Chapter 21
Future Trends in Digital Face Manipulation and Detection

Ruben Tolosana, Christian Rathgeb, Ruben Vera-Rodriguez, Christoph Busch, Luisa Verdoliva, Siwei Lyu, Huy H. Nguyen, Junichi Yamagishi, Isao Echizen, Peter Rot, Klemen Grm, Vitomir Štruc, Antitza Dantcheva, Zahid Akhtar, Sergio Romero-Tapiador, Julian Fierrez, Aythami Morales, Javier Ortega-Garcia, Els Kindt, Catherine Jasserand, Tarmo Kalvet, and Marek Tiits

Abstract Recently, digital face manipulation and its detection have sparked large interest in industry and academia around the world. Numerous approaches have been proposed in the literature to create realistic face manipulations, such as DeepFakes and face morphs. To the human eye manipulated images and videos can be almost indistinguishable from real content. Although impressive progress has been reported

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

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

S. Romero-Tapiador
e-mail: sergio.romerot@uam.es

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

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

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

C. Rathgeb · C. Busch
Hochschule Darmstadt, Darmstadt, Germany
e-mail: christian.rathgeb@h-da.de

C. Busch
e-mail: christoph.busch@h-da.de

L. Verdoliva
University of Naples Federico II, Naples, Italy
e-mail: verdoliv@unina.it

C. Rathgeb et al. (eds.), *Handbook of Digital Face Manipulation and Detection*,
Advances in Computer Vision and Pattern Recognition,
https://doi.org/10.1007/978-3-030-87664-7_21

in the automatic detection of such face manipulations, this research field is often considered to be a *cat and mouse game*. This chapter briefly discusses the state of the art of digital face manipulation and detection. Issues and challenges that need to be tackled by the research community are summarized, along with future trends in the field.

21.1 Introduction

Over the last couple of years, digital face manipulation and detection has become a highly active area of research. This is demonstrated through the increasing number of workshops in top conferences [1–5], international projects such as MediFor and the recent SemaFor funded by the Defense Advanced Research Project Agency

S. Lyu
University at Buffalo, Buffalo, USA
e-mail: siweilyu@buffalo.edu

H. H. Nguyen
The Graduate University for Advanced Studies, Hayama, Japan
e-mail: nhhuy@nii.ac.jp

J. Yamagishi · I. Echizen
National Institute of Informatics, Chiyoda City, Japan
e-mail: jyamagis@nii.ac.jp

I. Echizen
e-mail: iechizen@nii.ac.jp

P. Rot · K. Grm · V. Štruc
University of Ljubljana, Ljubljana, Slovenia
e-mail: peter.rot@fe.uni-lj.si

K. Grm
e-mail: klemen.grm@fe.uni-lj.si

V. Štruc
e-mail: vitomir.struc@fe.uni-lj.si

A. Dantcheva
Inria Sophia Antipolis, Biot, France
e-mail: antitza.dantcheva@inria.fr

(DARPA), and competitions such as the Media Forensics Challenge (MFC2018)[1] launched by the National Institute of Standards and Technology (NIST), the Deepfake Detection Challenge (DFDC)[2] launched by Facebook, and the recent DeeperForensics Challenge.[3]

Face manipulation techniques can erode trust in digital media through fake news and the spread of misinformation [6]. With the big impact of social networks on our daily life, disinformation can be easily widespread and influence the public opinion [7]. Its targets can be individuals, economy, or politics [8]. Manipulated videos have already been used to create political tensions, and the technology enabling their creation is being considered as a threat by various governments [9].

Motivated by those facts, researchers have proposed various techniques to detect digital face manipulations in the recent past [10, 11]. In addition, public databases have been made available and first benchmarks have been conducted by different research groups [12–17], proving the high potential of the latest manipulation detectors. Nonetheless, a reliable detection of manipulated face images and videos is still considered an unsolved problem. It is generally conceded that digital face manipulation detection is still a nascent field of research in which numerous issues and challenges have to be addressed in order to reliably deploy such methods in real-world applications.

This chapter concludes the book providing an overview of open issues and challenges in the field of digital face manipulation and detection. Limitations of state-of-the-art methods are pointed out and potential future research direction toward advancing both fields are summarized, including promising application areas as well as novel use-cases. Moreover, legal and societal aspects of digital face manipulation and detection are discussed, such as the legality and legitimacy of the use of the manipulation detection or the potentially conflicting right to "one's own image", among others. Listing currently unsolved problems in the field, this chapter is intended to serve as a starting point for new researchers in the field.

Z. Akhtar
State University of New York Polytechnic Institute, Utica, USA
e-mail: akhtarz@sunypoly.edu

E. Kindt
Universiteit Leiden, Leiden, Netherlands
e-mail: els.kindt@kuleuven.be

E. Kindt · C. Jasserand
KU Leuven, Leuven, Belgium
e-mail: catherine.jasserand@kuleuven.be

T. Kalvet · M. Tiits
Institute of Baltic Studies and TalTech, Tallinn, Estonia
e-mail: tarmo@ibs.ee

M. Tiits
e-mail: marek@ibs.ee

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

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

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

The remainder of this chapter is organized as follows: Sect. 21.2 briefly describes the current state of the art in face manipulation together with public available databases. The most relevant issues with respect to the detection of face manipulations are discussed in Sect. 21.3. In Sect. 21.4, future research directions and application areas are summarized. Subsequently, Sect. 21.5 discusses societal and legal aspects of face manipulation and detection. Finally, a summary is given in Sect. 21.6.

21.2 Realism of Face Manipulation and Databases

21.2.1 State of the Art

Face manipulation techniques have been improved significantly in the last years as discussed in Part II of the book. State-of-the-art techniques are able to generate fake images and videos that are indistinguishable to the human eye [10, 11, 18]. However, when considering automatic end-to-end manipulation techniques, the visual quality is not always stable and it depends severely on different aspects [17, 19–21]. Figure 21.1 shows some weaknesses that limit the naturalness and facilitate fake detection. We highlight next some of the most critical aspects:

- **Face detection and segmentation**, which is not 100% accurate [22, 23]: this problem becomes worse when input images or videos are in bad quality, e.g., bad lighting condition, noisy, blurry, or low resolution.
- **Blending manipulated faces into the original image or video**: although there have been improvements in the blending algorithms [19], artifacts at the edges of the manipulated and original regions still exist in many cases. In addition, mismatch between these two regions (e.g., lighting condition, skin color, or noise) can degrade the realism of the manipulated images/videos, making them easier to be detected.
- **Low-quality synthesized faces**: while progress has been made here thanks to Generative Adversarial Networks (GAN), for example, through the recent StyleGAN2 model [24] that is able to generate non-existent faces with high resolution, editing with such models through GAN inversion techniques is time consuming and computationally demanding. This computational complexity also hinders development of high-resolution video manipulation techniques. Basic techniques often generate low-resolution face images, typically between 64×64 and 256×256 pixels as discussed in Chap. 4 of the book.
- **Temporal inconsistencies along frames**: this is of special importance in face manipulations such as Audio-to-Video as discussed in Chap. 8 of the book: Are there any relationships between audio and other facial features such as eyes, teeth, and even head movements? Techniques based on 3D pose and expression could further benefit this research line [25].

Apart from the aspects commented before, it is also interesting to highlight that most face manipulation techniques are currently focused only on the visual quality at pixel level [10, 11], to the best of our knowledge. Biological aspects of the human

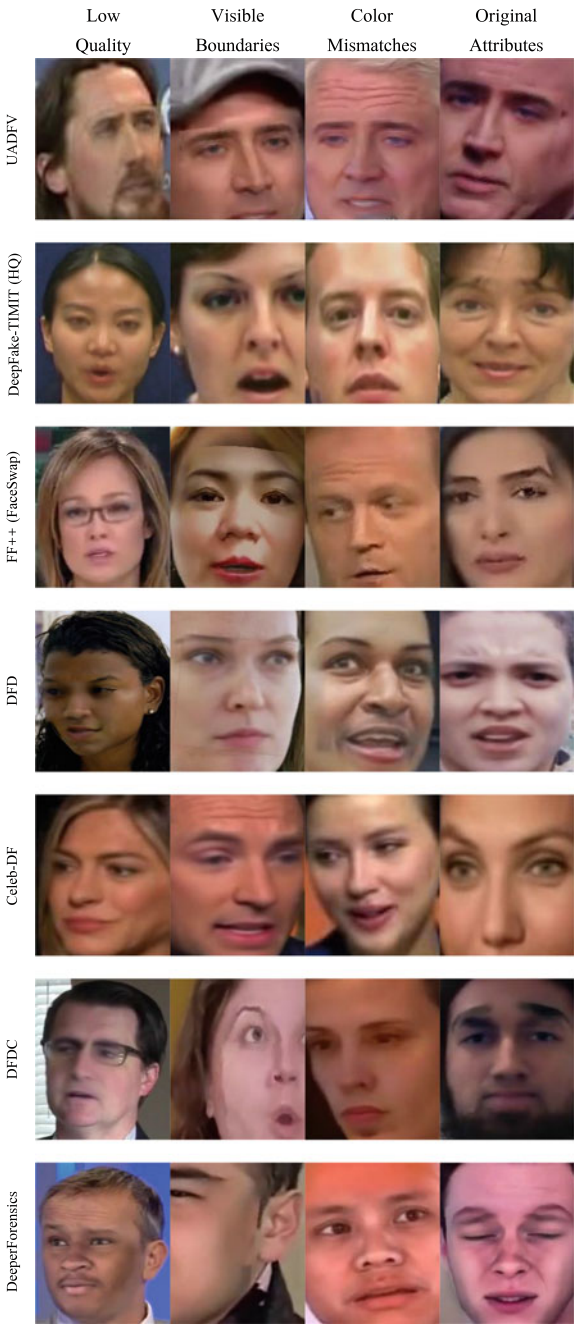

Fig. 21.1 Weaknesses of automatic end-to-end face manipulations that limit the naturalness and facilitate fake detection

being should be also taken into account in the manipulation process, e.g., blood circulation (heart rate) and eye blink rate could be automatically extracted from the faces to detect whether the video is real or fake [26, 27], as discussed in Chap. 12.

21.2.2 Missing Resources

Although new public databases are being released recently, these generally lack diversity and include low-quality videos. Specifically, for image databases, GAN-based models have been proved to be very effective in creating realistic faces, but these are usually generated with one model (e.g., StyleGAN2 [24]) with one specific set of parameters [28–30]. Video databases, on the other hand, are plagued with low-quality synthesis results, exhibiting visible blurring or boundary artifacts. As a result, it is easy for current fake detectors to over-adapt to the specific characteristics of the generation method and artifacts. In addition, high-accuracy fake detection performances on databases containing significant fraction of low-quality videos will not be representative for the performance in real life.[4] To make the detection more challenging, databases need to improve on the types and variants of the generation models, post-processing steps, codecs, and compression methods, as well as adversarial attacks.

Furthermore, current databases are still small and monotonic compared with those in other areas like image classification or speech, e.g., ImageNet [31] or VoxCeleb2 [32]. One of the largest databases, the DFDC dataset [12], only has 128,154 videos with less than 20 types of manipulation methods. Moreover, most databases only contain one or two subjects in an image or video except the recently released Face Forensics in the Wild (FFIW) database [33], and they are easily perceived (not a small subject in a crowd). It is also interesting to highlight that none of the databases contain manipulated or synthesized speech except the DFDC, but its manipulation is simple and is hard to be considered as "fake".

Finally, it is important to highlight two more missing resources in face manipulation: (i) the generation of manipulated face databases based on 3D manipulation techniques [34], and (ii) the generation of multimodal manipulated face databases as current ones are only focused on the manipulation of either audio or face visual information [35, 36].

21.3 Limitations of Face Manipulation Detection

21.3.1 Generalizability

The vast majority of existing face manipulation detection techniques have been evaluated only on known types of face manipulations or on a single database [10, 11]. In other words, the published empirical results showed performances of detectors under same train and test manipulation type/database. However, their performances

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

usually drop significantly when evaluated under cross-manipulation setting (where train and test sets are not from the same manipulation type or database) [12, 19, 37, 38]. Therefore, reported detection performance rates are over-optimistic.

Tackling the unknown emerging face manipulations is still a key challenge [30, 39]. In fact, generalization of detection techniques is crucial in attaining dependable accuracy in real-life scenarios. It is agreed upon researchers that face manipulation and detection is well described as a *cat and mouse game*, where improvements in one area trigger improvements in the other.

The generalization capabilities of existing detectors are still an open issue that is difficult to address with today's (mostly supervised) solutions. An evident example of this *generalization problem* was demonstrated by the recent Deepfake Game Competition (DFGC) [40], held in conjunction with the 2021 edition of the International Joint Conference on Biometrics (IJCB 2021[5]). The competition had multiple rounds of submissions, where participants first designed DeepFake detectors based on the training data provided by the organizers and then contributed novel DeepFake generation techniques to test the detectors. With most developed detection techniques, the performance deteriorated quickly with the introduction of novel DeepFakes not seen during training.

Beyond being able to generalize, it is important that current methods are robust to possible post-processing steps. In fact, media assets often undergo a number of not malicious operations, such as compression and resizing [41], that occurs every time they are uploaded over a social network or made available on a website. Now these operations tend to weaken the forensic traces and above all cause a misalignment between training and test data that can make the learning-based detectors not properly work [11]. Similar problems are also seen in other related areas, for example, Presentation Attack Detection (PAD) [42–45], which despite decades of research, issues with cross-dataset performance and robustness to unseen attacks is still a major issue of even the most advanced solutions.

21.3.2 Interpretability

Until now, very few studies have attempted to explore the interpretability and trustworthiness aspects of face manipulation detectors [46]. Many detection methods, particularly those based on deep neural networks, generally lack explainability owing to the black box nature of deep learning techniques. The fake detectors presently label a face sample with a fakeness probability score, occasionally detection confidence is provided, but little insight about such results is provided beyond simple numerical scores. It would be more beneficial to describe why a detector predicts a specific face as real or manipulated. For instance, which face parts are believed to be forged and where the detector is looking for label prediction [17]. For human, it is vital to comprehend and trust the opinion of a fake detector—however, the human expert operates

[5] http://ijcb2021.iapr-tc4.org/.

at the end of the processing chain and therefore wants to understand the decision. A numerical score or label not corroborated with decent reasoning and insights cannot be accepted in some critical applications like journalism or law enforcement, among others.

Furthermore, it is not easy to characterize the intent of a face manipulation. So far, learning-based detectors cannot distinguish between malicious and benign processing operations. For example, it is impossible to tell if a change of illumination was carried out only for the purpose of enhancement or to better fit a swapped face in a video. In general, characterizing the intent of a manipulation is extremely difficult and it will become even harder with the spread of deep learning-based methods in almost all the enhancement operations. In fact, how can a face manipulation detector realize that GAN content generated for super-resolution is acceptable while GAN content that modify a face attribution is not? In a world where most of the media are processed using deep learning-based tools, it is increasingly likely that something be manipulated, and the key to forensic performance is learning the intent behind a manipulation. A good forensic detector should be able to single out only malicious manipulations on faces, by leveraging not only on the single media but looking at the context and including all other related media and textual information.

21.3.3 Vulnerabilities

State-of-the-art detection methods make heavy use of deep learning, i.e., deep neural network models serve as the most popular backbone. Such approaches can suffer severely from the adversarial attacks as some recent works suggested [47–49]. Although real-world fake detectors may cope with various degradations like video/image noise, compression, etc., they can be vulnerable to adversarial examples with imperceptible additive noises [47, 50]. Prior studies have demonstrated that detectors based on neural networks are most susceptible to adversarial attacks [51]. Unfortunately, it has been noticed that all existing methods seem to fail against adversarial attacks, even the accuracy of some fake detectors is reduced to 0% [51].

Beyond adversarial attacks, it is worth observing that every detection algorithm should take into account the presence of an adversary to fool it. In fact, by relying on the knowledge of the specific clues exploited by a face manipulation detector, one can make it not work anymore. For example, if an adversary knows that the algorithm exploits the presence of the specific GAN fingerprints that characterize synthetic media, then it would be possible to remove them [30] and also to insert real fingerprints related to modern digital cameras [52]. Overall, researchers should be always aware about the two-player nature of this research and design a detector robust also to possible targeted attacks.

21.3.4 Human Capabilities

Detecting high-quality face manipulations by humans is already a highly challenging tasks, especially if the subject is not versed in this area. While researchers working on face manipulation are still often able to spot giveaways with the current generation of manipulation techniques, it is expected that this will change in the near future. Although humans have a limited capability to detect high-quality face manipulations, they are usually better at detecting manipulation patterns with little prior knowledge. Thus, they can still be included in the forensic applications' decision-making. Human in the loop systems will lead us to a better reliable detection of high-quality face manipulations. As the quality of digital face manipulation is improving so quickly, it might not be possible to detect them solely based on human visual inspection without an in-depth analysis of image characteristics. Two recent studies [53, 54] have shown that humans cannot reliably distinguish images generated by advanced GAN technologies from pristine images. The average accuracy turned out to be around 50% (coin tossing) for untrained observers, increasing to just 60% for trained observers with unlimited analysis time [54]. In ref. [53], experiments reveal that the realism of synthetic images even surpasses those of real images (68% for synthetic images versus 52% for real ones).

Fooling machines, on the other hand, is more challenging as long as examples of face manipulations are available for supervised training, which does not always simulate real-life scenarios.

21.3.5 Further Limitations

Standards in the field of face manipulation and detection represent the common rules for assembling, evaluating, storing, and sharing samples and detectors' performances. There are no international standards yet, although some inceptive efforts have been made toward this [21]. There is a strong need for standardized frameworks, which should be composed of protocols and tools for manipulation generation and detection, common criteria, and open platforms to transparently analyze systems against benchmarks. Such standardized frameworks will help operators and consumers of digital media to generate, evaluate, configure, or compare face manipulation and detection techniques.

21.4 Face Manipulation and Detection: The Path Forward

21.4.1 Application Areas for Face Manipulation

Face manipulation techniques could mark a milestone in many different application areas in the near future. We summarize next and in Fig. 21.2 some potential applications:

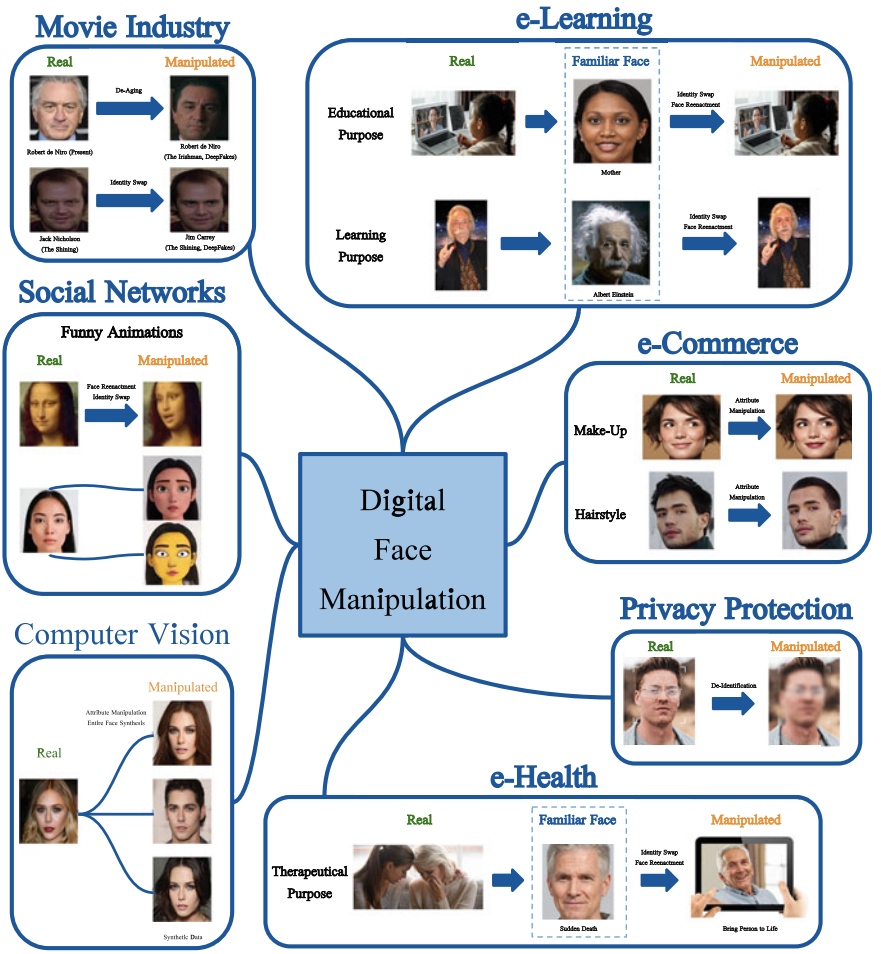

Fig. 21.2 Application areas of the face manipulation technology

- **Movie industry**: it is a simple and cheap way to do animations compared with traditional computer graphics techniques.[6] With some improvements in terms of quality and resolution, we can foresee that DeepFakes will revolutionize the movie industry, for example, allowing dead actors to act again and to speak seamlessly many languages, enhancing expressions, as well as allowing new settings and takes (e.g., 3D views anytime, without expensive equipment).

[6] https://www.youtube.com/watch?v=dHSTWepkp_M&t=76s.

- **Social networks and entertainment**: there already exist several startups focusing on building funny animations from still images using lip sync technology, e.g., Avatarify,[7] Wombo.ai,[8] DeepNostalgia.[9]
- **Privacy protection**: face manipulation techniques could conceal certain attributes of the subject from human observers or automatic techniques. For example, face de-identification techniques aim to make identity inference from images and video impossible by altering key facial properties [55, 56], and soft-biometric privacy-enhancing techniques [57] try to modify image characteristics to make it difficult to automatically infer potentially sensitive attribute information from facial images, e.g., age, gender, or ethnicity. These techniques could be very valuable, among others, to replace faces of subjects or witnesses who wish to conceal their identity in fear of prosecution and discrimination.
- **e-Commerce**: face attribute manipulations could further benefit the retail sector, for example, through popular applications such as FaceApp.[10] Consumers could use this technology to try on a broad range of products such as cosmetics and makeup, glasses, or hairstyles in a virtual and user-friendly environment.
- **e-Learning**: face manipulation techniques could enhance the process of remote education of children/students in many different scenarios, for example, swapping teacher's face with their parents as it is proved that familiarity enhances the rate of learning. Similarly, videos of historical figures could be generated, allowing students to learn about the topics in a more interactive way, generating more appealing learning scenarios.
- **e-Health**: bring a person to life using face manipulation techniques could be very valuable for therapeutic purposes, allowing patients to express their feelings and get over hard situations, e.g., sudden deaths.
- **Computer vision**: due to the nature of contemporary machine learning models (which are notoriously data hungry), larger and larger datasets are needed for training and ensuring competitive performance. A common, but questionably practice established by the computer vision community in recent years, is to address this demand for data collecting large-scale datasets from the web. However, no consent is obtained for such collections and the generated datasets are often associated with privacy concerns. Using synthetic data, generated from images of a small number of consenting subject and state-of-the-art manipulation techniques, may be a possibility to address the need for the enormous amount of data required by modern machine learning models and comply with existing data protection regulations, such as General Data Protection Regulation (GDPR) of the European Union.

[7] https://avatarify.ai/.

[8] https://www.wombo.ai/.

[9] https://www.myheritage.com/deep-nostalgia.

[10] https://www.faceapp.com/.

21.4.2 Promising Approaches

Fake detection technology will continue improving in the coming years. One evidence is that more and more publications have appeared in the last years in top conferences such as AAAI, CVPR, and ICCV. However, as the face manipulation technology will also improve simultaneously, we may still not see highly reliable detectors in a near future, especially those that can handle unseen face manipulations, which is currently one of the most challenging limitations of the detectors as discussed in Sect. 21.3.

Several research directions can be pursued to improve the generalization problem of current face manipulation detectors:

- We expect to see more interest in **one-class learning** models, commonly used in the anomaly and novelty detection literature [58, 59]. Such models learn from real examples and do not require examples of manipulated data to train fake detectors. As a result, they are expected to generalize better for the detection of novel (unseen) face manipulation techniques. Of course, such detection techniques come with their own set of problems that range from data representation and model design to learning objectives, among others.
- **Online learning** is also one promising way to deal with generalization [60]. Unfortunately, current databases are not optimal to conduct online learning research. Therefore, focusing on making better databases and applying online learning can be done together to improve face manipulation detection in the future.
- Recent studies suggest that no single feature/characteristic is adequate to build effective and robust detectors of face manipulations. On the other hand, many successful real-life machine learning solutions are based on **ensemble models** that fuse results from individual types of features or detectors and are calibrated for stronger collective performance [61, 62]. The most notable example is the recent fake detectors presented in the DeepFake Detection Challenge [12].
- Similar to the previous point, **multimodal approaches**, which are able to fuse multiple detection strategies including artifact analysis, identity-aware detection, as well as contextual information such as accompanying text, audio, and origin of data. Multimodal approaches also increase the interpretability and hence the understanding of the reasoning of deep neural networks [63].
- More recently, **identity-aware detection** mechanisms have been proposed which do not learn to detect specific artifacts but rather learn features of a subject [64]. However, such schemes additionally require reference data resulting in a differential detection approach [65, 66].

Apart from the promising fake detection approaches listed above, researchers working on the topic of face manipulation could incorporate mechanisms that intentionally include imperceivable elements (watermarks) into the manipulated images/videos in order to make the detection easier [67, 68]. While such idea does not address the general problem of detecting face manipulations, it could set the bar for

adversaries higher and make sharing face manipulation techniques (with legitimate use cases) with the research community less challenging.

Finally, face manipulation techniques could also improve privacy protection [69]. Research on privacy-enhancing techniques is increasingly looking at formal privacy schemes, such as k-Anonymity [70, 71] or ϵ-differential privacy [72, 73], which provide formal (mathematically proven) guarantees about the privacy levels ensured. We expect to see novel algorithms around these concepts in the near future.

21.5 Societal and Legal Aspects of Face Manipulation and Detection

Face manipulation brings an array of complex legal issues. There is no comprehensive legislation on the use of manipulated images, yet several aspects are already regulated in various countries. It should hence not surprise that the development of new manipulation technology and the detection thereof also leads to new issues and questions from a legal perspective which deserve further research.

If it is used to mislead, manipulated images can cause significant harm to the individuals they falsely portray. They can cause emotional distress and reputational damage. The victims of these fake images can try to find relief through torts and criminal laws. Beyond individuals, these digitally altered images can also affect society at large. The problem is that viewers are most of the time not aware that these images are not genuine. In some countries, altered (body) images used for commercial purposes (such as the fashion industry) need to be labeled. More generally, legislative proposals in several countries try to tackle the transparency issue by imposing an obligation to inform users that they interact with AI-generated content (such as DeepFakes). Besides this aspect, manipulated face images might also be subject to copyright protection. But only the photographer of the original images can benefit from it and object to their use without his or her authorization. On the other side, the subjects might benefit from image, publicity, and privacy rights for the alteration of their images without their consent. The rules are different from one country to another. In some jurisdictions, they will be balanced with individuals' freedom of speech (that could allow them to alter these images). But not every use of altered face images is intended to be malicious. Indeed, they can be very beneficial to some industries (such as entertainment or healthcare as discussed in Sect. 21.4.1). Therefore, it is very challenging to tackle the complexity of the use of digitally manipulated face images with a single piece of legislation while technically it would be possible to apply cryptographic techniques to ensure the integrity and authenticity of image data. Finally, there is room to investigate the rules applicable to the digital alteration of the face images for research purposes.

Focusing on face manipulation, one shall keep in mind that in several countries individuals have a right to "one's own image". This implies that individuals are entitled to control their representation and the reproduction of their images, especially

face, to the outside world. In an increasingly digitized world, these individuals may also choose to protect their digital images, for example, to prevent profiling. Detecting manipulations, based on this right to control your own image by protecting it, should not have adverse effects for these individuals, unless there is a clear legal rule that this would be forbidden for legitimate reasons, e.g., on identity documents. The potential conflict between this specific right to "one's own image" and other needs, e.g., of public authorities deserves further discussion and debate, based on researched arguments.

All new digital technology used in a societal context raises inevitably new questions, also from the legal side. The reasons why digital technologies are often under close review also by the regulator is that such technologies may change existing (power) relations and affect prior balances once established, for example, when investigating crime or when spreading news information.

Once the manipulation technologies are more widely used, for Example, for spreading fake news over digital platforms, the owners of such platform will face the need for a delicate exercise of assessing whether and removing any information was manipulated. This exercise risks to collide with some fundamental principles in democratic societies, such as the right of freedom of speech, but also the rights to respect for privacy and data protection. For instance, there are currently several proposals for the regulation of digital content on platforms, including an EU Commission's proposal of a Digital Services Act, setting a common set of rules on obligations for so-called intermediaries offering digital services. These services would include video messages, which could be manipulated as to the identities of the actors therein, leading to identity theft or spreading false information.

In case manipulation detection methods are used by public authorities competent for preventing, investigating, detecting, or prosecuting criminal offences this shall be done in a lawful and fair manner. While these are broad concepts, case law further explains how to apply these concepts. Lawfulness refers to the need—in accordance with the rule of law principle—to adopt adequate, accessible, and foreseeable laws with sufficient precision and sufficient safeguards whenever the use of the detection technology, which may be considered as a part of or sub process to, e.g., for face recognition, could interfere with fundamental rights and freedoms. When used for forensics, explainability of the algorithms used, also in court, will be high on the agenda. Fairness points to the need for being transparent about the use of the technology. Furthermore, it is obvious that the use of the detection methods should be restricted to well-defined legitimate purposes, such as, e.g., preventing illegitimate migration or detecting identity fraud. From an organizational point, one should also know that decisions purely and solely based on automated processing, producing adverse legal effects or significantly effecting subjects, are prohibited, unless authorized by law, and subject to appropriate safeguards, including at least human oversight and intervention. Again, according technical solutions to assure the authenticity of data would need to be implemented as a prerequisite.

As argued in Chap. 20, the risk of harming integrity of personal identity and misuse of it as well as the risk of privacy invasion and function creep represent major issues. In the case of integrity of subject's identity, during the identity theft or loss more than

privacy will be harmed, the subject could be refused access to services, lose control over their identity, and face damages which are done in their name.

Regarding privacy and function creep, the main issues are related to solutions where a subject's data are used without his or her authorisation. In these cases, how the data is leaked—whether it is from a data leak or insecure service, hackers (adversaries), or vulnerable data systems—is not as important as what the consequences were [74, 75].

The absence of a unified approach, common regulatory framework, and commonly accepted practices has resulted in a situation where different initiatives emerge across countries which share some common elements but also numerous differences that can lead to challenges related to interoperability. It is recommended to share between countries next to technical know-how additionally how social and ethical risks have been and are being managed.

Lastly, it is important to note that face manipulation techniques are also expected to have positive impact on society and economy. For instance, face manipulation techniques can help to address privacy issues through privacy-enhancing techniques, they facilitate the training of machine learning models with synthetic data (without images scrapped from the web), they can help with sustainability by facilitating virtual fitting rooms for the beauty and fashion industries and drive economic development with (high added value) mobile e-commerce, entertainment, and social media applications.

21.6 Summary

This concluding chapter has given an overview of different unsolved issues in (and surrounding) the research field of digital face manipulation and detection. It summarizes the opinions of several distinguished researchers from academia and industry of different backgrounds, including computer vision, pattern recognition, media forensics as well as social and legal research, regarding the future trends in said field. Moreover, this chapter has listed various avenues which should be considered in future research and, thus, serves as good reference point for researchers working in the area of digital face manipulation and detection.

This research work has been funded by

- The German Federal Ministry of Education and Research and the Hessian Ministry of Higher Education, Research, Science and the Arts within their joint support of the National Research Center for Applied Cybersecurity ATHENE.
- The ARRS Research Project J2–1734 "Face deidentification with generative deep models", and ARRS Research Programs P2–0250 (B) "Metrology and Biometric Systems" and P2–0214 (A) "Computer Vision".
- PRIMA (H2020-MSCA-ITN-2019-860315), TRESPASS-ETN (H2020-MSCA-ITN-2019-860813), BIBECA (MINECO/FEDER RTI2018-101248-B-I00), REAVIPERO (RED2018-102511-T), and COST CA16101 (MULTI-FORESEE).
- The Defense Advanced Research Projects Agency (DARPA) and the Air Force Research Laboratory (AFRL) under agreement number FA8750-20-2-1004. The U.S. Government is authorized to reproduce and distribute reprints for Governmental purposes notwithstanding any copyright notation thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as necessarily representing the official policies or endorsements, either expressed or implied, of DARPA and AFRL or the U.S. Government.
- The PREMIER project, funded by the Italian Ministry of Education, University, and Research within the PRIN 2017 program and by a Google gift.
- The European Commission, grant number 883356—Image Manipulation Attack Resolving Solutions (iMARS).
- The Cybersecurity Initiative Flanders, Strategic Research Program (CIF).

References

1. Barni M, Battiato S, Boato G, Farid H, Memon N (2020) MultiMedia forensics in the wild. In: International conference on pattern recognition
2. Biggio B, Korshunov P, Mensink T, Patrini G, Rao D, Sadhu A (2019) Synthetic realities: deep learning for detecting AudioVisual fakes. In: International conference on machine learning
3. Gregory S, Cristian C, Leal-Taixé L, Christoph B, Hany F, Matthias N, Sergio E, Edward D, McCloskey S, Isabelle G, Arslan B, Justus T, Luisa V, Hugo Jair E, Christa S, Andreas R, Jun W, Davide C, Guo G (2020) Workshop on media forensics. In: Conference on computer vision and pattern recognition
4. Kiran R, Naser D, Cunjian C, Antitza D, Adam C, Hu H, Raghavendra R (2020) Workshop on Deepfakes and presentation attacks in biometrics. In: Winter conference on applications of computer vision
5. Verdoliva L, Bestagini P. Multimedia forensics. In: ACM multimedia
6. Citron D (2019) How DeepFake undermine truth and threaten democracy. https://www.youtube.com/watch?v=pg5WtBjox-Y
7. Allcott Hunt, Gentzkow Matthew (2017) Social media and fake news in the 2016 election. J Econ Perspect 31(2):211–36
8. Suwajanakorn Supasorn, Seitz Steven M, Kemelmacher-Shlizerman Ira (2017) Synthesizing obama: learning lip sync from audio. ACM Trans Graph 36(4):1–13
9. Kietzmann J, Lee LW, McCarthy IP, Kietzmann TC (2020) Deepfakes: Trick or Treat? Bus Horiz 63(2):135–146

10. Tolosana Ruben, Vera-Rodriguez Ruben, Fierrez Julian, Morales Aythami, Ortega-Garcia Javier (2020) DeepFakes and beyond: a survey of face manipulation and fake detection. Inf Fusion 64:131–148
11. Verdoliva Luisa (2020) Media forensics and DeepFakes: an overview. IEEE J Sel Top Signal Process 14:910–932
12. Dolhansky B, Bitton J, Pflaum B, Lu J, Howes R, Wang M, Ferrer CC (2020) The DeepFake detection challenge (DFDC) dataset. arXiv:2006.07397
13. Jiang L, Li R, Wu W, Qian C, Loy CC (2020) DeeperForensics-1.0: a large-scale dataset for real-world face forgery detection. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition
14. Raja K, Ferrara M, Franco A, Spreeuwers L, Batskos I, Gomez-Barrero FD, Scherhag U, Fischer D, Venkatesh S (2020) In: Singh JM, Li G, Loïc B, Sergey I, Raghavendra R, Christian R, Dinusha F, Uwe S, Fons K, Raymond V, Davide M, Christoph B (eds) Evaluation platform and benchmarking. IEEE transactions on information forensics and security, morphing attack detection-database
15. Rössler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Nießner M (2018) FaceForensics: a large-scale video dataset for forgery detection in human faces. arXiv:1803.09179
16. Rössler A, Cozzolino D, Verdoliva L, Riess C, Thies J, Nießner M (2019) FaceForensics++: learning to detect manipulated facial images. In: Proceeding of the IEEE/CVF international conference on computer vision
17. Ruben Tolosana, Sergio Romero-Tapiador, Julian Fierrez, and Ruben Vera-Rodriguez. DeepFakes Evolution: Analysis of Facial Regions and Fake Detection Performance. In *Proc. International Conference on Pattern Recognition Workshops*, 2020
18. Mirsky Yisroel, Lee Wenke (2021) The Creation and Detection of Deepfakes: A Survey. ACM Computing Surveys 54(1):1–41
19. Li Y, Yang X, Sun P, Qi H, Lyu S (2020) Celeb-DF: a large-scale challenging dataset for DeepFake forensics. In: Proceeding of the IEEE/CVF conference on computer vision and pattern recognition
20. Scherhag Ulrich, Rathgeb Christian, Merkle Johannes, Breithaupt Ralph, Busch Christoph (2019) Face recognition systems under morphing attacks: a survey. IEEE Access 7:23012–23026
21. Venkatesh S, Ramachandra R, Raja K, Busch C (2021) Face morphing attack generation & detection: a comprehensive survey. In: IEEE transactions on technology and society
22. Zhang S, Chi C, Lei Z, Li SZ (2020) Refineface: refinement neural network for high performance face detection. In: IEEE transactions on pattern analysis and machine intelligence
23. Zhou Y, Liu D, Huang T (2018) Survey of face detection on low-quality images. In: Proceedings of the IEEE international conference on automatic face & gesture recognition, pp 769–773
24. Karras T, Laine S, Aittala M, Hellsten J, Lehtinen J, Aila T (2020) Analyzing and improving the image quality of StyleGAN. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition
25. Deng Y, Yang J, Chen D, Wen F, Tong X (2020) Disentangled and controllable face image generation via 3D imitative-contrastive learning. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition
26. Ciftci UA, Demir I, Yin L (2020) Fakecatcher: detection of synthetic portrait videos using biological signals. In: IEEE transactions on pattern analysis and machine intelligence
27. Hernandez-Ortega J, Tolosana R, Fierrez J, Morales A (2021) DeepFakesON-Phys: DeepFakes detection based on heart rate estimation. In: Proceedings of the 35th AAAI conference on artificial intelligence workshops
28. Gragnaniello D, Cozzolino D, Marra F, Poggi G, Verdoliva L (2021) Are GAN generated images easy to detect? A critical analysis of the state-of-the-art. In: Proceedings of the IEEE international conference on multimedia and expo
29. Marra F, Gragnaniello D, Verdoliva L, Poggi G (2019) Do GANs leave artificial fingerprints? In: Proceeding of the IEEE conference on multimedia information processing and retrieval

30. Neves Joã C, Tolosana Ruben, Vera-Rodriguez Ruben, Lopes Vasco, Proenca Hugo, Fierrez Julian (2020) GANprintR: improved fakes and evaluation of the state of the art in face manipulation detection. IEEE J Sel Top Signal Process 14(5):1038–1048
31. Deng J, Dong W, Socher R, Li LJ, Li K, Fei-Fei L (2009) ImageNet: a large-scale hierarchical image database. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition
32. Chung JS, Nagrani A, Zisserman A (2018) VoxCeleb2: deep speaker recognition. arXiv:1806.05622
33. Zhou T, Wang W, Liang Z, Shen J (2021) Face forensics in the wild. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition
34. Pang Min, He Ligang, Kuang Liqun, Chang Min, He Zhiying, Han Xie (2020) Developing a parametric 3D face model editing algorithm. IEEE Access 8:167209–167224
35. Giachanou A, Zhang G, Rosso P (2020) Multimodal multi-image fake news detection. In: Proceedings of the IEEE international conference on data science and advanced analytics
36. Singhal S, Kabra A, Sharma M, Shah RR, Chakraborty T, Kumaraguru P (2020) Spotfake+: a multimodal framework for fake news detection via transfer learning. In: Proceedings of the AAAI conference on artificial intelligence
37. Du M, Pentyala S, Li Y, Hu X (2020) Towards generalizable Deepfake detection with locality-aware AutoEncoder. In: Proceedings of the ACM international conference on information & knowledge management
38. Nguyen HH, Fang F, Yamagishi J, Echizen I (2019) Multi-task learning for detecting and segmenting manipulated facial images and videos. In: Proceedings of the IEEE international conference on biometrics theory, applications and systems
39. Cozzolino D, Thies J, Rössler A, Riess C, Nießner M, Verdoliva L (2018) ForensicTransfer: weakly-supervised domain adaptation for forgery detection. arXiv:1812.02510
40. Peng B, Fan H, Wang W, Dong J, Li Y, Lyu S, Li Q, Sun Z, Chen H, Chen B et al (2021) DFGC 2021: a DeepFake game competition. arXiv:2106.01217
41. Rathgeb C, Bernardo K, Haryanto NE, Busch C (2021) Effects of image compression on face image manipulation detection: a case study on facial retouching. IET Biom 10
42. Galbally Javier, Marcel Sebastien, Fierrez Julian (2014) Biometric anti-spoofing methods: a survey in face recognition. IEEE Access 2:1530–1552
43. Marcel S, Nixon MS, Fierrez J, Evans N (2019) Handbook of biometric anti-spoofing, 2nd edn
44. Ramachandra Raghavendra, Busch Christoph (2017) Presentation attack detection methods for face recognition systems: a comprehensive survey. ACM Comput Surv 50(1):1–37
45. Tolosana Ruben, Gomez-Barrero Marta, Busch Christoph, Ortega-Garcia Javier (2019) Biometric presentation attack detection: beyond the visible spectrum. IEEE Trans Inf Forensics Secur 15:1261–1275
46. Trinh L, Tsang M, Rambhatla S, Liu Y (2021) Interpretable and trustworthy DeepFake detection via dynamic prototypes. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision
47. Carlini N, Farid H (2020) Evading Deepfake-image detectors with white-and black-box attacks. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops
48. Gandhi A, Jain S (2020) Adversarial perturbations fool Deepfake detectors. In: Proceedings of the international joint conference on neural networks
49. Huang Y, Juefei-Xu F, Wang R, Xie X, Ma L, Li J, Miao W, Liu Y, Pu G (2020) FakeLocator: robust localization of GAN-based face manipulations via semantic segmentation networks with bells and whistles. arXiv:2001.09598
50. Huang R, Fang F, Nguyen HH, Yamagishi J, Echizen I (2020) Security of facial forensics models against adversarial attacks. In: Proceedings of the IEEE international conference on image processing
51. Hussain S, Neekhara P, Jere M, Koushanfar F, McAuley J (2021) Adversarial Deepfakes: evaluating vulnerability of Deepfake detectors to adversarial examples. In: Proceedings of the IEEE/CVF winter conference on applications of computer vision

52. Cozzolino D, Thies J, Rössler A, Nießner M, Verdoliva L (2021) SpoC: spoofing camera fingerprints. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops
53. Lago F, Pasquini C, Böhme R, Dumont H, Goffaux V, Boato G (2021) More real than real: a study on human visual perception of synthetic faces. arXiv:2106.07226v1
54. Nightingale Sophie J, Agarwal Shruti, Härkönen Erik, Lehtinen Jaakko, Farid Hany (2021) Synthetic faces: how perceptually convincing are they? Vision Sciences Society (VSS) meeting, In Proc
55. Meden Blaž, Emeršič Žiga, Štruc Vitomir, Peer Peter (2018) k-same-net: k-anonymity with generative deep neural networks for face deidentification. Entropy 20(1):60
56. Meden B, Mallı RC, Fabijan S, Ekenel HK, Štruc V, Peer P (2017) Face deidentification with generative deep neural networks. IET Signal Process 11(9):1046–1054
57. Mirjalili Vahid, Raschka Sebastian, Ross Arun (2019) FlowSAN: privacy-enhancing semi-adversarial networks to confound arbitrary face-based gender classifiers. IEEE Access 7:99735–99745
58. Giovanni C, Luisa P, Verdoliva D (2019) Extracting camera-based fingerprints for video forensics. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops
59. Perera P, Oza P, Patel VM (2021) One–class classification: a survey, pp 1–19. arXiv:2101.03064
60. Hoi SC, Sahoo D, Lu J, Zhao P (2021) A comprehensive survey. Neurocomputing, Online Learning
61. Dong Xibin, Zhiwen Yu, Cao Wenming, Shi Yifan, Ma Qianli (2020) A survey on ensemble learning. Front Comput Sci 14(2):241–258
62. Sagi O, Rokach L (2018) Ensemble learning: a survey. Wiley Interdiscip Rev Data Min Knowl Discov 8(4)
63. Montavon G, Binder A, Lapuschkin S, Samek W, Müller KR (2019) Explainable AI: interpreting, explaining and visualizing deep learning, vol 11700. Springer Nature
64. Cozzolino D, Rössler A, Thies J, Nießner M, Verdoliva L (2021) ID-reveal: identity-aware DeepFake video detection. arXiv:2012.02512
65. Rathgeb C, Satnoianu C-I, Haryanto NE, Bernardo K, Busch C (2020) Differential detection of facial retouching: a multi-biometric approach. IEEE Access 8:106373–106385
66. Scherhag U, Rathgeb C, Merkle J, Busch C (2020) Deep face representations for differential morphing attack detection. In: IEEE transactions on information forensics and security
67. Hsu LY, Hu HT (2020) Blind watermarking for color images using EMMQ based on QDFT. Expert Syst Appl 149
68. Khare P, Srivastava VK (2021) A secured and robust medical image watermarking approach for protecting Iitegrity of medical images. Trans Emerg Telecommun Technol 32(2)
69. Terhörst P, Huber M, Damer N, Rot P, Kirchbuchner F, Struc V, Kuijper A (2020) Privacy evaluation protocols for the evaluation of soft-biometric privacy-enhancing technologies. In: 2020 International conference of the biometrics special interest group (BIOSIG), pp 1–5
70. Newton EM, Sweeney L, Malin B (2005) Preserving privacy by de-identifying face images. IEEE Trans Knowl Data Eng 17(2):232–243
71. Sweeney Latanya (2002) K-anonymity: a model for protecting privacy. Int J Uncertain Fuzziness Knowl-Based Syst 10(5):557–570
72. Croft WL, Sack JR, Shi W (2019) Differentially private obfuscation of facial images. In: Proceedings of the international cross-domain conference for machine learning and knowledge extraction
73. Dwork C (2008) Differential privacy: a survey of results. In: Proceedings of the international conference on theory and applications of models of computation
74. Tiits Marek, Kalvet Tarmo, Mikko Katrin-Laas (2014) Analysis of the e-passport readiness in the EU. Institute of Baltic Studies, Technical report, Tartu
75. Tiits M, Kalvet T, Mikko K-L (2014) Social acceptance of e-passports. In: Proceedings of the international conference of the biometrics special interest group