	VidTIMIT	87.60%
	ResNet152 [50]	VidTIMIT	99.40%
	Y-shaped Autoencoder [51]	FF, FF++	93.01%
	MesoNet [52]	FF++	95.23%
	Meso-4 [52]	VidTIMIT	87.80%
	MesoInception-4 [52]	VidTIMIT	80.40%
	DeepRhythm [53]	FF++	98.00%
	DeepRhythm [53]	DFDC	64.10%
	Convolutional Attention Network(CAN) [55]	Celeb-DF	99.90%
	Convolutional Attention Network(CAN) [55]	DFDC	98.20%
	PixelCNN++ [56]	FF	96.20%
	EfficentNet [57]	FF++	95.10%
	3D CNN [58]	FF++	88.57%
	FD2Net [59]	FF++	99.45%
	FD2Net [59]	DFD	78.65%
	FD2Net [59]	DFDC	66.09%
	DFT-MF [60]	Celeb-DF	71.25%
	DFT-MF [60]	VidTIMIT	98.70%
RNN	BiLSTM [94]	FF++	99.34%
	FacenetLSTM [62]	FF++	97.00%
	Neural-ODE [95]	COHFACE	99.01%
	Neural-ODE [95]	VidTIMIT	99.02%
MTCNN+RNN	CLRNet [61]	FF++	96.00%
Transformer	Transformer-based model (EfficentNet+ViT) [57]	FF++	95.10%

STAR-GAN [33] is a cutting-edge generative adversarial network that successfully trains on image data from all categories and learns the mappings between diverse aspects with only one generator as well as a discriminator. This technique is capable of performing image-to-image translation and can translate images between several domains with just one model, as shown in Figure 12. It has therefore been contrasted with other currently used techniques [36,102] and demonstrates how STAR-GAN is able to produce images of greater graphical excellence.

STYLE-GAN [34] or style generative adversarial network, modifies the generator model of STAR-GAN by mapping points into latent space to an intermediary latent space that regulates the *style* output at each point of the process of generation, as shown in Figure 13. Additionally, adding noise as a source of fluctuation in the previously described elements produces superior outcomes.

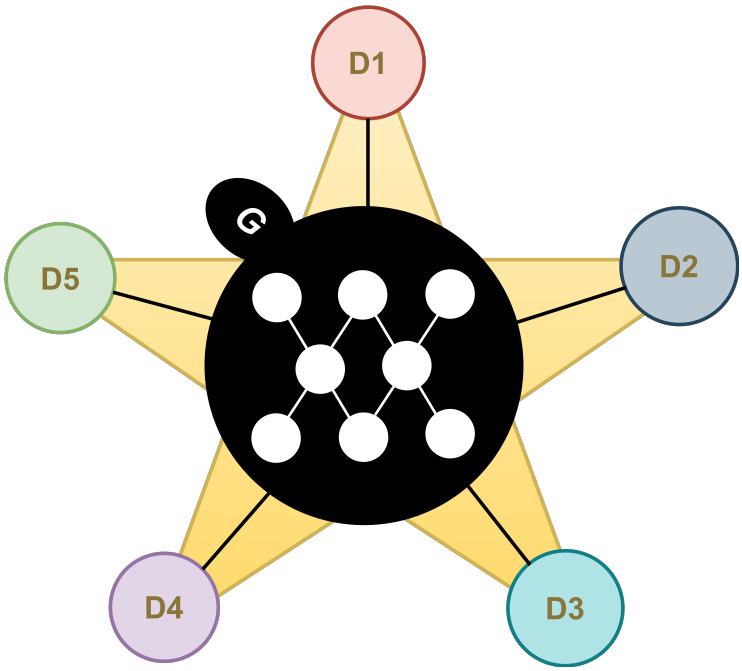

Figure 12. StarGAN learning framework.

Table 7. Overview of GAN models.

GAN Model	Process of the Model	Application Area	Limitation	Used in Deepfake
STARGAN [33]	To use a particular model, image-to-image translations at several domains.	Image-to-image translations (e.g., facial attribute, facial expression).	StarGAN tries to manipulate the age of the source images and is unable to generate facial expressions when the incorrect mask vector is utilized.	[103–106]
STYLEGAN [34]	Sends semantic information to a target domain with a distinctive style from the source domain.	Create incredibly genuine, high-resolution photographs of people's faces.	It is obvious from looking at the distribution of training data that low-density regions are underrepresented, making it more challenging for the generator to learn in those areas.	[106–113]
ATTGAN [35]	Constrained categorization and the transmission of facial traits.	Attribute intensity control, attribute style manipulation	Cannot manipulate style attribute.	[105,111,114]
CycleGAN [36]	To convert a visual representation from one domain to another when there are not any paired examples.	Style transfer, object transfiguration, season transfer, photo enhancement.	Many of their outcomes are rendered hazy and do not keep the level of clarity as observed in the input, failing to keep the identification of the input.	[115–117]
GDWCT [37]	Increases the capacity for styling.	Specially designed for image translations.	It makes mistakes while classifying gender.	[37,105,111]
Ic-GAN [38]	Mapping a real picture into a conditional interpretation and latent space.	Apply to recreate and alter real-world pictures of faces based on random attributes.	By producing photographs of men, the self-identity in the picture is not preserved.	[38]

Table 7. *Cont.*

GAN Model	Process of the Model	Application Area	Limitation	Used in Deepfake
VAE/GAN [39]	While translating an image, swaps element-wise mistakes for feature-wise defects to effectively capture the distribution of data.	Use it on pictures of faces and identify element-wise matching.	The relationship between the latent representation and the characteristics cannot be modeled.	[39,118]
FS-GAN [31]	Face swapping and reenactment.	Adjustments for changes in attitude and expressions in an image or a video clip.	The texture is blurred when the face recreation generator is used too frequently and the sparse landmark tracking technique fails to capture the complexity of facial emotions.	[31,110,119]

As a result, STYLE-GAN is able to create images of people's faces that are not only astonishingly lifelike and of high quality but that also provide control settings for the image's overall *style* at various levels of detail. Even if it is possible to produce pseudo-portraits that seem realistic, tiny details may point out that the photos are unreal. Karras et al. [120] proposed STYLE-GAN2 to address these flaws in STYLEGAN. They improved the generator by redesigning the regularization, multi-resolution, and normalizing approaches.

In ***ATT-GAN*** [35], a revolutionary technique, an attribute categorization constraint is added to the generated image to ensure that only the necessary characteristics are changed in the proper ways. This technique avoids limiting latent representation. Results show that ATTGAN outperforms the state-of-the-art techniques, i.e., STARGAN [33], CycleGAN [36], Ic- GAN [38], Fader Networks [121], VAE/GAN [39], in terms of realistically modifying facial characteristics. Figure 14 represents the learning framework for attention-based GAN. The formula for attention GAN where the objective for the encoder and decoder [35] is as follows:

$$\min_{G_{\text{enc}},G_{\text{dec}}} \mathcal{L}\text{enc,dec} = \lambda_1 \mathcal{L}\text{rec} + \lambda_2 \mathcal{L}\text{clsg} + \mathcal{L}\text{advg} \quad (1)$$

and the objective for the discriminator and the attribute classifier:

$$\min_{D,C} \mathcal{L}_{\text{dis,cls}} = \lambda_3 \mathcal{L}_{\text{clsc}} + \mathcal{L}_{\text{advd}} \quad (2)$$

Here, in Equations (1) and (2), G_{enc} and G_{dec} represent the encoder and decoder networks, respectively. $\mathcal{L}_{\text{rec}}$ is the reconstruction loss, which measures the difference between the input and the reconstructed output. $\mathcal{L}_{\text{clsg}}$ is the attribute classification constraint loss, which encourages the network to classify the attributes of the input correctly. $\mathcal{L}_{\text{advg}}$ is the adversarial loss for the generator (encoder–decoder), which encourages the generated output to be realistic and attribute-preserving. D is the discriminator network, which distinguishes between real and generated samples. C is the attribute classifier network, which predicts the attributes of the input. $\mathcal{L}_{\text{clsc}}$ is the attribute classification loss, which measures the difference between the predicted and true attribute labels. $\mathcal{L}_{\text{advd}}$ is the adversarial loss for the discriminator and attribute classifier, which encourages them to distinguish between real and generated samples accurately. λ_1, λ_2, and λ_3 are hyperparameters that balance the different losses in the objectives.

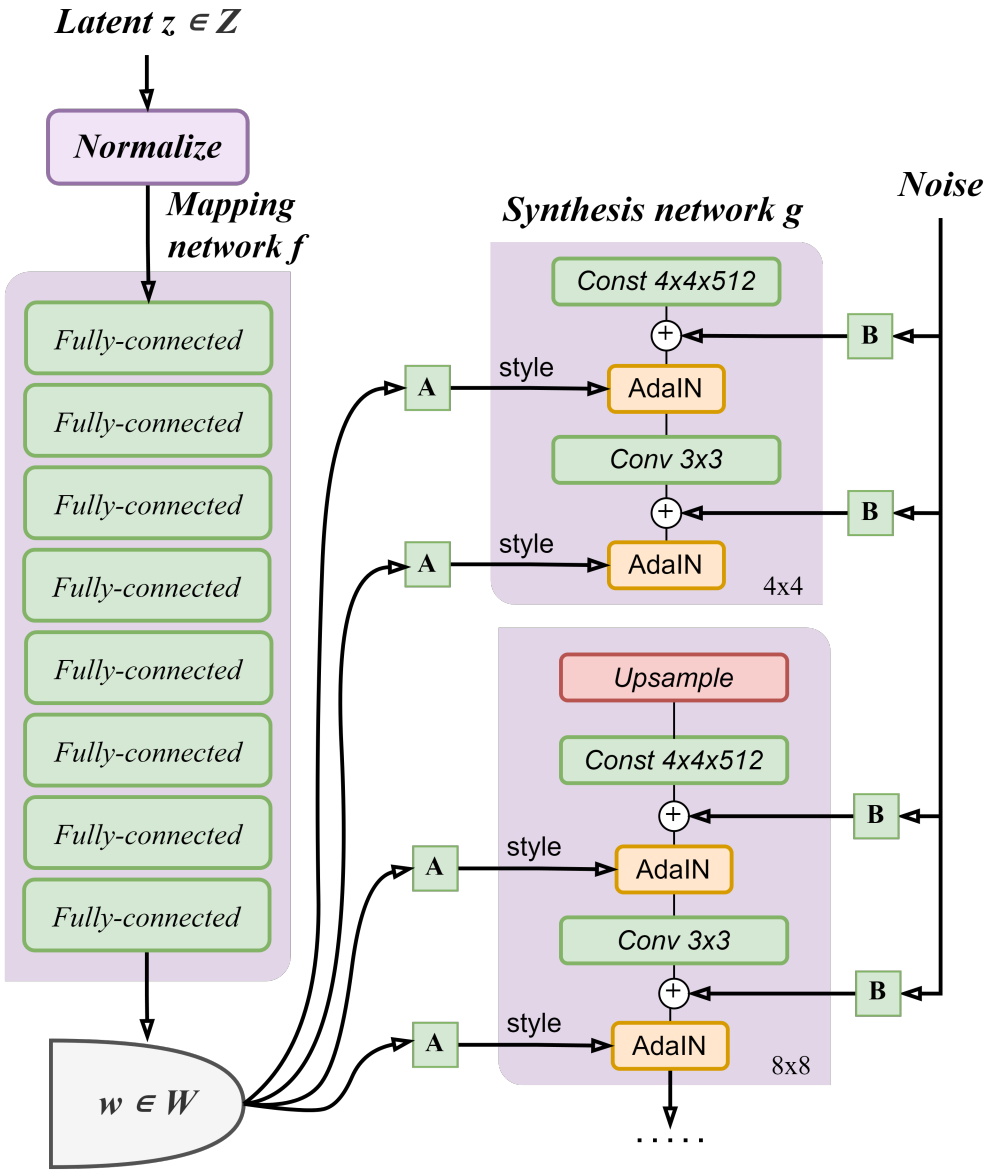

Figure 13. Style-GAN learning framework.

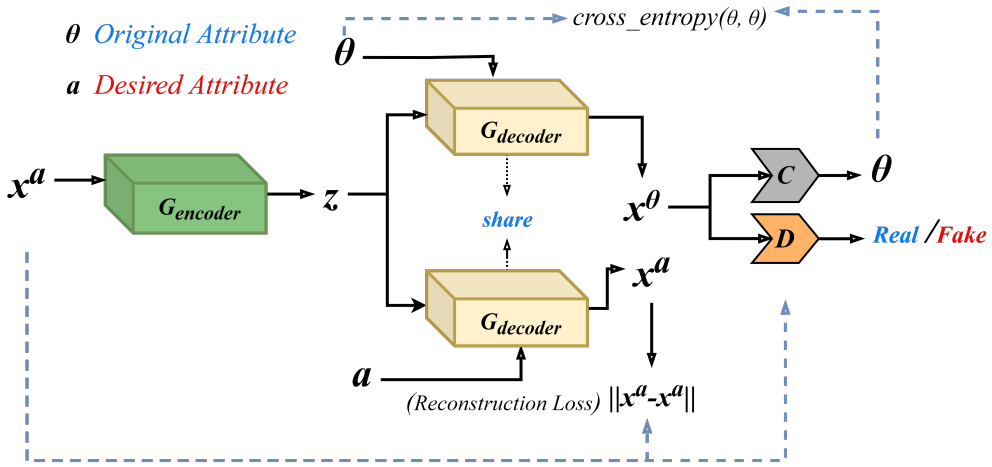

Figure 14. Attention-GAN learning framework.

6.2.2. Summary on GAN

Table 7 presents variants of GAN-based deepfake creation models. The table first shows the process of how these GAN models apply deepfake in images. These processes largely use image-to-image translation, sending semantic information to a target with a style, and make use of latent and conditional visual space, face swapping, etc. We also present major application areas such as genuine image creation, intensity control, style transfer, season transfer, etc., of these GAN-based deepfake generative models. In addition to this, we highlight major limitations of these models in the table. We also present different independent studies where these generative models have been used depending on the nature of the problem.

7. Deepfake Tools

This section covers different tools and methods utilized in creating and detecting deepfakes across various types of media, including videos, images, audio, and text. Additionally, the deepfake tools are divided into two main categories: those used for creating deepfakes and those used for detecting them.

7.1. Deepfake Creation Tools

Table 8 provides a comparison of various deepfake creation tools based on their effectiveness parameters such as accuracy, speed, usability, security, and availability. The tools listed in the table are FaceSwap-Gan, SimSwap, Fewshot FT GAN, FaceShifter, DiscoFaceGAN, Faceapp, StarGan, StarGan-V2, ATTGAN, Style-Gan, Style-Gan2, Style-Gan3, and CycleGAN.

The accuracy of the deepfake tools has been evaluated based on their ability to create high-quality deepfakes, with a high accuracy rating being indicative of better performance. SimSwap, Fewshot FT GAN, FaceShifter, DiscoFaceGAN, Faceapp, StarGan-V2, ATTGAN, Style-Gan, Style-Gan2, Style-Gan3, and CycleGAN have been rated high in terms of accuracy, while StarGan and FaceSwap-Gan have been rated moderate.

The speed of the deepfake creation tools refers to the time taken to generate the deepfakes, with a faster speed rating being better. SimSwap, FaceShifter, and CycleGAN have been rated high in terms of speed, while FaceSwap-Gan, StarGan, Style-Gan, Style-Gan2, and Style-Gan3 have been rated slow.

The usability of the tools is based on their ease of use and the availability of tutorials or documentation to assist users. Faceapp has been rated high in terms of usability, while FaceShifter and ATTGAN have been rated moderate. DiscoFaceGAN and the StarGan variants have been rated low due to the lack of documentation.

The security of the deepfake creation tools is evaluated based on their ability to prevent the creation of malicious deepfakes or to protect the privacy of users. Most of the tools listed in the table have low security ratings, with only Fewshot FT GAN and SimSwap having a moderate rating.

Finally, the availability of deepfake tools refers to their accessibility, with open-source tools being more widely available than paid tools. The table shows that most of the deepfake creation tools are open source, with only Faceapp and SimSwap being paid tools.

Table 8. Effectiveness of deepfake creation tools.

Tool	Underlying Model	Main Focus	Accuracy	Speed	Usabilty	Security	Availability	Environment
FaceSwap-GAN [122]	HEAR-Net+AEINet	Face-swapping and reenactment approach can be applied to pairs of faces.	High	Slow	Moderate	Low	Open source	TensorFlow
SimSwap [123]	Encoder-Decoder + GAN	Randomized face swapping on both still pictures and moving pictures.	High	Fast	Moderate	Low	Paid	PyTorch1.5+
Fewshot FT GAN	Few-Shot GAN	Generating faces with glasses, hair, geometric distortion, and fixed gaze using consistent faces perform poorly when converting to Asian faces.	High	Moderate	Moderate	Moderate	Paid	TensorFlow
FaceShifter [124]	HEAR-Net	Unique two-stage face-swapping method that allows for excellent accuracy and occlusion sensitivity.	High	Fast	Low	Low	Paid	Pytorch
DiscofaceGAN [125]	Disentangled StyleGAN	Can be controlled and completely disconnected using three-dimensional representational learning.	High	Slow	Low	Low	Open source	TensorFlow
Faceapp [126]		Enables changing the face, hairstyle, gender, age, as well as other physical traits.	Low	Fast	High	Low	Paid	Web
StarGAN [33]	StarGAN	Disentangled and controllable face image generation via 3D imitative-contrastive learning.	Moderate	Slow	Low	Low	Open source	Pytorch
StarGAN-V2	StarGAN-V2	Meets the requirements for various produced graphics and scalability across several domains.	High	Slow	Low	Low	Open source	Pytorch
ATTGAN [35]	Attribute-GAN	Transmission of facial characteristics under classification restrictions.	High	Moderate	Moderate	Low	Open source	TensorFlow
Style-Gan [34]	Style-Gan	Style-based GAN that produces deepfakes.	High	Slow	Low	Low	Open source	TensorFlow
Style-Gan2	Style-Gan2	Proposes weight modification, regularizes path length, modifies the generator, and drops continued expansion to enhance quality images.	High	Slow	Low	Low	Open source	TensorFlow

Table 8. *Cont.*

Tool	Underlying Model	Main Focus	Accuracy	Speed	Usabilty	Security	Availability	Environment
Style-Gan3	Style-Gan3	Better adapted for videos and animations since it is fully consistent and efficient for rotation and translation also at sub-pixel ranges.	High	Slow	Low	Low	Open source	Pytorch
CycleGAN [36]	CycleGAN	Without paired samples, converts a picture from input data X to destination domain Y.	High	Fast	Low	Low	Open source	Pytorch

Efficiency rate: High, Fast → (80–100%); Moderate → (50–79%); Low, Slow → (0–49%).

Comparative Analysis of Deepfake Creation Tools

In this discussion, we examine and compare various deepfake creation tools based on their computational efficiency, resilience against adversarial attacks, scalability, and usability. We also explore the advantages and limitations associated with these tools.

Computational efficiency

CycleGAN stands out among these techniques for its comparatively efficient computational performance. It achieves this by utilizing a cycle-consistency loss and by not requiring paired training data. CycleGAN is particularly effective in transferring facial attributes and performing image-to-image tasks. On the other hand, Style-GAN, Style-GAN2, and Style-GAN3 are well known for their impressive image-synthesis capabilities, but they tend to be computationally demanding due to their complex architecture and high-resolution image generation.

For face swapping and modification, tools such as FaceSwap-GAN, SimSwap, Fewshot FT GAN, FaceShifter, DiscoFaceGAN, and Faceapp offer varying levels of computational efficiency. SimSwap, for example, stands out for its lightweight network design, delivering good results with faster inference times.

Tools such as StarGan and StarGan-V2 excel in attribute transfer across different domains but may require moderate to high computational resources, especially for large datasets or high-resolution images. ATTGAN strikes a balance between computational efficiency and result quality, specifically designed for attribute manipulation in face photographs.

Overall, the computational efficiency of these deepfake tools depends on factors such as model complexity, dataset size, image resolution, and specific task requirements. Researchers and practitioners should consider their computing needs when selecting a tool based on available resources and desired performance levels. Through the integration of cutting-edge techniques with edge-cloud services and energy-harvesting methods, it has been observed that computational complexity can be significantly reduced (e.g., [127]).

Robustness against adversarial attack

When considering the resistance to adversarial attacks, deepfake techniques are vulnerable to such attacks due to the inherent characteristics of their generative models. Adversarial attacks can manipulate or deceive these models, compromising the credibility and authenticity of the generated content. However, there are several noteworthy observations regarding the resilience of the mentioned tools.

CycleGAN: CycleGAN is recognized for its capability to handle unpaired data and to perform image-to-image translation. Although it may lack specific defenses against adversarial attacks, its reliance on cycle-consistency loss contributes to preserving the overall integrity of the generated images to some extent.

Style-GAN, Style-GAN2, Style-GAN3: These models are extensively used for producing high-quality images but may be more susceptible to adversarial attacks. Their intricate architectures and the generation of high-resolution images make them potential targets for manipulation and alteration.

FaceSwap-GAN, SimSwap, Fewshot FT GAN, FaceShifter, DiscoFaceGAN, Faceapp: These tools primarily focus on face swapping and modification, and their resistance to adversarial attacks can vary depending on the specific techniques employed. It is important to consider the underlying defenses and precautions implemented in each tool to mitigate potential vulnerabilities.

StarGan, StarGan-V2: These models are designed for attribute transfer between different domains and may not possess explicit defenses against adversarial attacks. Their susceptibility to such attacks can vary depending on the particular implementation and the precautions taken during the training process.

ATTGAN: ATTGAN specializes in manipulating facial attributes and may incorporate certain levels of robustness against adversarial attacks. However, the specific defenses implemented may vary depending on the implementation and training strategies employed.

In summary, the robustness against adversarial attacks in deepfake tools is influenced by several factors, including the underlying architecture, training techniques, and the extent to which specific defenses are integrated. Evaluating the robustness of these tools in real-world scenarios requires careful consideration of potential vulnerabilities and countermeasures. Ongoing research in adversarial attacks and defense mechanisms continues to enhance the resilience of these tools against potential threats. On the other hand, various blockchain-based techniques (e.g., [128]) are used to protect against such kinds of issues. The decentralized and distributed characteristics of blockchain technology provide a level of resistance against tampering with the data stored on the blockchain, making it challenging for adversaries to manipulate the information.

Scalability and usability

When considering the scalability and usability of the mentioned deepfake tools, there are important aspects to consider. In terms of scalability, CycleGAN stands out, as it can handle unpaired data and perform various image-to-image translation tasks effectively. However, models such as Style-GAN, Style-GAN2, and Style-GAN3 may face scalability limitations due to their complex architectures and high-resolution image generation, demanding substantial computational resources. The scalability of face swapping and attribute manipulation tools, such as FaceSwap-GAN, SimSwap, Fewshot FT GAN, FaceShifter, DiscoFaceGAN, Faceapp, StarGan, StarGan-V2, and ATTGAN, varies depending on their specific implementation and the complexity of the tasks they handle. Regarding usability, CycleGAN is user-friendly, does not require paired training data, and accommodates unpaired image datasets seamlessly. In contrast, models such as Style-GAN, Style-GAN2, and Style-GAN3 can be more challenging to use due to their intricate architectures and advanced image-synthesis capabilities, often necessitating deep learning expertise and sufficient computational resources. The usability of face swapping and attribute manipulation tools depends on factors such as user interface, documentation, and the technical expertise level required. Ultimately, practitioners and researchers should carefully evaluate scalability and usability factors to select the most suitable deepfake tool for their specific needs and available resources.

Advantages and Limitations

Each deepfake tool has its own set of advantages and limitations, making them suitable for different use cases. It is important to consider the specific requirements, desired outcomes, and available resources when selecting the most appropriate tool for a given task. Additionally, ongoing research and advancements in deepfake technologies may lead to further improvements in the advantages and limitations of these tools. Table 9 presents a comprehensive overview of the advantages and limitations associated with deepfake creation tools across different scenarios.

Table 9. Advantages and limitations of deepfake creation tools.

Tools	Advantages	Limitations
FaceSwap-GAN	Realistic face swapping results.	Limited handling of complex facial transformations.
SimSwap	Fast inference times, lightweight design.	Performance limitations for extreme facial transformations and diverse datasets.
Fewshot FT GAN	Realistic few-shot face swapping.	Fine-tuning requirements, challenges with highly diverse face attributes.
FaceShifter	Attribute manipulation and controllable results.	Extensive training requirements, limitations in extreme facial transformations.

Table 9. *Cont.*

Tools	Advantages	Limitations
DiscoFace-GAN	Facial attribute editing with control.	Data-handling considerations, limitations with complex facial expressions.
Faceapp	User-friendly interface, a range of transformation filters.	Limited customization and fine-grained control over facial transformations.
StarGan, StarGan-V2	Attribute transfer across domains.	Computational resource requirements, challenges with high-resolution images.
ATTGAN	Balance between computational efficiency and result quality.	Hyperparameter tuning, limitations with extreme facial transformations.
Style-GAN, Style-GAN2, Style-GAN3	High-quality image synthesis, fine-grained control.	Computational intensity, limitations with large-scale datasets.
CycleGAN	Unpaired image-to-image translation, versatility.	Hyperparameter tuning, challenges in preserving fine details during image translation.

7.2. Deepfake Detection Tools

This section covers an evaluation of the effectiveness of deepfake-detection tools. We have examined several widely used and effective deepfake-detection tools in the field of deepfake.

a. ***Sensity AI.*** A deepfake detection software called Sensity AI makes use of machine learning and artificial intelligence methods to spot altered material. It is renowned for finding deepfakes quickly and precisely, even in huge datasets. To stop the spread of deepfakes, the technique has been used by a number of groups, including social networking sites and law enforcement organizations. Sensity AI is made to scan vast volumes of data rapidly and has demonstrated great levels of accuracy, with accuracy scores of up to 95%. The tool's easy-to-use layout makes it usable to both technical and lay users, and it is scalable and adaptable to various sectors and user needs. In conclusion, Sensity AI is a useful tool for recognizing and mitigating the risks of deepfakes.

b. ***Truepic.*** This is a platform that provides services to verify the authenticity of photos and videos and to detect any manipulation or tampering in the media. It offers various features, including cryptographic techniques to verify authenticity, advanced forensic analysis capabilities, real-time verification, integration with different platforms, a user-friendly interface, and high accuracy rates in detecting manipulated media. Truepic can verify media from various sources and formats, ensuring wide coverage. Additionally, it provides a transparent and auditable trail of the verification process for reliable validation of the media's authenticity.

c. ***D-ID.*** This is a tool that utilizes advanced algorithms to protect users' privacy by transforming photos and videos in a way that prevents facial recognition systems from identifying the individual in the media. The tool is effective in achieving this goal, with high accuracy rates in facial anonymization. It is easy to integrate into various applications, works quickly, and is compatible with various platforms such as iOS, Android, and web applications. Additionally, D-ID's algorithm is robust against attacks by adversarial machine learning techniques, ensuring that anonymization remains effective even in the face of such attacks.

d. ***Amber Video.*** This is a deepfake-detection tool that is known for its high accuracy in detecting deepfakes, particularly those created using generative models. The tool's effectiveness parameters include the use of multiple levels of detection, continuous learning to adapt to emerging threats, real-time detection with scalability, a cloud-based solution for easy deployment and integration, and customization options for detection rules and thresholds to meet specific needs.

e. ***Deeptrace.*** This is a tool for detecting deepfake videos that uses machine learning algorithms. The tool is effective in detecting deepfakes due to its high accuracy rate, which is achieved through the use of state-of-the-art machine learning algorithms. It can detect deepfakes in real-time, making it useful for identifying fake videos as they are being shared. Deeptrace uses a combination of audio-, visual-, and text-based analysis to detect deepfakes, which makes it more comprehensive than other tools. The tool is scalable and can analyze large datasets, meeting the needs of different organizations. Additionally, Deeptrace continuously learns and adapts to new deepfake techniques, ensuring that it can effectively detect the latest types of deepfakes.

f. ***FooSpidy's Fake Finder.*** This is a deepfake-detection tool that uses image forensics and deep neural networks to identify manipulations in images and videos. Its effectiveness parameters include high accuracy, real-time detection, a user-friendly interface, and customizable settings. The tool has an accuracy rate of over 90% and allows users to quickly upload and scan media files for deepfakes. Additionally, users can customize detection settings to enhance accuracy and precision in identifying deepfakes.

g. ***DeepSecure.ai.*** This is a deepfake-detection tool that uses a unique approach of analyzing the semantic content of videos to detect deepfakes. Some of its effectiveness parameters include high accuracy, real-time speed, scalability, ease of use, and versatility. It has a reported accuracy rate of 96% in detecting deepfakes and can handle large volumes of videos. Additionally, it can detect a wide range of deepfake techniques, including face-swapping and voice cloning, among others. The tool is user-friendly and can be easily integrated into existing workflows, making it accessible to a wide range of users.

h. ***HooYu,*** This is a platform for digital identity verification that offers fraud protection, customer onboarding, and identity authentication solutions. To assure high accuracy, the platform's verification technology makes use of a variety of verification methods and data sources. HooYu's automatic verification procedure is rapid and effective, allowing companies to quickly onboard clients. The platform is built to provide clients with a seamless and user-friendly experience while adhering to regulatory regulations. It is adaptable and may be tailored to fit the unique requirements of different businesses, including e-commerce and financial services.

i. ***iProov.*** This is a tool designed to detect deepfakes by using a proprietary technology called Flashmark to identify any signs of manipulation in facial biometric data. This tool's effectiveness lies in its ability to accurately detect even the most sophisticated deepfake attempts, providing real-time verification of users' faces, and it has a user-friendly interface. It is platform-agnostic, working seamlessly across iOS, Android, and web browsers, with a high level of trust from various government agencies that use it to verify identities and prevent fraud.

j. ***Blackbird.AI.*** This is a tool designed for deepfake detection that combines machine learning algorithms with human intelligence to identify and classify manipulated media. Its effectiveness parameters include a high detection accuracy of 98%, advanced machine learning algorithms to detect signs of manipulation, and real-time monitoring of various online sources. Blackbird.AI also uses a team of trained analysts to review flagged videos and confirm their authenticity. Users can customize their settings to meet their specific needs, such as setting the threshold for deepfake detection

or excluding certain sources from monitoring. Overall, Blackbird.AI is a reliable and effective solution to combat the proliferation of manipulated media online.

k. ***Cogito***. This is an AI-based behavioral analytics platform that uses machine learning algorithms to identify and prevent deepfakes in real time. Its effectiveness parameters include high accuracy and precision in detecting even the most advanced deepfakes, real-time detection capabilities for immediate flagging and minimization of potential harm, user-friendly and seamless integration into business workflows, customization of deepfake detection protocols, and scalability to address evolving threats and challenges. Overall, Cogito's platform is an effective and adaptable solution for businesses seeking to combat the spread of manipulated media online.

l. ***Veracity.ai.*** This is a deepfake-detection tool that uses algorithms based on ML and AI to accurately pinpoint media that has been altered. The program is especially helpful for use in live video streams since it can identify deepfakes in real-time. In order to provide thorough coverage, it can also identify deepfakes across a variety of modalities, including video, audio, and pictures. Veracity.ai is simple to use and does not require any technical knowledge to use. To further its efficiency, it is additionally regularly updated with the most recent deepfake detecting methods and methodologies.

m. ***XRVision Sentinel.*** A deep learning technology called XRVision Sentinel analyzes the structure and composition of facial photos and videos to detect those that have been altered. It is capable of spotting deep fakes in a variety of contexts, including political campaigns, news media, and social media. Advanced machine learning techniques are used by the instrument to identify minute variations in facial expressions, lip movements, and eye movements. Additionally, it has the ability to recognize deep fakes created using a variety of methods, such as GAN-based models and facial reenactment techniques. Testing of XRVision Sentinel on datasets such as the Deepfake Detection Challenge dataset showed that it had a high degree of accuracy in identifying deepfakes and a low proportion of false positives.

n. ***Amber Authenticate.*** This is a tool that utilizes cryptographic techniques to validate the authenticity of image and video content and to prevent the spread of deepfakes. The tool has been shown to be highly accurate and efficient in detecting deepfakes in real time. It is compatible with various media file formats and has a user-friendly interface, making it easy for users of all technical levels to operate. Overall, Amber Authenticate is a versatile and reliable tool for detecting deepfakes across various platforms and applications.

o. ***FaceForensics++.*** A deepfake detection program called FaceForensics++ focuses on identifying face exchanges in videos. It is highly effective at identifying deepfakes by studying minute facial movements and has the capacity to recognize deep fakes produced using a variety of techniques. FaceForensics++ features an easy-to-use interface that requires little technical knowledge, is open-source, and may be freely used and modified by researchers. Finally, it is a scalable tool that can be incorporated into real-time deepfake detection systems because it can effectively analyze large datasets of videos.

p. ***FakeSpot.*** This is an AI-powered web-based tool that can identify fake reviews on e-commerce websites with an accuracy rate of 90%. It can analyze thousands of reviews in seconds and is easy to use, even for non-technical users. FakeSpot also offers a browser extension that can be installed on Chrome, Firefox, and Safari, making it more accessible to users. Additionally, it is compatible with multiple e-commerce platforms including Amazon, Yelp, TripAdvisor, and Walmart, making it a versatile tool for detecting fake reviews across various websites.

Comparative Analysis of Deepfake Detection Tools

In this section, we compare the tools discussed above in terms of computational efficiency, scalability, robustness against adversarial attacks, and usability. Table 10 provides

an overview of various deepfake-detection tools and their effectiveness parameters. Each tool is evaluated based on detection accuracy, speed, user-friendliness, scalability, and integration. Most of the tools show high accuracy in detecting deepfakes, with fast speed and easy-to-use interfaces. They are scalable and can be integrated using APIs and SDKs. FaceForensics++, an open-source tool, shows high detection accuracy but slow speed, and its use requires technical expertise. FakeSpot, a web-based tool, has moderate detection accuracy but fast speed, with an easy-to-use interface and high scalability. Overall, the table demonstrates the effectiveness of these deepfake-detection tools, which can be useful in combating the spread of fake content. We can consider the following insights.

Computational Efficiency

Sensity AI, D-ID, Amber Video, Deeptrace, and XRVision Sentinel are known for their efficient computational capabilities. They employ advanced algorithms and optimizations to process and analyze large volumes of data efficiently. Tools such as Truepic, HooYu, and Veracity.ai also prioritize computational efficiency but may not offer the same level of optimization as the aforementioned tools.

Scalability

Scalability can vary among these tools depending on their architecture and infrastructure. Tools such as Sensity AI, Amber Video, and Deeptrace have built scalable platforms that can handle high volumes of data and scale with increasing demands. Truepic, XRVision Sentinel, and Veracity.ai also focus on scalability, providing solutions that can adapt to varying data volumes and user requirements.

Robustness against Adversarial Attacks

Robustness against adversarial attacks refers to the ability of the tools to detect and mitigate attempts to manipulate or deceive the system. Tools such as Deeptrace, XRVision Sentinel, and FaceForensics++ have advanced techniques to detect deepfakes and other forms of manipulated media, showcasing strong robustness against adversarial attacks. Truepic, iProov, and Amber Authenticate also prioritize robustness by implementing various verification and authentication mechanisms.

Usability

Usability considers how user-friendly and intuitive the tools are in terms of their interfaces, integration capabilities, and ease of adoption. Tools such as Truepic, HooYu, and Veracity.ai focus on providing user-friendly interfaces and seamless integration options for easy adoption and usage. D-ID, DeepSecure.ai, and Cogito also prioritize usability by offering intuitive workflows and easy-to-understand features.

Advantages and Limitations

In the preceding discussion, we have examined the benefits and drawbacks of the deepfake-detection tools mentioned earlier.

Sensity AI is renowned for its deepfake detection and proactive monitoring capabilities, while Truepic specializes in image and video verification for supply chain authentication. D-ID focuses on anonymizing and securing facial images to protect personally identifiable information (PII), and Amber Video offers AI-powered video verification and real-time monitoring. Deeptrace provides comprehensive deepfake detection and analysis with advanced attribution capabilities, while FooSpidy's Fake Finder offers customizable solutions for specific industries. DeepSecure.ai specializes in securing AI models from adversarial attacks, HooYu focuses on identity verification, and iProov offers biometric authentication. Blackbird.AI focuses on disinformation detection, Cogito offers emotional analysis and sentiment analysis, and Veracity.ai provides media verification and fact-checking solutions. XRVision Sentinel specializes in deepfake detection and video analytics. Limitations include limited coverage of different types of media manipulation, narrow focuses on specific areas, scalability constraints, and potentially higher implementation costs for enterprise-level solutions.

Table 10. Effectiveness of deepfake detection tools.

Tools	Accuracy	Speed	User-Friendliness	Scalability	Integration
Sensity AI [129]	High	Fast	Easy	High	API, SDK
Trupic [130]	High	Fast	Easy	High	API
D-DI [131]	High	Fast	Easy	High	API
Deeptrace [132]	High	Fast	Moderate	High	API
DeepSecure.ai [133]	High	Fast	Easy	High	API
iProov [134]	High	Fast	Easy	High	API
Blackbird AI [135]	High	Fast	Easy	High	API
XRVision Sentinel [136]	High	Fast	Easy	High	API
Amber Authenticate [137,138]	High	Fast	Easy	High	API
FaceForensics++ [139]	High	Fast	Easy	Low	Open-source
FakeSpot [140]	Moderate	Fast	Easy	High	Web-based

Efficiency rate = High, Fast → (80–100%), Moderate → (50–79%), Low, Slow → (0–49%).

8. DL Based-Deepfake Tools

In order to create deepfake photos and videos, a variety of machine learning approaches have been applied. However, these methodologies have only been studied for research purposes and thus only exist in theory. On the contrary, a wide range of deepfake generating and detecting tools are available. However, we did not find any that make use of conventional machine learning models. These tools are based on different popular deep learning models because deep learning models can easily outperform conventional machine learning models.

From the literature, we have observed that CNN, RNN, GAN, and transformer-based models are mostly used in deepfake-detection tools. Therefore, in the taxonomy of deepfake tools (shown in Figure 15), we categorized the tools into four categories. Since traditional machine learning models are not used in detection tools, there is no branch for that category. From Figure 15, we can also observe that CNN and GAN are mainly used for developing these tools. Hence, in the following, we summarize a variety of CNN and RNN based deepfake tools.

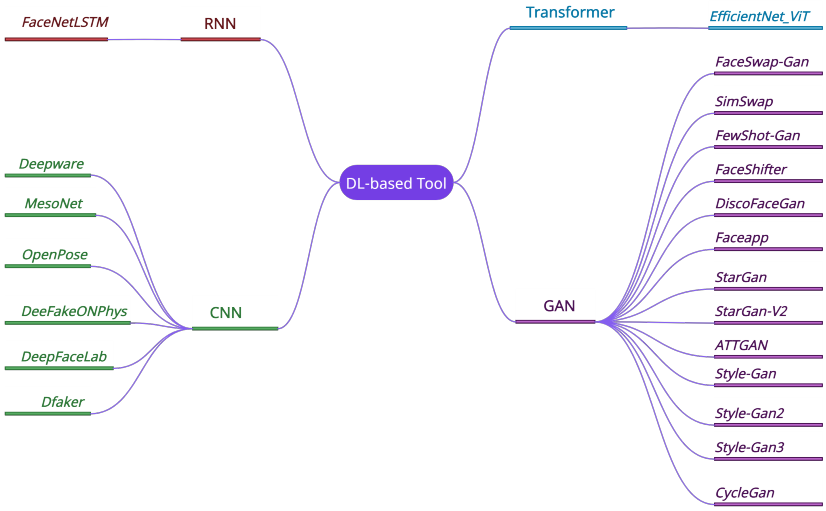

Figure 15. Taxonomy of deep-learning-based deepfake tools.

8.1. CNN-Based Tools

We discuss a few of the popular CNN-based modeling tools that we located on the internet.

- Face2Face [2,141] shows a much more accurate real-time facial emotion exchange from a source to a target film. It displays the effects of live manipulation of a target YouTube video using a webcam-captured source video stream. Additionally, it is compared to cutting-edge reenactment techniques and exceeds in terms of the final video quality and run-time.
- Face swapping is carried out by FaceSwap [3] using picture blending, Gauss–Newton optimization, and a deep neural network-based face alignment. The detected face and features for a given input photo are first found by the algorithm. Additionally, a 3D model matches the features whose edges are mapped to the picture space and are transformed into textural positions.
- Deepfake Faceswap [142] is a platform for swapping face applications that consist of a set of encoder–decoder-based deep learning models. The goal of developing FaceSwap is to reduce its abuse potential while enhancing its usefulness as a tool for research, experimentation, and legal face swapping.
- FaceSwap_Nirkin [143,144] is an automatic image-swapping tool. It demonstrates that, rather than designing algorithms specifically for face segmentation, a typical fully convolutional network (FCN) can perform amazingly quick and precise segmentation if trained on a large enough number of rich sample sets. It makes use of specialized image segmentation to provide face identification under unusual circumstances, to fit 3D facial features, and to assess the impact of intra-subject and inter-subject face swapping on identification. It gives a face-swapping accuracy of around 98.12% in the COFW dataset [145].
- Deepware Scanner [6] is a deepfake-detection tool that produces results on a variety of deepfake sets of data, together with natural deepfake and actual videos. Here, an EfficientNet B7 [57,146] model that has been pre-trained on the ImageNet dataset is used, and the classification algorithm is trained using only Facebook's DFDC [110] dataset, which contains 120k videos. Then, the model is trained to work in production, with an emphasis on fewer false positives. The model is a frame-based classifier, which means it does not take into account temporal coherence. Because video is a temporal medium, we believe this is a significant shortcoming that must be addressed.
- DFace [7,147] is a face-recognition and identification toolkit with attention to efficiency and usability. With certain enhancements, most importantly on storage overflows, this is a narrowed version of Timesler's FaceNet [148] (constructed using Inception Resnet (V1) models that have undergone VGGFace2 and CASIA-Webface pretraining) repository. FaceNet is employed to create facial embeddings, and MTCNN [149] is applied to detect faces.
- MesoNet [8] is a compact facial video detection techniques network. In [52], they examined a technique for dynamically spotting altered faces in video recordings. Deepfake and Face2Face are two contemporary methods used to produce forged clips that are incredibly lifelike. Clips typically do not lend themselves well to classical visual forensic approaches because of how tightly compressed they are, which severely affects the data. As a result, they use deep learning and build two networks with a few layers each to concentrate on the mesoscopic characteristics of the image. Utilizing both a new dataset and an existing dataset we created from web videos, we evaluated those rapid networks. For Face2Face and deepfake, our testing shows a success rate of over 98% and 95%, respectively.
- OpenPose [150] is the initial genuine multi-person technology that identifies 135 feature points overall on the facial, human body, hand, and foot feature points on a single image. Zhe Cao et al. [151] offer a real-time method for spotting numerous 2D poses in a picture. The method learns to link parts of bodies with persons in the picture, which is represented using a nonparametric method known as part affinity

fields (PAFs). No matter how many people are in the image, our bottom-up approach delivers great accuracy and real-time performance. In earlier research, PAFs and part of the body position measure were improved concurrently during the training stages. Here, PAF-only refinement rather than PAF and part of the body position adjustment leads to a significant improvement in accuracy and runtime efficiency. Initially, an integrated body and foot keypoint detector is also presented, and it is based on a privately published internal foot dataset. In the end, it trains a multi-stage CNN model with a deepfake-detection accuracy of about 84.9%, utilizing data from the COCO 2016 keypoints challenge [152] and MPII human multi-person [153] datasets.

- A framework called DeepfakesONPhys [154] uses a physiological assessment to identify deepfake. Utilizing remote photoplethysmography, it specifically takes into account data regarding the heartbeat (rPPG). In order to more effectively identify fraudulent films, DeepfakesON-Phys employs a convolutional attention network (CAN) [55], which pulls out both spatial and temporal data from video sequences. Utilizing the most recent open datasets in the industry, Celeb-DF and DFDC, it has been systematically assessed. The findings obtained, approximately 98% AUC with both datasets, surpass the current state of science and demonstrate the effectiveness of physiologically based fake classifiers for spotting the most recent Deepfake movies.
- For the purpose of detecting deep fake videos, EfficientNet_ViT [57,155] combines EfficientNet and Vision Transformer. The method does not employ either distillation or ensemble methods, in contrast to cutting-edge techniques. In addition, the technique provides a simple voting-based inference approach for managing many faces in a single video frame. On the DFDC dataset, the best model had an accuracy of 95.10%.
- DeepFaceLab [25,156] is the most popular program for making deepfakes. DeepFaceLab is used to make more than 95% of deepfake videos and is used by well-known YouTube and TikTok channels (e.g., deeptomcruise [157], arnoldschwarzneggar [158], diepnep [159], deepcaprio [160], VFXChris Ume [161], Sham00k [162], NextFace [163], Deepfaker [164], Deepfakes in movie [165], DeepfakeCreator [166], Jarkan [167]). It is possible to substitute faces, reverse aging, replace heads, and even manipulate politicians' lips using this tool. S3FD [168] is used to detect faces in DFL, 2DFAN [169] and PRNet [170] are used to align faces in DFL, and a fine-grained Face-Segmentation network (TernausNet [171]) is used to segment faces. To train the DFL model, the FF++ dataset is used and gains deepfake detection accuracy of around 99% better than Face2Face, FaceSwap, and deepfake.
- FakeApp [54,172] is a computer program that enables the production of what is now referred to as "deepfakes".
- The Deepfakesweb [173] app is a cloud-based deepfake tool. This app handles everything else; the user only needs t upload clips and photos and then press a button. This app allows the model to be used again after training. By doing so, users can create new films or enhance the outcome's face-swapping quality without having to train a model again. The excellence and duration of the films determine how good a deepfake is.

8.2. GAN-Based Tools

Here, we have covered some of the most popular GAN-based modeling tools we could find online.

- FaceShifter [29,124] is a special two-stage face-swapping technique for high accuracy and occlusion-sensitive face swapping. In contrast to earlier techniques, it can handle facial occlusions utilizing a second synthesis step that consists of a heuristic error acknowledging refinement network (HEAR-Net).
  It is capable of producing high-quality identity-preserving face-swapping outcomes. First, a high-quality face-swapping outcome is produced using an adaptive embedding integration network (AEINet) based on the integration of information. Then, it produces a heuristic error recognizing network (HEAR-Net) to handle the difficult facial occlusions. The datasets CelebA-HQ [174], and FFHQ [34] are used to train

AEI-NET. HEAR-Net, on the other hand, makes use of the faces' upper half. It gives a fake classification accuracy of around 97.38%.

- SimSwap [30,123] is a highly effective face-swapping tool. In order to effectively aid their system in implicitly preserving the face attributes, Simswap presents the Weak Feature Matching Loss. According to experimental findings, they are more capable of preserving qualities than earlier state-of-the-art techniques. It employs a GAN-based model and trains the model using VGGFace2 [175] and FF++ [54] datasets. Encoder, ID Injection Module (IIM), and Decoder are the three components that make up the generator. It produces deepfakes with a 96.57 percent accuracy rate.
- FaceSwap-GAN [31,122] is one of the GAN-based deepfake-detection techniques. It can produce accurate and reliable eye motions, and it produces videos with better face orientation and improved quality. With a deepfake-detection accuracy of over 99%, it uses the Segmentation CNN + Recurrent Reenactment Generator model and is trained on the FF++ [54] dataset.
- DiscoFaceGAN [32,125] is a technique used for producing synthetic faces of people using perfectly adjustable, completely separated latent representations of their identity, appearance, position, and lighting. Adversarial learning is used in this case to incorporate three-dimensional priors, and the network is trained to mimic the picture creation of an analytical three-dimensional facial image modification and rendering procedure.
- Faceapp [126] is a mobile application that enables users to make customized deepfake videos. This is a deep-learning-based tool that uses cycleGAN as a model with extremely accurate results. It simply creates amazingly realistic facial changes for pictures. Using a mobile phone, it is possible to alter the face, haircut, age, gender, as well as other characteristics.

Table 11 shows specific models used with tools such as EfficientNet of CNN model, HEAR-Net, etc. The table also presents popular datasets and shows the accuracy of these tools.

Table 11. Accuracy of some deepfake tools based on models and datasets.

Deepfake Tools	Model Used in This Tool	Dataset	Accuracy
FaceSwap_Nirkin [144]	Fully Convolutional Network (FCN)	COFW	98.12%
FaceSwap-Gan [31]	Segmentation CNN + Recurrent Reenactment Generator	FF++, Figar	99.00%
Deepware [6]	EfficientNet B7	FF++ DFDC 4000 DFDC 5000 Celeb-DF Real Celeb-DF Fake Celeb-DF YouTube FaceForensics Deepfake-Detection FaceForensics Actors FaceForensics Deepfakes FaceForensics YouTube	99.26% 87.10% 91.30% 93.10% 85.60% 89.00% 84.20% 99.70% 92.70% 90.90%
MesoNet [8]	CNN(Meso-4) CNN(MesoInception-4 [8])	FaceForensics FaceForensics	89.10% 91.70%
OpenPose [150]	multi-stage CNN (1st stage → 10 layers VGG-19)	MPII, COCO 2016	84.90%
DeepfakesONPhys [154]	Convolutional Attention Network (CAN)	Celeb-DF, DFDC	98.00%
FaceShifter [124]	HEAR-Net + AEINet	CelebA-HQ, FFHQ, VGGFace	97.38%

Table 11. *Cont.*

Deepfake Tools	Model Used in This Tool	Dataset	Accuracy
SimSwap [123]	GAN → Generator (Encoder, ID Injection Module (IIM), and Decoder) + Discriminator	VGGFace2, FF++	96.57%
EfficientNet_ViT [57]	EfficientNet + Vision Transformer	DFDC, FF++	95.10%
DeepFaceLab [25]	S3FD + 2DFAN + PRNet + TernausNet	FF++	99.00%

9. Challenges

During the development of this study, we encountered a number of difficulties, which is covered in this section. Figure 16 represents five crucial issues and challenges in deepfake.

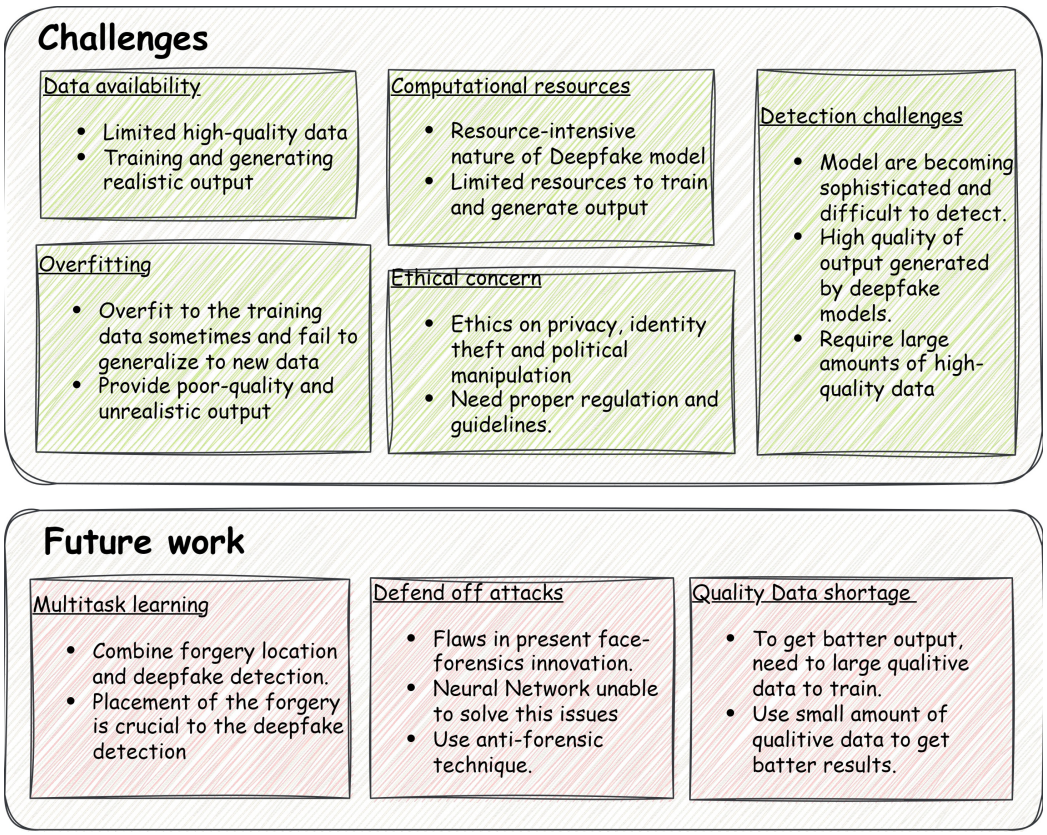

Figure 16. Summary of challenges and future works.

- It is associated with the body of research. In this research, we compile the related papers from various conferences, journals, websites, and archives of numerous e-libraries. There is still a chance that our database of studies lacks some of the relevant papers. Additionally, we might have made a few errors while categorizing these experiments using the selection or rejection criteria we employed. In order to remedy such inaccuracies, we double-checked our evaluation of the papers in our collection.
- When encountering low-quality films compared to high-resolution videos, detection algorithms frequently show a performance decline. Videos may also undergo procedures such as picture reshaping and rotations in addition to compression techniques. When constructing detection algorithms, flexibility becomes a crucial quality that must be taken into account.

- When used in a real-world setting, time consumption assumes substantial significance. Deepfake-detection techniques will be broadly applied to media services in the near future to minimize the harm that deepfake films cause to social security. Moreover, because of their extensive time requirements, existing detection techniques are still far from being widely used in real-world situations.
- When we wish to create deepfake movies using a character, deepfake models are frequently trained on a specific collection of datasets, but the model is unable to produce an accurate output since there is not enough data for this character. Finding sufficient information for a single character, however, could be challenging. It takes time to retrain the model to recognize each distinct target.
- The majority of datasets are developed in highly favorable conditions (such as ideal lighting, flawless facial expression, high-quality photographs or videos, etc.), but in the testing phase, we give data that do not keep this quality. This makes dataset quality one of the challenging areas.

Despite the variety of deepfake-generation tools available, they are not flawless. In actuality, the tools at hand are specially created and solely concentrate on specific traits. Because of the above difficulties, developing deepfake-generation tools needs additional study to boost efficiency. Consequently, creating a deepfake-generation tool is a difficult process.

10. Open Research and Future Work

We also foresee several potential research possibilities for deepfake generation and detection to solve issues with the methods now in use. Figure 16 represents three main future scopes for deepfake to develop.

- Because of flaws in present face-forensic innovation, antiforensic technology has been invented. Neural networks are frequently employed in the area of deepfake detection to identify fake videos. Neural networks are unable to fend off attacks from adversarial samples because of inherent flaws [176]. Researchers must develop more flexible strategies that can withstand prospective threats that are identified in order to prevent these attacks in certain situations.
- It has been demonstrated that multitask learning, which involves carrying out several tasks at once, improves prediction performance when compared to single-task learning. It has been discovered that combining forgery location and deepfake-detection tasks can increase deepfake-detection task accuracy. The model may complete two jobs at once while taking into account the losses incurred by each, significantly enhancing the performance of the model. The authors in [177,178] demonstrate how the placement of a forgery is crucial to the deepfake detection challenge. Thus, there is a lot of room for improving deepfake detection using multitasking.
- One of the key areas where researchers can focus their research is to enhance deepfake photo- or video-generation models that cannot produce deepfake movies more realistically using one or a few photographs or videos. Because the majority of deepfake creation models are trained on such a qualitative dataset, users face many challenges in managing many quality images or videos as the testing or source data.

In summary, the advancements in deepfake technology have highlighted the need for more robust detection methods to counter the flaws in existing face-forensic technology. Antiforensic techniques have emerged to exploit the vulnerabilities of neural networks used for deepfake detection. Multitask learning has shown promise in improving detection accuracy by combining forgery location and deepfake detection tasks. Enhancing deepfake-generation models to create more realistic videos from limited source data is another important research direction. Overall, addressing these challenges is crucial to effectively detect and mitigate the risks associated with deepfakes.

11. Conclusions

Deep-learning-based falsified innovations have been growing at an entirely unexpected rate during this period. The worldwide spread of the Internet makes it possible for illegal face-altered films produced by deepfake technologies to spread quickly, harming social stability and individual rights. In order to mitigate the harmful effects of deepfake films on individuals, business enterprises and various scientific organizations across the globe are conducting a significant amount of studies. In this paper, we discussed various deepfake models and the models that are employed in the development of well-known online deepfake tools. Here, we presented examples of numerous well-known deepfake tools, together with their traits, accuracy of deepfake models and tools, and model-based taxonomy. Finally, we covered the existing issues, gave insights into unresolved problems, and addressed the next research on deepfake production and detection technologies.

Author Contributions: Conceptualization, M.S.H.M., S.A., S.I., M.J. and M.E.A.; methodology, M.S.H.M., J.A. and M.A.K.R.; validation, M.S.H.M., S.A., S.I., M.E.A. and M.J.; formal analysis, M.S.H.M., J.A. and M.A.K.R.; investigation, M.S.H.M., J.A. and M.A.K.R.; writing—original draft preparation, M.S.H.M., J.A. and M.A.K.R.; writing—review and editing, M.S.H.M., S.A., S.I., M.E.A. and M.J.; supervision, M.S.H.M., S.A., S.I., M.E.A. and M.J.; project administration, M.S.H.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research is funded by the Institute for Advanced Research Publication Grant of the United International University, ref. no. IAR-2023-Pub-016.

Data Availability Statement: Not applicable.

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

References

1. Westerlund, M. The emergence of deepfake technology: A review. *Technol. Innov. Manag. Rev.* **2019**, *9*, 39–52. [CrossRef]
2. 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.
3. Kowalski, M. FaceSwap. 2021. Available online: https://github.com/MarekKowalski/FaceSwap (accessed on 6 November 2022).
4. Singh, R.; Shrivastava, S.; Jatain, A.; Bajaj, S.B. Deepfake Images, Videos Generation, and Detection Techniques Using Deep Learning. In *Machine Intelligence and Smart Systems: Proceedings of MISS 2021*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 501–514.
5. Vaccari, C.; Chadwick, A. Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Soc. Media Soc.* **2020**, *6*, 2056305120903408. [CrossRef]
6. Mertyyanik. Deepware Scanner (CLI). 2020. Available online: https://github.com/deepware/deepfake-scanner (accessed on 6 November 2022).
7. Dodobyte. dFace. 2019. Available online: https://github.com/deepware/dface (accessed on 6 November 2022).
8. DariusAf. MesoNet. 2018. Available online: https://github.com/DariusAf/MesoNet (accessed on 6 November 2022).
9. Mirsky, Y.; Lee, W. The creation and detection of deepfakes: A survey. *ACM Comput. Surv. CSUR* **2021**, *54*, 1–41. [CrossRef]
10. Ahmed, S.R.; Sonuç, E.; Ahmed, M.R.; Duru, A.D. Analysis survey on deepfake detection and recognition with convolutional neural networks. In Proceedings of the 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), Virtual, 9–11 June 2022; pp. 1–7.
11. Nguyen, T.T.; Nguyen, Q.V.H.; Nguyen, D.T.; Nguyen, D.T.; Huynh-The, T.; Nahavandi, S.; Nguyen, T.T.; Pham, Q.V.; Nguyen, C.M. Deep learning for deepfakes creation and detection: A survey. *Comput. Vis. Image Underst.* **2022**, *223*, 103525. [CrossRef]
12. Kugler, M.B.; Pace, C. Deepfake privacy: Attitudes and regulation. *Nw. UL Rev.* **2021**, *116*, 611. [CrossRef]
13. Gerstner, E. Face/off: "Deepfake" face swaps and privacy laws. *Def. Couns. J.* **2020**, *87*, 1.
14. Harris, K.R. Video on demand: What deepfakes do and how they harm. *Synthese* **2021**, *199*, 13373–13391. [CrossRef]
15. Sharma, M.; Kaur, M. A Review of Deepfake Technology: An Emerging AI Threat. *Soft Computing for Security Applications*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 605–619.
16. Woo, S. ADD: Frequency Attention and Multi-View Based Knowledge Distillation to Detect Low-Quality Compressed Deepfake Images. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 20–27 February 2022; Volume 36, pp. 122–130.
17. Lyu, S. Deepfake detection: Current challenges and next Steps. In Proceedings of the 2020 IEEE International Conference on Multimedia & Expo Workshops (ICMEW), London, UK, 6–10 July 2020; pp. 1–6.

18. 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.
19. Felixrosberg. FaceDancer. 2021. Available online: https://github.com/felixrosberg/FaceDancer (accessed on 27 November 2022).
20. Rosberg, F.; Aksoy, E.E.; Alonso-Fernandez, F.; Englund, C. FaceDancer: Pose-and Occlusion-Aware High Fidelity Face Swapping. *arXiv* **2022**, arXiv:2210.10473.
21. Kingma, D.P.; Welling, M. An introduction to variational autoencoders. *Found. Trends Mach. Learn.* **2019**, *12*, 307–392. [CrossRef]
22. Khalid, H.; Woo, S.S. OC-FakeDect: Classifying deepfakes using one-class variational autoencoder. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 656–657.
23. Du, M.; Pentyala, S.; Li, Y.; Hu, X. Towards generalizable deepfake detection with locality-aware autoencoder. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Virtual Event, 19–23 October 2020; pp. 325–334.
24. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *Commun. ACM* **2020**, *63*, 139–144. [CrossRef]
25. Perov, I.; Gao, D.; Chervoniy, N.; Liu, K.; Marangonda, S.; Umé, C.; Dpfks, M.; Facenheim, C.S.; RP, L.; Jiang, J.; et al. DeepFaceLab: Integrated, flexible and extensible face-swapping framework. *arXiv* **2020**, arXiv:2005.05535.
26. Harris, D. Deepfakes: False pornography is here and the law cannot protect you. *Duke L. Tech. Rev.* **2018**, *17*, 99.
27. Chen, T.; Kumar, A.; Nagarsheth, P.; Sivaraman, G.; Khoury, E. Generalization of Audio Deepfake Detection. In Proceedings of the Odyssey 2020, The Speaker and Language Recognition Workshop, Tokyo, Japam, 1–5 November 2020; pp. 132–137.
28. Pilares, I.C.A.; Azam, S.; Akbulut, S.; Jonkman, M.; Shanmugam, B. Addressing the challenges of electronic health records using blockchain and ipfs. *Sensors* **2022**, *22*, 4032. [CrossRef]
29. Li, L.; Bao, J.; Yang, H.; Chen, D.; Wen, F. Faceshifter: Towards high fidelity and occlusion aware face swapping. *arXiv* **2019**, arXiv:1912.13457.
30. 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.
31. 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.
32. Deng, Y.; Yang, J.; Chen, D.; Wen, F.; Tong, X. Disentangled and Controllable Face Image Generation via 3D Imitative-Contrastive Learning. In Proceedings of the IEEE Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020.
33. 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 Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 8789–8797.
34. 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, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 4401–4410.
35. 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]
36. 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.
37. Cho, W.; Choi, S.; Park, D.K.; Shin, I.; Choo, J. Image-to-image translation via group-wise deep whitening-and-coloring transformation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 10639–10647.
38. Perarnau, G.; van de Weijer, J.; Raducanu, B.; Álvarez, J.M. Invertible conditional gans for image editing. *arXiv* **2016**, arXiv:1611.06355.
39. Larsen, A.B.L.; Sønderby, S.K.; Larochelle, H.; Winther, O. Autoencoding beyond pixels using a learned similarity metric. In Proceedings of the International Conference on Machine Learning, New York, NY, USA, 24–26 June 2016; pp. 1558–1566.
40. Kaliyar, R.K.; Goswami, A.; Narang, P. DeepFakE: Improving fake news detection using tensor decomposition-based deep neural network. *J. Supercomput.* **2021**, *77*, 1015–1037. [CrossRef]
41. Narayan, K.; Agarwal, H.; Mittal, S.; Thakral, K.; Kundu, S.; Vatsa, M.; Singh, R. DeSI: Deepfake Source Identifier for Social Media. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 2858–2867.
42. Agarwal, H.; Singh, A.; Rajeswari, D. Deepfake Detection using SVM. In Proceedings of the 2021 Second International Conference on Electronics and Sustainable Communication Systems (ICESC), Coimbatore, India, 4–6 August 2021; pp. 1245–1249.
43. Fagni, T.; Falchi, F.; Gambini, M.; Martella, A.; Tesconi, M. TweepFake: About detecting deepfake tweets. *PLoS ONE* **2021**, *16*, e0251415. [CrossRef]
44. Durall, R.; Keuper, M.; Pfreundt, F.J.; Keuper, J. Unmasking deepfakes with simple features. *arXiv* **2019**, arXiv:1911.00686.
45. Ismail, A.; Elpeltagy, M.S.; Zaki, M.; Eldahshan, K. A New Deep Learning-Based Methodology for Video Deepfake Detection Using XGBoost. *Sensors* **2021**, *21*, 5413. [CrossRef]

46. Rupapara, V.; Rustam, F.; Amaar, A.; Washington, P.B.; Lee, E.; Ashraf, I. Deepfake tweets classification using stacked Bi-LSTM and words embedding. *PeerJ Comput. Sci.* **2021**, *7*, e745. [CrossRef] [PubMed]
47. Chollet, F. Xception: Deep learning with depthwise separable convolutions. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 1251–1258.
48. 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.
49. Nguyen, H.H.; Yamagishi, J.; Echizen, I. Capsule-forensics: Using capsule networks to detect forged images and videos. In Proceedings of the ICASSP 2019—2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Brighton, UK, 12–17 May 2019; pp. 2307–2311.
50. Li, Y.; Lyu, S. Exposing deepfake videos by detecting face warping artifacts. *arXiv* **2018**, arXiv:1811.00656.
51. 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 2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS), Tampa, FL, USA, 23–26 September 2019; pp. 1–8.
52. 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.
53. Qi, H.; Guo, Q.; Juefei-Xu, F.; Xie, X.; Ma, L.; Feng, W.; Liu, Y.; Zhao, J. Deeprhythm: Exposing deepfakes with attentional visual heartbeat rhythms. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 4318–4327.
54. 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.
55. Hernandez-Ortega, J.; Tolosana, R.; Fierrez, J.; Morales, A. Deepfakeson-phys: Deepfakes detection based on heart rate estimation. *arXiv* **2020**, arXiv:2010.00400.
56. Khodabakhsh, A.; Busch, C. A generalizable deepfake detector based on neural conditional distribution modelling. In Proceedings of the 2020 International Conference of the Biometrics Special Interest Group (BIOSIG), Online, 16–18 September 2020; pp. 1–5.
57. Coccomini, D.A.; Messina, N.; Gennaro, C.; Falchi, F. Combining efficientnet and vision transformers for video deepfake detection. In *Proceedings of the International Conference on Image Analysis and Processing*; Springer: Berlin/Heidelberg, Germany, 2022; pp. 219–229.
58. Ganiyusufoglu, I.; Ngô, L.M.; Savov, N.; Karaoglu, S.; Gevers, T. Spatio-temporal features for generalized detection of deepfake videos. *arXiv* **2020**, arXiv:2010.11844.
59. Zhu, X.; Wang, H.; Fei, H.; Lei, Z.; Li, S.Z. Face forgery detection by 3d decomposition. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Virtual, 19–25 June 2021; pp. 2929–2939.
60. Jafar, M.T.; Ababneh, M.; Al-Zoube, M.; Elhassan, A. Forensics and analysis of deepfake videos. In Proceedings of the 2020 11th International Conference on Information and Communication Systems (ICICS), Irbid, Jordan, 7–9 April 2020; pp. 053–058.
61. Zhang, K.; Zhang, Z.; Li, Z.; Qiao, Y. Joint face detection and alignment using multitask cascaded convolutional networks. *IEEE Signal Process. Lett.* **2016**, *23*, 1499–1503. [CrossRef]
62. Sohrawardi, S.J.; Chintha, A.; Thai, B.; Seng, S.; Hickerson, A.; Ptucha, R.; Wright, M. Poster: Towards robust open-world detection of deepfakes. In Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security, London, UK, 11–15 November 2019; pp. 2613–2615.
63. Wang, J.; Wu, Z.; Ouyang, W.; Han, X.; Chen, J.; Jiang, Y.G.; Li, S.N. M2tr: Multi-modal multi-scale transformers for deepfake detection. In Proceedings of the 2022 International Conference on Multimedia Retrieval, Newark, NJ, USA, 27–30 June 2022; pp. 615–623.
64. Heo, Y.J.; Choi, Y.J.; Lee, Y.W.; Kim, B.G. Deepfake detection scheme based on vision transformer and distillation. *arXiv* **2021**, arXiv:2104.01353.
65. Wodajo, D.; Atnafu, S. Deepfake video detection using convolutional vision transformer. *arXiv* **2021**, arXiv:2102.11126.
66. Stroebel, L.; Llewellyn, M.; Hartley, T.; Ip, T.S.; Ahmed, M. A systematic literature review on the effectiveness of deepfake detection techniques. *J. Cyber Secur. Technol.* **2023**, *7*, 83–113. [CrossRef]
67. Rana, M.S.; Nobi, M.N.; Murali, B.; Sung, A.H. Deepfake detection: A systematic literature review. *IEEE Access* **2022**, *10*, 25494–25513. [CrossRef]
68. 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]
69. Deshmukh, A.; Wankhade, S.B. Deepfake Detection Approaches Using Deep Learning: A Systematic Review. In *Intelligent Computing and Networking: Proceedings of IC-ICN 2020*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 293–302.
70. Shahzad, H.F.; Rustam, F.; Flores, E.S.; Luís Vidal Mazón, J.; de la Torre Diez, I.; Ashraf, I. A Review of Image Processing Techniques for Deepfakes. *Sensors* **2022**, *22*, 4556. [CrossRef]
71. Mahmud, B.U.; Sharmin, A. Deep insights of deepfake technology: A review. *arXiv* **2021**, arXiv:2105.00192.

72. Kan, M. This AI Can Recreate Podcast Host Joe Rogan's Voice to Say Anything, 2019. Available online: https://www.pcmag.com/news/this-ai-can-recreate-podcast-host-joe-rogans-voice-to-say-anything#:~:text=A%20group%20of%20engineers%20has,to%20almost%20every%20word%20said (accessed on 2 May 2023)
73. Solsman, J.E. Samsung Deepfake AI Could Fabricate a Video of You from a Single Profile Pic, 2019. Available online: https://www.cnet.com/tech/computing/samsung-ai-deepfake-can-fabricate-a-video-of-you-from-a-single-photo-mona-lisa-cheapfake-dumbfake/ (accessed on 2 May 2023)
74. Evans, C. Spotting Fake News in a World with Manipulated Video, 2018. Available online: https://www.cbsnews.com/news/spotting-fake-news-in-a-world-with-manipulated-video (accessed on 2 May 2023)
75. Baron, K. Digital Doubles: The Deepfake Tech Nourishing New Wave Retail, 2019. Available online: https://www.forbes.com/sites/katiebaron/2019/07/29/digital-doubles-the-deepfake-tech-nourishing-new-wave-retail/?sh=5428ce694cc7 (accessed on 2 May 2023)
76. Brandon, J. Terrifying High-Tech Porn: Creepy 'Deepfake' Videos Are on the Rise, 2018. Available online: https://www.foxnews.com/tech/terrifying-high-tech-porn-creepy-deepfake-videos-are-on-the-rise (accessed on 2 May 2023)
77. Dickson, B. When AI Blurs the Line between Reality and Fiction. 2018. Available online: https://www.pcmag.com/news/when-ai-blurs-the-line-between-reality-and-fiction (accessed on 6 November 2022).
78. Chivers, T. What Do We Do about Deepfake Video? 2019. Available online: https://www.theguardian.com/technology/2019/jun/23/what-do-we-do-about-deepfake-video-ai-facebook (accessed on 6 November 2022).
79. Singh, D. WGoogle, Facebook, Twitter Put on Notice about Deepfakes in 2020 Election. 2019. Available online: https://www.cnet.com/tech/mobile/google-facebook-and-twitter-sent-letters-about-deepfakes-by-rep-schiff/ (accessed on 6 November 2022).
80. Dietmar, J. GANs and Deepfakes Could Revolutionize the Fashion Industry. 2019. Available online: https://www.forbes.com/sites/forbestechcouncil/2019/05/21/gans-and-deepfakes-could-revolutionize-the-fashion-industry/?sh=6f22c1723d17 (accessed on 6 November 2022).
81. Bell, K. The Most Urgent Threat of Deepfakes Isn't Politics. 2020. Available online: https://www.youtube.com/watch?v=hHHCrf2-x6w&t=2s (accessed on 6 November 2022).
82. Karasavva, V.; Noorbhai, A. The real threat of deepfake pornography: A review of canadian policy. *Cyberpsychology Behav. Soc. Netw.* **2021**, *24*, 203–209. [CrossRef] [PubMed]
83. Kerner, C.; Risse, M. Beyond porn and discreditation: Epistemic promises and perils of deepfake technology in digital lifeworlds. *Moral Philos. Politics* **2021**, *8*, 81–108. [CrossRef]
84. Fido, D.; Rao, J.; Harper, C.A. Celebrity status, sex, and variation in psychopathy predicts judgements of and proclivity to generate and distribute deepfake pornography. *Comput. Hum. Behav.* **2022**, *129*, 107141. [CrossRef]
85. Diakopoulos, N.; Johnson, D. Anticipating and addressing the ethical implications of deepfakes in the context of elections. *New Media Soc.* **2021**, *23*, 2072–2098. [CrossRef]
86. Hoven, J.v.d. Responsible innovation: A new look at technology and ethics. In *Responsible Innovation 1*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 3–13.
87. Siegel, D.; Kraetzer, C.; Seidlitz, S.; Dittmann, J. Media forensics considerations on deepfake detection with hand-crafted features. *J. Imaging* **2021**, *7*, 108. [CrossRef]
88. Wang, G.; Jiang, Q.; Jin, X.; Cui, X. FFR_FD: Effective and Fast Detection of Deepfakes Based on Feature Point Defects. *arXiv* **2021**, arXiv:2107.02016.
89. 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.
90. Korshunov, P.; Marcel, S. Deepfakes: A new threat to face recognition? assessment and detection. *arXiv* **2018**, arXiv:1812.08685.
91. Chen, H.S.; Rouhsedaghat, M.; Ghani, H.; Hu, S.; You, S.; Kuo, C.C.J. Defakehop: A light-weight high-performance deepfake detector. In Proceedings of the 2021 IEEE International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021; pp. 1–6.
92. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–20 June 2016; pp. 770–778.
93. Szegedy, C.; Ioffe, S.; Vanhoucke, V.; Alemi, A.A. Inception-v4, inception-resnet and the impact of residual connections on learning. In Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence, San Francisco, CA, USA, 4–9 February 2017.
94. Masi, I.; Killekar, A.; Mascarenhas, R.M.; Gurudatt, S.P.; AbdAlmageed, W. Two-branch recurrent network for isolating deepfakes in videos. In *Proceedings of the European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2020; pp. 667–684.
95. Fernandes, S.; Raj, S.; Ortiz, E.; Vintila, I.; Salter, M.; Urosevic, G.; Jha, S. Predicting heart rate variations of deepfake videos using neural ode. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Republic of Korea, 27–28 October 2019.
96. Tariq, S.; Lee, S.; Woo, S.S. A convolutional LSTM based residual network for deepfake video detection. *arXiv* **2020**, arXiv:2009.07480.
97. Güera, D.; Delp, E.J. Deepfake video detection using recurrent neural networks. In Proceedings of the 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), Auckland, New Zealand, 17–30 November 2018; pp. 1–6.

98. Chintha, A.; Thai, B.; Sohrawardi, S.J.; Bhatt, K.; Hickerson, A.; Wright, M.; Ptucha, R. Recurrent convolutional structures for audio spoof and video deepfake detection. *IEEE J. Sel. Top. Signal Process.* **2020**, *14*, 1024–1037. [CrossRef]
99. Montserrat, D.M.; Hao, H.; Yarlagadda, S.K.; Baireddy, S.; Shao, R.; Horváth, J.; Bartusiak, E.; Yang, J.; Guera, D.; Zhu, F.; et al. Deepfakes detection with automatic face weighting. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 668–669.
100. Khormali, A.; Yuan, J.S. DFDT: An End-to-End Deepfake Detection Framework Using Vision Transformer. *Appl. Sci.* **2022**, *12*, 2953. [CrossRef]
101. Khan, S.A.; Dai, H. Video transformer for deepfake detection with incremental learning. In Proceedings of the 29th ACM International Conference on Multimedia, Virtual Event, 20–24 October 2021; pp. 1821–1828.
102. Li, M.; Zuo, W.; Zhang, D. Deep identity-aware transfer of facial attributes. *arXiv* **2016**, arXiv:1610.05586.
103. Wang, X.; Huang, J.; Ma, S.; Nepal, S.; Xu, C. Deepfake Disrupter: The Detector of Deepfake Is My Friend. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 14920–14929.
104. Guarnera, L.; Giudice, O.; Guarnera, F.; Ortis, A.; Puglisi, G.; Paratore, A.; Bui, L.M.; Fontani, M.; Coccomini, D.A.; Caldelli, R.; et al. The Face Deepfake Detection Challenge. *J. Imaging* **2022**, *8*, 263. [CrossRef] [PubMed]
105. 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.
106. Wang, S.Y.; Wang, O.; Zhang, R.; Owens, A.; Efros, A.A. CNN-generated images are surprisingly easy to spot... for now. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 8695–8704.
107. Yang, C.; Lim, S.N. One-shot domain adaptation for face generation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 5921–5930.
108. Yang, C.; Lim, S.N. Unconstrained facial expression transfer using style-based generator. *arXiv* **2019**, arXiv:1912.06253.
109. Songsri-in, K.; Zafeiriou, S. Complement face forensic detection and localization with faciallandmarks. *arXiv* **2019**, arXiv:1910.05455.
110. Dolhansky, B.; Bitton, J.; Pflaum, B.; Lu, J.; Howes, R.; Wang, M.; Ferrer, C.C. The deepfake detection challenge (dfdc) dataset. *arXiv* **2020**, arXiv:2006.07397.
111. Guarnera, L.; Giudice, O.; Battiato, S. Fighting deepfake by exposing the convolutional traces on images. *IEEE Access* **2020**, *8*, 165085–165098. [CrossRef]
112. Frank, J.; Eisenhofer, T.; Schönherr, L.; Fischer, A.; Kolossa, D.; Holz, T. Leveraging frequency analysis for deep fake image recognition. In Proceedings of the International Conference on Machine Learning, Virtual, 13–18 July 2020; pp. 3247–3258.
113. Wolter, M.; Blanke, F.; Hoyt, C.T.; Garcke, J. Wavelet-packet powered deepfake image detection. *arXiv* **2021**, arXiv:2106.09369.
114. Fernandes, S.; Raj, S.; Ewetz, R.; Pannu, J.S.; Jha, S.K.; Ortiz, E.; Vintila, I.; Salter, M. Detecting deepfake videos using attribution-based confidence metric. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 308–309.
115. Huang, Y.; Juefei-Xu, F.; Wang, R.; Guo, Q.; Ma, L.; Xie, X.; Li, J.; Miao, W.; Liu, Y.; Pu, G. Fakepolisher: Making deepfakes more detection-evasive by shallow reconstruction. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 1217–1226.
116. Pu, J.; Mangaokar, N.; Wang, B.; Reddy, C.K.; Viswanath, B. Noisescope: Detecting deepfake images in a blind setting. In Proceedings of the Annual Computer Security Applications Conference, Austin, TX, USA, 6–10 December 2020; pp. 913–927.
117. Mitra, A.; Mohanty, S.P.; Corcoran, P.; Kougianos, E. EasyDeep: An IoT Friendly Robust Detection Method for GAN Generated Deepfake Images in Social Media. In *Proceedings of the IFIP International Internet of Things Conference*; Springer: Berlin/Heidelberg, Germany, 2021; pp. 217–236.
118. Zendran, M.; Rusiecki, A. Swapping Face Images with Generative Neural Networks for Deepfake Technology–Experimental Study. *Procedia Comput. Sci.* **2021**, *192*, 834–843. [CrossRef]
119. Narayan, K.; Agarwal, H.; Thakral, K.; Mittal, S.; Vatsa, M.; Singh, R. DeePhy: On Deepfake Phylogeny. *arXiv* **2022**, arXiv:2209.09111.
120. 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.
121. Lample, G.; Zeghidour, N.; Usunier, N.; Bordes, A.; Denoyer, L.; Ranzato, M. Fader networks: Manipulating images by sliding attributes. *Adv. Neural Inf. Process. Syst.* **2017**, *30*.
122. Shaoanlu. Faceswap-GAN. 2018. Available online: https://github.com/shaoanlu/faceswap-GAN (accessed on 6 November 2022).
123. Neuralchen. SimSwap. 2021. Available online: https://github.com/neuralchen/SimSwap (accessed on 6 November 2022).
124. Usingcolor. Faceshifter. 2020. Available online: https://github.com/mindslab-ai/faceshifter (accessed on 6 November 2022).
125. YuDeng. DiscoFaceGan. 2020. Available online: https://github.com/microsoft/DiscoFaceGan (accessed on 6 November 2022).
126. Faceapp. 2019. Available online: https://www.faceapp.com/ (accessed on 6 November 2022).
127. Heidari, A.; Navimipour, N.J.; Jamali, M.A.J.; Akbarpour, S. A green, secure, and deep intelligent method for dynamic IoT-edge-cloud offloading scenarios. *Sustain. Comput. Inform. Syst.* **2023**, *38*, 100859. [CrossRef]

128. Heidari, A.; Javaheri, D.; Toumaj, S.; Navimipour, N.J.; Rezaei, M.; Unal, M. A new lung cancer detection method based on the chest CT images using Federated Learning and blockchain systems. *Artif. Intell. Med.* **2023**, *141*, 102572. [CrossRef]
129. Dependabot. Sensity AI, 2022. Available online: https://github.com/sensity-ai/dot (accessed on 3 April 2023).
130. Truepic. Truepic. 2022. Available online: https://truepic.com/ (accessed on 3 April 2023).
131. Ddi. DDI. 2022. Available online: https://www.d-id.com/ (accessed on 3 April 2023).
132. Mgongwer. DeepTraCE. 2022. Available online: https://github.com/DeNardoLab/DeepTraCE (accessed on 3 April 2023).
133. DSA. Deep Secure AI. 2023. Available online: https://tracxn.com/d/companies/deep-secure-ai/__Vg5KA9H7Is7wzbVWluIoNcwc_XaTgx1t3WSjzigbEE4 (accessed on 3 April 2023).
134. Iproov. Iproov. 2022. Available online: https://www.iproov.com/blog/deepfakes-statistics-solutions-biometric-protection (accessed on 3 April 2023).
135. Blackbird. Blackbird. 2023. Available online: https://www.blackbird.ai/blog/2023/04/navigating-the-warped-realities-of-generative-ai (accessed on 3 April 2023).
136. Sentinel. Sentinel. 2021. Available online: https://thesentinel.ai/ (accessed on 3 April 2023).
137. Amber. Amber. 2020. Available online: https://www.wired.com/story/amber-authenticate-video-validation-blockchain-tampering-deepfakes/ (accessed on 3 April 2023).
138. Amberapp. Amberapp. 2020. Available online: https://app.ambervideo.co/public (accessed on 3 April 2023).
139. FaceForensics. FaceForensics. 2020. Available online: https://github.com/ondyari/FaceForensics (accessed on 3 April 2023).
140. Fakespot. Fakespot. 2023. Available online: https://www.fakespot.com/ (accessed on 3 April 2023).
141. Datitran. Face2face. 2018. Available online: https://github.com/datitran/face2face-demo (accessed on 6 November 2022).
142. Torzdf. Faceswap. 2018. Available online: https://github.com/deepfakes/faceswap (accessed on 6 November 2022).
143. YuvalNirkin. FaceSwap. 2017. Available online: https://github.com/YuvalNirkin/face_swap (accessed on 6 November 2022).
144. Nirkin, Y.; Masi, I.; Tuan, A.T.; Hassner, T.; Medioni, G. On face segmentation, face swapping, and face perception. In Proceedings of the 2018 13th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2018), Xi'an, China, 15–19 May 2018; pp. 98–105.
145. Burgos-Artizzu, X.P.; Perona, P.; Dollár, P. Robust face landmark estimation under occlusion. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 1513–1520.
146. Tan, M.; Le, Q. Efficientnet: Rethinking model scaling for convolutional neural networks. In Proceedings of the International Conference on Machine Learning, Long Beach, CA, USA, 9–15 June 2019; pp. 6105–6114.
147. Mitchell, T.; Buchanan, B.; DeJong, G.; Dietterich, T.; Rosenbloom, P.; Waibel, A. Machine learning. *Annu. Rev. Comput. Sci.* **1990**, *4*, 417–433. [CrossRef]
148. Timesler. Facenet-Pytorch. 2018. Available online: https://github.com/timesler/facenet-pytorch (accessed on 6 November 2022).
149. Xiang, J.; Zhu, G. Joint face detection and facial expression recognition with MTCNN. In Proceedings of the 2017 4th International Conference on Information Science and Control Engineering (ICISCE), Changsha, China, 21–23 July 2017; pp. 424–427.
150. matkob. OpenPose. 2018. Available online: https://github.com/CMU-Perceptual-Computing-Lab/openpose (accessed on 6 November 2022).
151. Cao, Z.; Simon, T.; Wei, S.E.; Sheikh, Y. Realtime multi-person 2d pose estimation using part affinity fields. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 7291–7299.
152. Lin, T.Y.; Maire, M.; Belongie, S.; Hays, J.; Perona, P.; Ramanan, D.; Dollár, P.; Zitnick, C.L. Microsoft coco: Common objects in context. In *Proceedings of the European Conference on Computer Vision*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 740–755.
153. Andriluka, M.; Pishchulin, L.; Gehler, P.; Schiele, B. 2d human pose estimation: New benchmark and state of the art analysis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbus, OH, USA, 23–28 June 2014; pp. 3686–3693.
154. BiDAlab. DeepfakesON-Phys. 2020. Available online: https://github.com/BiDAlab/DeepFakesON-Phys (accessed on 6 November 2022).
155. Coccomini, D. Combining EfficientNet and Vision Transformers for Video Deepfake Detection. 2021. Available online: https://github.com/davide-coccomini/Combining-EfficientNet-and-Vision-Transformers-for-Video-Deepfake-Detection (accessed on 6 November 2022).
156. Iperov. DeepFaceLab. 2018. Available online: https://github.com/iperov/DeepFaceLab (accessed on 6 November 2022).
157. Younger, P. Deeptomcruise. Available online: https://www.tiktok.com/@deeptomcruise (accessed on 6 November 2022).
158. Schwarz, L. Arnoldschwarzneggar. Available online: https://www.tiktok.com/@arnoldschwarzneggar (accessed on 6 November 2022).
159. Diepnep. Available online: https://www.tiktok.com/@diepnep (accessed on 6 November 2022).
160. Deepcaprio. 2018. Available online: https://www.tiktok.com/@deepcaprio (accessed on 6 November 2022).
161. vfx. VFXChrisUme. 2019. Available online: https://www.youtube.com/c/VFXChrisUme (accessed on 6 November 2022).
162. Shamook. Shamook. 2018. Available online: https://www.youtube.com/channel/UCZXbWcv7fSZFTAZV4beckyw/videos (accessed on 6 November 2022).
163. NextFace. 2018. Available online: https://www.youtube.com/c/GuusDeKroon (accessed on 6 November 2022).

164. Deepfaker. 2018. Available online: https://www.youtube.com/channel/UCkHecfDTcSazNZSKPEhtPVQ (accessed on 6 November 2022).
165. Deepfakes in Movie. 2019. Available online: https://www.youtube.com/c/DeepFakesinmovie (accessed on 6 November 2022).
166. DeepfakeCreator. 2020. Available online: https://www.youtube.com/c/DeepfakeCreator (accessed on 6 November 2022).
167. Jarkancio. Jarkan. 2007. Available online: https://www.youtube.com/c/Jarkan (accessed on 6 November 2022).
168. Zhang, S.; Zhu, X.; Lei, Z.; Shi, H.; Wang, X.; Li, S.Z. S3fd: Single shot scale-invariant face detector. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 192–201.
169. Bulat, A.; Tzimiropoulos, G. How far are we from solving the 2d & 3d face alignment problem? (and a dataset of 230,000 3d facial landmarks). In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 1021–1030.
170. Feng, Y.; Wu, F.; Shao, X.; Wang, Y.; Zhou, X. Joint 3d face reconstruction and dense alignment with position map regression network. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 534–551.
171. Iglovikov, V.; Shvets, A. Ternausnet: U-net with vgg11 encoder pre-trained on imagenet for image segmentation. *arXiv* **2018**, arXiv:1801.05746.
172. Fakeapp. Available online: https://www.fakeapp.com/ (accessed on 6 November 2022).
173. Deepfakesweb. Available online: https://deepfakesweb.com/ (accessed on 6 November 2022).
174. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive growing of gans for improved quality, stability, and variation. *arXiv* **2017**, arXiv:1710.10196.
175. 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.
176. Carlini, N.; Farid, H. Evading deepfake-image detectors with white-and black-box attacks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 658–659.
177. Li, L.; Bao, J.; Zhang, T.; Yang, H.; Chen, D.; Wen, F.; Guo, B. Face X-ray for more general face forgery detection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 5001–5010.
178. 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.