



Hybrid Identity Deepfakes: A Comprehensive Review of Multi-Person Facial Blending and Its Impact on Forensic Detection

Ashwini M^{1*}, Sai Shivani²

^{1,2}B.Sc. Hons., Digital Forensics, Malla Reddy University, Maisammaguda, Dulapally, Hyderabad-500043, Telangana, India

ashwinireddyym@gmail.com*, shivanikesireddy@gmail.com

Abstract

Deep learning techniques have brought a revolution to digital media synthesis that can generate hyper realistic manipulated content, called deepfakes. Earlier image manipulation techniques included face morphing and manual compositing operated through basic geometric and pixel level processes, which limited both visual realism and practical scalability. The development of deep generative models through VAEs and GANs led to a major shift towards using data to learn complex visual and temporal patterns. The integration of attention mechanisms with transformer architectures and diffusion-based models has led to better synthesis quality for images and videos and multimodal content, which has significantly reduced the difference between artificial and real media. Scientists have worked to develop forensic methods that detect differences between authentic data and manipulated content through the development of handcrafted feature-based analysis, which has progressed into deep neural models that detect spatial and temporal inconsistencies. The rapid technology has created an ongoing competition that reduces the power of conventional detection indicators while showing how models perform when operating in real-time environments. Deepfake systems use hybrid synthesis methods to create new identities through face swapping by combining different facial features from various people. The multiple identity deepfakes create problems for forensic models because their generated faces lack any connection to actual people while showing signs of digital distortion. The research evaluates multiple-person image blending through a classification system which includes four main approaches: morphing-based, latent-space-based, temporal-based based and part-based facial compositing. The research demonstrates how hybrid identities break down three essential identity verification methods, which include artefact-based, identity-consistency checks and temporal cues. The paper also sketches future research directions that focus on developing identity coherence models that are resistant to attacks.

Keywords: Deepfake detection, Deep Learning Models- GANs & VAEs, Face morphing, Multi-identity synthesis, Digital media forensics.

1. Introduction

The development of artificial intelligence over the last ten years has brought continuous transformations to digital media production and editing processes, which now include visual content manipulation through deepfake technology. Deep learning models produce deepfakes, which use their capabilities to create perfect duplicates of human faces, voices and body movements in audio, video and image files ^[1].

The first attempts at digital image editing is through morphing, face swapping, and texture blending that required users to perform manual work which uses basic mathematical operations ^[2]. The methods introduced new approaches, but they failed to generate expressions that matched real human behaviour, while they also struggled to produce realistic lighting effects and maintain continuous video playback.

Deep learning technology brought a complete transformation to digital media synthesis because it allowed models to discover complex patterns from extensive visual and audiovisual datasets through self-learning processes ^[3]. The development of Variational Autoencoders (VAEs) and Generative Adversarial Networks (GANs) marked a crucial point in this evolutionary process. VAEs brought a probabilistic learning approach, which learns the statistical patterns from datasets; their output results showed less detailed information than GAN-based methods, which followed later but still produced realistic samples through reconstruction control ^[4].

The GANs system introduced an adversarial learning approach, which used a generator network to create artificial content while a discriminator network worked to distinguish between authentic and artificially produced data ^[5]. The competitive optimisation system operates to enhance both networks through continuous improvement, which produces realistic yet, synthetic output results. GAN-based architectures now serve as the primary method to generate high-quality synthetic media because they allow deepfake applications to perform facial replacement, attribute manipulation, expression transfer, and complete scene synthesis with superior realism ^[6].

Research following this study showed that generative models achieve better results when scientists add advanced architectural elements, which include attention mechanisms, transformer networks, and diffusion-based processes. The integrated systems now enable generative systems to produce more than static images because they can generate videos, speech and create content that combines different media types ^[3]. Research based on surveys demonstrates that modern deepfake creation methods now use advanced techniques, which include conditional GANs, progressive training methods, attention-based generation and diffusion models to produce realistic facial details, natural body movements and small behavioural indicators ^[7].

The development of new technologies has made it possible to generate synthetic media that duplicates human appearance and behaviour patterns at a level that threatens to compromise forensic evidence analysis ^[8]. The research community has worked to create deepfake detection methods that identify fake content from authentic media while generative modelling continues

to advance. The detection methods extend from basic handcrafted feature-based detection systems to current deep learning systems, which detect synthetic content through its spatial, temporal and cross-modal inconsistencies^[9].

The identification of spatial artifacts in videos, use Convolutional Neural Networks (CNNs) as a standard approach, while Recurrent Neural Networks (RNNs) and temporal models detect anomalies that appear between consecutive video frames^[10]. Research has introduced transformer-based architectures combined with hybrid multimodal frameworks that use visual, audio, and physiological signals to enhance system performance and adaptability^[11].

Detection systems were somewhat effective against early deepfake methods because they were able to spot obvious visual artifacts, colour mismatches, and temporal inconsistencies. Such vulnerable systems could not last long against more sophisticated deepfake methods, which taught substitutes to pull off more advanced temporal synthesis and accurate emotion-matching against targeted faces; they prompted attempts to reveal weak forensic synthesis cues by using more efficient systems for varying texture faces, illuminations and motion synthesis^[12].

Present deepfake methods and systems can pull off demanding manipulations a lot more easily, but are way more difficult to spot compared to systems of accuracy and manipulation level of the GANs with more prominent motion interplay. Detection models are often built and tested on null benchmark datasets such as FaceForensics, the Deep Fake Detection Challenge and CelebDF, which severely limit the ability of the models to work under crippling yet real-world scenarios^[13],^[14]. These artifacts form the basis of real-world scenarios, like compression artifacts, resolution, re-encoding and adversarial post-processing used to obscure forensic traces of detection systems. These datasets train detection systems, are dependent on detection architectures and adversarial generation techniques, which only offer a narrow scope and limited performance generalization of detection systems^[15].

2. Literature Review

2.1 Multiple Image Blending

The process of deepfake synthesis requires image blending operations to merge visual elements, spatial positions and boundary connections between different source images, which are used to create a single target frame or video sequence. This process becomes much more complicated when multiple people participate in it. The system needs to handle multiple identities simultaneously while it solves the problem of body part, facial area intersections, and it needs to keep faces in front of bodies while preserving lighting effects, colour accuracy and movement patterns throughout the sequence. The blending and harmonization stages of deepfake creation which researchers call seamless cloning and composite refinement determines the level of naturalness and detectability in synthetic content.

2.2 Challenges in Detecting Deepfakes Created Through Multi-Person Image Blending

The process of multi-person image blending for deepfake creation leads to multiple obstacles when trying to detect these fake videos.

The detection of deepfakes, which combine visual elements from different people, becomes challenging because these fake videos deviate from the typical identity substitution pattern, which most forensic studies have focused on^[16],^[17]. The current detection methods operate

under the assumption that manipulated samples contain only one donor identity, which gets applied to a target person [18]. These systems operate to detect irregularities, which include abnormal face borders, wrong light effects and inconsistent 3D shapes [19]. The assembly of synthetic faces from different sources makes these standard indicators completely unreliable because they no longer meet their original assumptions [16], [20].

The main challenge arises because multi-identity blending creates synthetic facial images which lack any connection to actual human beings since these generated faces exist only as artificial creations thus creating an entirely new identity [21]. The detection models which use face embeddings to identify known identities and check face synthesis inconsistencies cannot function because they lack a reference identity to compare against [16]. The forged identity exists beyond the predicted embedding space which creates difficulties for detectors to identify irregularities [16], [22].

The process of uniting multiple facial components from different people results in a distributed artifact pattern which creates additional challenges. The system generates localized distortions which tend to appear in specific areas of the system (e.g., the facial errors appear as small irregularities which spread across the entire face, while the main distortion occurs near the jawline or hairline [17], [23]. The convolution-based forensic methods detect concentrated irregularities like warped edges, blending seams and colour inconsistencies which fail to identify these artifacts, because they are spread out across the entire image. The filter or kernel in convolutional layers examine individual elements within image patches instead of trying to detect diffused artifact patterns [23], [24]. The process of merging information from different sources makes it harder to analyse time-based evidence which deepfake detection systems depend on for their analysis.

Standard single-identity manipulations produce temporal artifacts because the models either become unstable or their features do not match between consecutive frames [25]. The output becomes more refined when the generative model receives input from multiple users because their facial movements create less restrictive patterns [24]. The system produces an identity which shows all standard time-based visual inconsistencies including texture flickering, expression instability and geometric distortions [23], [25].

The detection of physiological signals including pulse-induced colour changes, natural blinking rhythms and micro-expression coherence becomes difficult because of multi-person composites [26]. In general, such biological patterns of a single person create the cues which people use to identify them. The combination of different identities results in physical evidence that becomes unstable and produces weak signals which models interpret as being in between realistic and unrealistic levels [26], [27].

The lack of accessible multi-person blended forgery datasets in public databases makes the challenge even more difficult to overcome [17], [21]. The detection models lack the ability to learn multi-identity composite statistical biometrics, because benchmark datasets including FaceForensics++, Celeb-DF and DFDC present only single-source manipulation data. The combination of these problems makes it harder to detect deepfakes which combine footage from different people so researchers need to develop new forensic methods to analyse hybrid identities, distributed artifacts, complex temporal and physiological patterns [20], [21].

3. Taxonomy of Multi-Individual Deepfake Image Synthesis

The process of deepfake image synthesis involves multiple people, requires a systematic

approach to identify when different sources of identity information get merged during the generation pipeline. Different methods operate at different levels because they either perform identity fusion within pixel data or they do it at the latent representation level or semantic facial component level or by analysing video sequences through time. The following section examines the main deepfake synthesis methods which involve multiple individuals while explaining their working principles and their restricted capabilities [31]-[33].

a. Face Morphing-Based Synthesis

Face morphing stands as one of the first and most researched methods for creating multiple identity images [34]. The morphing-based generation method uses two or more facial images which it aligns and combines to produce a single face that keeps all the biometric features from each original face. The main goal of this process involves creating images which look both realistic and also make it difficult for automated systems to identify the person in the image [35].

The system performs two main operations which include facial landmark detection and alignment of source images through eye, nose, mouth and jawline point identification followed by geometric normalization [34]. The process of shape interpolation uses landmark coordinate averaging or weighting to create a single facial geometry that represents all faces. The system also performs operations through its texture or pixel interpolation process which combines matching sections from each identity within the image space. It performs global colour and illumination normalization to minimize acquisition-related differences between images.

The basic structure of these pipelines remains easy to understand but modern morphing attacks now use multiple additional processing steps. This process includes three main steps which are spatial smoothing, manual retouching and neural enhancement/refinement networks that work to eliminate visual artifacts like ghosting, edge discontinuities and texture inconsistencies [36]. Face morphing technology creates an extreme security risk when people need to enroll their identities for passport and ID card programs. A single morphed image when enters a biometric database system will later verify different people thus breaking the fundamental link between personal identity and biometric data [35].

b. Latent-Space Identity Blending Using Generative Models

Deep generative modelling has evolved to create new methods which allow better control over multi-person synthesis through latent-space identity blending [32], [37]. The methods use different approaches than morphing-based methods because they perform identity combination through learned representation spaces which stems from Generative Adversarial Networks (GANs) and Variational Auto Encoders (VANS) based architectures.

The standard latent-space blending framework operates as follows: The first step of processing each source face involves using an encoder or inversion process to create a small latent identity representation [38], [39]. The research combines identity embeddings from different people through three methods which include interpolation, linear combination and vector concatenation. The network transforms the merged latent information into a generated facial image.

The operating system provides multiple benefits to users who work in latent space instead of using pixel-level blending. The system provides users with continuous and complete control to adjust the amount of influence each identity has on the facial attributes. The process of blending

takes place inside a facial structure which people learn to recognize. So, the method minimizes the typical problems which occur in traditional morphing techniques [38]. The system benefits from these advantages, but its ability to separate identities from each other remains limited because of the challenge to achieve stable identity separation. The hidden representations in data systems combine personal identity information with other non-personal attributes which include body positions, facial emotions and environmental illumination [39]. The blending process creates two problems, which make it difficult to control synthesis and perform forensic analysis: transferring unwanted properties and making one identity dominate the other [40].

c. Part-Based Face Compositing with Neural Refinement

The deepfake synthesis of multiple individuals requires a different approach which involves building facial components as individual elements. These methods choose specific facial areas from different people to create a new face by combining the eyes, nose, mouth and jaw [42].

The system uses two methods to achieve spatial alignment of extracted facial components through landmark-based alignment and dense correspondence models. The system performs direct compositing of chosen areas which get combined with a target facial canvas. The system applies neural refinement networks to enhance visual coherence in its results. The process of neural refinement serves as a vital component for this system [42].

The process of seam removal, texture, shading, colour distribution and reconciliation in composited areas uses specialized networks which were trained for image inpainting, boundary smoothing and illumination harmonization. The successful execution of this process results in a realistic face that unites identity characteristics which stems from different information sources. The attack method of part-based composition enables attackers to choose which biometric elements they choose to modify. The system remains sensitive to occlusions which occur when parts of the face becomes hidden from view and it also struggles with complicated facial movements including hair covering the face, wearing glasses, having facial hair and making strong facial expressions. The extraction process for components becomes more difficult because of these elements which create hidden forensic traces that detection systems can identify [40].

d. Temporally Consistent Multi-Identity

The process of video synthesis becomes more complicated when multiple individuals need to generate video content because it requires maintaining consistent timing between frames. The independent application of identity blending to each frame through identity blending results in temporal artifacts, which produce flickering textures, facial geometry inconsistencies and unstable identity proportions during time progression [41].

The solution of these problems in modern video-based synthesis pipelines depends on the following mechanisms: temporal discriminators that enforce consistency across consecutive frames, optical-flow-guided warping to transfer facial structure and appearance data between consecutive video frames [41]. The system also uses recurrent or attention-based refinement modules, which explicitly model temporal dependencies.

The methods work to preserve stable blended identities when changes occur in head position, facial expressions and lighting conditions, while preserving authentic movement patterns. The research community has not found a solution to maintain long-term temporal stability when videos show rapid motion, objects hide from view. The process of preserving continuous identity contributions through long sequences becomes challenging when different source

identities present themselves with distinct facial structures and visual appearances^{[31], [40]}.

4. Open Issues and Future Directions

Most current forensic models face major obstacles when trying to detect multi-person image blending deepfakes because they rely on single-identity assumptions, which do not hold. The solution for blended face identification needs to focus on identity coherence analysis because blended faces create entirely new synthetic identities that differ from actual identities. Real face analysis enables detectors to understand how different facial elements relate to each other through statistical methods which helps them check if the generated face has proper anatomical and biometric measurements. The internal relationships between facial features become disrupted when researchers combine multiple individual faces into hybrid faces which appear realistic to the human eye thus making coherence-based analysis an appropriate substitute for identity matching.

The process of multi-source blending creates artifacts which spreads throughout the output data. The system produces blended identities through its method which disperse small facial irregularities throughout the entire face instead of creating traditional deepfake boundary distortions. The solution requires detection models to use global feature learning methods which examine complete face texture information instead of analysing limited image sections. The combination of multi-scale spatial and frequency-domain analysis enables researchers to identify unnatural smoothing, texture homogenization and blending-related distortions which standard local convolutional filters cannot detect.

The reduced visibility of temporal artifacts creates an additional problem because multi-identity synthesis generates videos that has more consistent transitions between frames. The analysis of identity-related features across extended time periods should replace the current methods which measure visual flicker during short periods. The detection of blended identities becomes possible through the analysis of facial geometry ratios and feature embeddings which span multiple video segments to detect the slow changes that occur in blended identities but not in actual video footage.

The process of detecting physiological cues becomes more difficult when faces are blended, because biological signals show up as inconsistent or diluted patterns instead of obvious fake signals. The detection systems should concentrate on facial region signal coherence instead of seeking individual-specific strong patterns. The natural synchronization of pulse-related colour changes, blinking rhythms and micro-expression activation occurs in authentic facial expressions. The analysis of multi-person blended faces reveals biological response inconsistencies which can serve as important evidence for forensic investigations.

The absence of multi-person blended deepfake datasets creates a major challenge because it prevents models from learning to recognize these types of fake content. The problem can be solved through specific synthetic data creation which involves controlled multi-identity blending through morphing, latent-space interpolation and part-based compositing techniques. The training process of detectors with these different hybrid samples helps them understand how blended identities appear statistically instead of becoming specialized in detecting individual manipulation techniques.

The development of new blending methods creates ongoing challenges for scientists who want to create detection systems that can identify all existing types of forgeries. Self-consistency and anomaly-based learning approaches should be used to solve this problem. The detection of blended deepfakes becomes possible through model training which uses only real facial data to recognize typical human behaviour that occurs under different lighting conditions, facial expressions and head positions. The detection system becomes more resistant to attacks through this approach because it continues to function properly when facing new multi-person blending techniques.

5. Conclusion

Advancements in the field of deepfake video generation and the challenges against detection, particularly the blending of multiple subjects, continue to grow. Early methods of deepfake video generation, such as image face morphing and overlay compositing, suffered from lack of realism. However, modern deepfake videos using GANS and other diffusion based autoencoder models provide far greater levels of realism, resulting in new challenges in video forensics. The challenges are especially concerning as identities are blended to create hybrid faces that do not represent any actual people.

The research constructed a taxonomy for the multi-person deepfake videos, such as morphing, latent space identity blending, compositing of facial features, and the generation of temporally consistent videos. These videos also show that conventional methods of detection are circumvented by the creation of attenuated temporal artifacts, spatial incoherence, absent physiological signals, and the blending of faces, ultimately impacting the ability of the detection systems to rely on identity or component-artifact localization.

The absence of a well curated, multi-identity (also known as multi-actor) benchmark datasets also serves as a major limitation on research in this field. Therefore, detection efforts should aim for multi-scale and holistic identity coherent analysis; consistency in long-term timelines; and physiological region interdependence within signals. Training self-consistency and anomaly detection models, grounded on real data, can yield greater detection of the latest multi-person deep fake video synthesis methods.

Acknowledgement:

We would like to express our sincere gratitude to **Mr. Dhanunjaya Rao Reddy**, Department of Digital Forensic Science, Malla Reddy University, for his valuable guidance, constructive suggestions, and continuous support throughout the course of this research. His academic expertise and feedback significantly contributed to the direction, methodology, and overall quality of the study. We also acknowledge the academic environment and resources provided by Malla Reddy University, which facilitated the successful completion of this work.

References

- [1] Y. Mirsky and W. Lee, "The Creation and Detection of Deepfakes: A Survey," *ACM Computing Surveys*, vol. 54, no. 1, 2021.
- [2] V. Blanz and T. Vetter, "A Morphable Model for the Synthesis of 3D Faces," *SIGGRAPH*, 1999.
- [3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning*, MIT Press, 2016.
- [4] D. P. Kingma and M. Welling, "Auto-Encoding Variational Bayes," *arXiv:1312.6114*, 2013.
- [5] I. Goodfellow et al., "Generative Adversarial Nets," *NeurIPS*, 2014.
- [6] A. Tolosana et al., "DeepFakes and Beyond: A Survey of Face Manipulation," *Information Fusion*, vol. 64, 2020.
- [7] G. Pei et al., "Deepfake Generation and Detection: A Benchmark and Survey," *arXiv:2403.17881*, 2024.
- [8] H. Farid, "Digital Forensics in a Post-Truth Age," *Forensic Science International*, 2019.
- [9] S. Agarwal et al., "A Comprehensive Survey on Deepfake Detection," *Electronics*, vol. 13, no. 3, 2024.
- [10] D. Güera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," *AVSS*, 2018.
- [11] J. Yang et al., "Audio-Visual Deepfake Detection Using Transformers," *IEEE TIFS*, 2023.
- [12] A. Rossler et al., "FaceForensics++," *ICCV*, 2019.
- [13] B. Dolhansky et al., "The Deepfake Detection Challenge Dataset," *arXiv:2006.07397*, 2020.
- [14] Y. Li et al., "Celeb-DF: A New Dataset for DeepFake Forensics," *CVPR Workshops*, 2020.
- [15] S. Popovic et al., "Deepfake Detection: Generalization Challenges," *Digital Investigation*, 2025.
- [16] S. Dong, S. Wang, R. Ji, J. Li, and S. Wang, "Implicit Identity Leakage: The Stumbling Block to Improving Deepfake Detection Generalization," *arXiv preprint arXiv:2210.14457*, 2022.
- [17] G. Pei, X. Yang, L. Li, X. Sun, and J. Huang, "Deepfake Generation and Detection: A Benchmark and Survey," *arXiv preprint arXiv:2403.17881*, 2024.

- [18] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A Large-Scale Challenging Dataset for DeepFake Forensics," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, Seattle, WA, USA, 2020, pp. 3207–3216.
- [19] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, Seoul, South Korea, 2019, pp. 1–11.
- [20] L. Verdoliva, "Media Forensics and DeepFakes: An Overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 910–932, Aug. 2020.
- [21] F. Marra, D. Gagnaniello, D. Cozzolino, and L. Verdoliva, "Detection of GAN-Generated Fake Images over Social Networks," in *Proc. IEEE Conf. Multimedia Information Processing and Retrieval (MIPR)*, Miami, FL, USA, 2018, pp. 384–389.
- [22] P. Korshunov and S. Marcel, "Deepfakes: A New Threat to Face Recognition? Assessment and Detection," *arXiv preprint arXiv:1812.08685*, 2018.
- [23] D. Cozzolino, G. Poggi, and L. Verdoliva, "Recasting Residual-Based Local Descriptors as Convolutional Neural Networks: An Application to Image Forgery Detection," in *Proc. ACM Workshop on Information Hiding and Multimedia Security*, Philadelphia, PA, USA, 2017, pp. 159–164.
- [24] X. Wang, Y. Wu, and P. Xue, "CNN-Based Detection of Diffused Image Manipulation Artifacts," *Pattern Recognition*, vol. 107, pp. 107504, 2020.
- [25] Y. Li and S. Lyu, "Exposing DeepFake Videos by Detecting Face Warping Artifacts," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition Workshops (CVPRW)*, Long Beach, CA, USA, 2019, pp. 46–52.
- [26] H. Guera and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks," in *Proc. IEEE Int. Conf. Advanced Video and Signal Based Surveillance (AVSS)*, Auckland, New Zealand, 2018, pp. 1–6.
- [27] T. B. Nguyen, C. M. Nguyen, D. T. Nguyen, D. T. Nguyen, and S. Nahavandi, "Deep Learning for Deepfakes Creation and Detection: A Survey," *Computers*, vol. 8, no. 3, pp. 1–23, 2019.
- [28] S. Lyu, "Detecting DeepFake Videos in the Wild," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, Taipei, Taiwan, 2019, pp. 1–5.
- [29] Y. Li, M. Chang, and S. Lyu, "In Ictu Oculi: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking," in *Proc. IEEE Int. Workshop on Information Forensics and Security (WIFS)*, Hong Kong, 2018, pp. 1–7.
- [30] B. Dolhansky *et al.*, "The DeepFake Detection Challenge (DFDC) Dataset," *arXiv preprint arXiv:2006.07397*, 2020.
- [31] Y. Mirsky and W. Lee, "The creation and detection of deepfakes: A survey," *ACM Computing Surveys*, vol. 54, no. 1, pp. 1–41, 2021.

- [32] A. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, "Deepfakes and beyond: A survey of face manipulation and fake detection," *Information Fusion*, vol. 64, pp. 131–148, 2020.
- [33] G. Pei, X. Yang, L. Li, X. Sun, and J. Huang, "Deepfake generation and detection: A benchmark and survey," *arXiv preprint arXiv:2403.17881*, 2024.
- [34] V. Blanz and T. Vetter, "A morphable model for the synthesis of 3D faces," in *Proc. ACM SIGGRAPH*, Los Angeles, CA, USA, 1999, pp. 187–194.
- [35] M. Ferrara, A. Franco, and D. Maltoni, "The magic passport," in *Proc. IEEE Int. Joint Conf. Biometrics (IJCB)*, Clearwater, FL, USA, 2014, pp. 1–7.
- [36] U. Scherhag, C. Rathgeb, J. Merkle, and C. Busch, "Detection of face morphing attacks based on deep learning," *IEEE Transactions on Information Forensics and Security*, vol. 14, no. 4, pp. 1–16, 2019.
- [37] I. Goodfellow, J. Pouget-Abadie, M. Mirza, et al., "Generative adversarial nets," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2014, pp. 2672–2680.
- [38] T. Karras, S. Laine, and T. Aila, "A style-based generator architecture for generative adversarial networks," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 4401–4410.
- [39] R. Abdal, Y. Qin, and P. Wonka, "Image2StyleGAN: How to embed images into the StyleGAN latent space," in *Proc. IEEE/CVF Int. Conf. Computer Vision (ICCV)*, 2019, pp. 4431–4440.
- [40] L. Verdoliva, "Media forensics and deepfakes: An overview," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 5, pp. 910–932, Aug. 2020.
- [41] H. Guera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in *Proc. IEEE Int. Conf. Advanced Video and Signal Based Surveillance (AVSS)*, 2018, pp. 1–6.
- [42] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner, "Neural textures: Learning-based rendering for real-time facial reenactment," in *Proc. IEEE/CVF Conf. Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 3387–3396.

Open Access This chapter is licensed under the terms of the Creative Commons Attribution-NonCommercial 4.0 International License (<http://creativecommons.org/licenses/by-nc/4.0/>), which permits any noncommercial use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license and indicate if changes were made.

The images or other third party material in this chapter are included in the chapter's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the chapter's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder.

