Int. J. Nonlinear Anal. Appl. 13 (2022) No. 1, 745-755
ISSN: 2008-6822 (electronic)
http://dx.doi.org/10.22075/ijnaa.2022.5580

A landscape view of deepfake techniques and detection methods

Ahmed S. Abdulreda[a,*], Ahmed J. Obaid[a]

[a]*Faculty of Computer Science and Mathematics, University of Kufa, Iraq*

(Communicated by Madjid Eshaghi Gordji)

Abstract

Deep fakes is the process of changing the information of the image or video with different techniques and methods that start with humor and fun and sometimes reach economic, political and social goals such as counterfeiting, financial fraud or impersonation. The data for this field is still increasing at a very high rate. And therefore. The process of combating and exploring them is a very difficult task. In this paper, we conducted a review of previous studies and what researchers dealt with on the subject of deep fakes. Explain the concepts of deepfakes. Counterfeiting methods and techniques and patterns through the techniques and algorithms used in counterfeiting. Some deepfake detection algorithms.

Keywords: Manipulation, faking, deepfakes, counterfeiting.

1. Introduction

Recent public outrage over forgeries and digital modification of videos and pictures, particularly utilizing DeepFake techniques [23, 35, 46]. Phrase "DeepFake" refers to a technique based on deep learning that generates false films in which one person's face is swapped with that of another. The phrase gained popularity when Reddit user deepfakes claimed in late 2017 that he had created a machine learning algorithm capable of transforming celebrities' visage into pornographic films [4]. Along with false pornography, some of the more pernicious applications of this kind of material Its content fake news, frauds, And counterfeiting in economic and financial matters. As a consequence, the area of study formerly dedicated to public multimedia forensics has been revived [27, 24], with an emphasis on identifying face alteration in pictures and video [40]. A portion of these reinvigorated

*Corresponding author
Email addresses: Ahmeds.albudairi@student.uokufa.edu.iq (Ahmed S. Abdulreda), Ahmedj.aljanaby@uokufa.edu.iq (Ahmed J. Obaid)

Received: August 2021 *Accepted:* September 2021

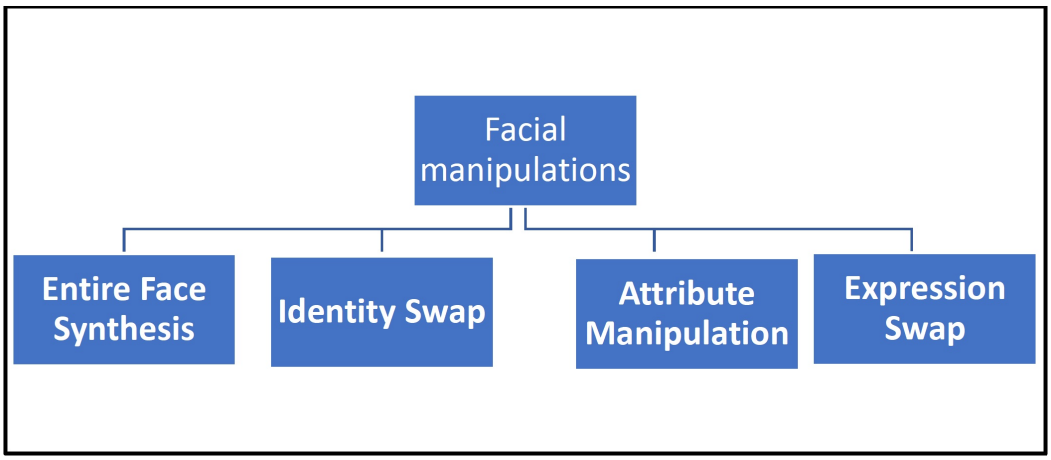

Figure 1: manipulation categories

efforts in detecting fake faces is based on prior research on biometric spoofing [11, 29] and deep learning by data-driven [34, 8]. The rising number of seminars at key conferences demonstrates the growing in fake facial detection [38, 10].

Face manipulation's quantity and realism have been limited by a lack of advanced editing tools, subject knowledge, and a complicated and time-consuming procedure. For example, early research on this subject [5] shown that it is possible to modify the lip movement of a speaking person by establishing links between the voice route sounds and the contour of the person's face. However, many things have changed dramatically in the years since these early works at the moment, it has become easier to automatically create non-existent faces or manipulate a single person's natural face in a photo/video, owing to 1) increased access to large-scale public data and 2) the development of deep learning techniques that eliminate many manual editing steps, such as automated programmers (AE) and adversarial generative networks (GAN) [22, 26].

However, in the field of facial recognition, large-scale public datasets have been scarce, and as a result, the majority of recent advancements in the community have been confined to Internet behemoths such as Facebook and Google. For instance, Google's most current technique for facial recognition was trained using 200 million pictures and eight million unique identities. This collection is almost three orders of magnitude bigger in size than any other publicly accessible face dataset. Needless to say, most worldwide research organizations, especially those in academia, are incapable of assembling a dataset this big [38].

Facial alterations may be classified into four distinct categories based on their degree of manipulation. Each face modification group is summarized visually in Figure 1. Each of them is described below, in order of increasing manipulation difficulty:

2. Techniques of facial modification

2.1. Entire Face Synthesis

This technique produces complete face pictures that do not exist, often using a powerful GAN, such as the recently suggested StyleGAN method in [20]. These technologies provide outstanding results, resulting in high-quality face pictures that are very realistic. In Figure 1, many instances of full-face synthesis produced using StyleGAN5 are shown. This manipulation may help a variety of businesses, including video games and 3D modeling, but it can also be used maliciously, such as generating very convincing false profiles.

It is easy for a person to describe a picture, which we learn to do from a young age. Machine learning, these are discriminative classification/regression, predictions Recent advances in ML/AI models, in special learning models, in distinct forms of character communication, as demonstrated in tasks such as visual object recognition (for example, from AlexNet to ResNet to ResNet (for ImageNet classification) and detection/segmentation of objects (eg, from RCNN to YOLO in a COCO dataset), etc.
However, the reverse task of wholesaling realistic images than description is more complex and requires many years of training in graphic design. In machine learning, this is a generative task, and it is more generative complex than tasks from division, where the generative model must generate more information (for example, a complete picture at a certain level of variance) based on the data of the first smaller details.
Despite the complexity of such applications, generative (with some control) is useful in a lot of cases:

2.1.1. Content Creation

The designer seeks inspiration by ordering the algorithm to retail 20 styles of brand shoes from her "Comfort", "Summer" and "Sentimental" batteries. New game development Create realistic avatars from a simple description.
Intelligent content-based editing: Photographer changes facial expressions, top wrinkles and hair styling in a photo with just a few clicks. get up

2.1.2. Data Augmentation

The drone developer can aggregate video data. The bank can be presented from the range of fraudulently presented types of operations poorly in the current data set

2.2. Identity Swap

This manipulation swaps a person's face in a video with another person's face. Typically, two methods are considered: 1) conventional methods based on computer graphics, such as FaceSwap6, and 2) new deep learning techniques dubbed DeepFakes7, such as the current ZAO smartphone application. Ultra-realistic films demonstrating this kind of modification are available on YouTube 8.
This kind of manipulation may help a variety of industries, particularly the film business; but, it can also be used for malicious reasons such as producing pornographic films of celebrities, frauds, and financial fraud, to name a few. Previously, to create a believable deep fake, hours of source video showing the target's face were required. So, at first, its use was restricted to celebrities, politicians and other public figures. Recent advances in machine learning have allowed fakes to be created using a single image of a target and only 5 seconds of its sound. Nowadays, it is common for people to post pictures and videos of themselves on social media, which is all an attacker needs to create a realistic fake. Does that sound scary? Yes it is. The goal has changed.
This disinformation or manipulation can influence public opinion by targeting political figures by creating fake footage of them saying things they never spoke, or committing actions that were never done before. What many of those interested in this topic do not know, is that deep-fake techniques may be used in phishing operations through voice calls or video calls.

2.3. Attribute Manipulation

Its also known facial modification or facial retouching, this technique entails altering some facial characteristics such as hair or skin color, beard, gender, moustache, and age, as well as the addition of spectacles. Typically, it is done using a GAN, and the StarGAN method proposed in [7]. The

popular FaceApp smartphone application is an example of this kind of manipulation. Consumers may use this technique to virtually using on a variety of goods, including cosmetics, make-up, eyeglasses, and haircuts.
Rapid advances in "deepfake" technology are making your face just another piece of personal data that you need to protect from theft. An iPhone app that has gone viral, period, makes creating fake videos just as easy as taking selfies.
The Chinese face swapped "Zao" massively recently, and it reached the top in China's iOS App Store, and hasn't budged from the number one spot since then. Not bad with it ironed out over a short period.
According to US website Gizmodo, Bloomberg News reported that the app belongs to Momo Inc.

2.4. Expression Swap

Also known as facial re-enactment, this kind of manipulation involves altering a person's facial expressions. Although many processing methods have been suggested in the literature, for example, at the picture level through standard GAN architectures [28], this category focuses on the most widely used techniques Neural-Textures and Face2Face [44, 45]. The face is the expression of a person in a video in comparison to the facial expression of another person. This kind of deception may have severe repercussions, as shown by the well-known video of Mark Zuckerberg stating things he never uttered.
Deepfake technology has been used in many ways to target people in all walks of life. Not only has it been used to create fake photos and videos of celebrities and politicians, but this technology has also been used to defraud businesses and steal their money, for example: in late 2019, a German energy company was defrauded of $220,000 after the deepfake was able to Voice imitation to create the voice of a high-profile executive character demanding immediate payment.
A spokesperson for the company's insurance company told the Washington Post, "The software was able to mimic the voice, not just the voice: tone, phonetic punctuation and stops, and the German accent." Not only was the audio perfectly recreated, but the phone call was matched along with a deep fake email mimicking the targeted CEO, adding another layer of legitimacy.
Some harmful picture alterations are produced - These are created using standard photo editing programs, such as Adobe Photoshop. There is a technique for identifying a very frequent Photoshop modification - bending pictures of human faces - by training a model solely on false images produced automatically by the Photoshop software. It has been shown that the model outperforms humans at identifying altered pictures, can determine the precise site of changes, and may in certain instances be used to "undo" the modifications and recreate the original, unedited image. He shows that the technique may be used effectively to manipulate the artist's actual picture [48].

3. Techniques for Detecting Manipulation

Several publications propose examining the internal GAN pipeline to identify artifacts associated with genuine and false pictures. For instance, the authors of [32] predict that the color of genuine camera pictures and false synthesis images is significantly different. They presented a detection method based on color characteristics and a "Linear Support Vector Machine" (SVM) for final classification, achieving a final AUC of 70.0 percent on the NIST MFC2018 dataset [12]. In this vein, [47] suggested another intriguing method. Wang, S., et al. It is hypothesized that monitoring neural behavior may also aid in identifying fake faces, since layer-by-layer neuronal activity patterns may pick up on more subtle and significant characteristics associated with face alteration.
System of detection. Their method, dubbed FakeSpotter, retrieved the neural covering behaviors of

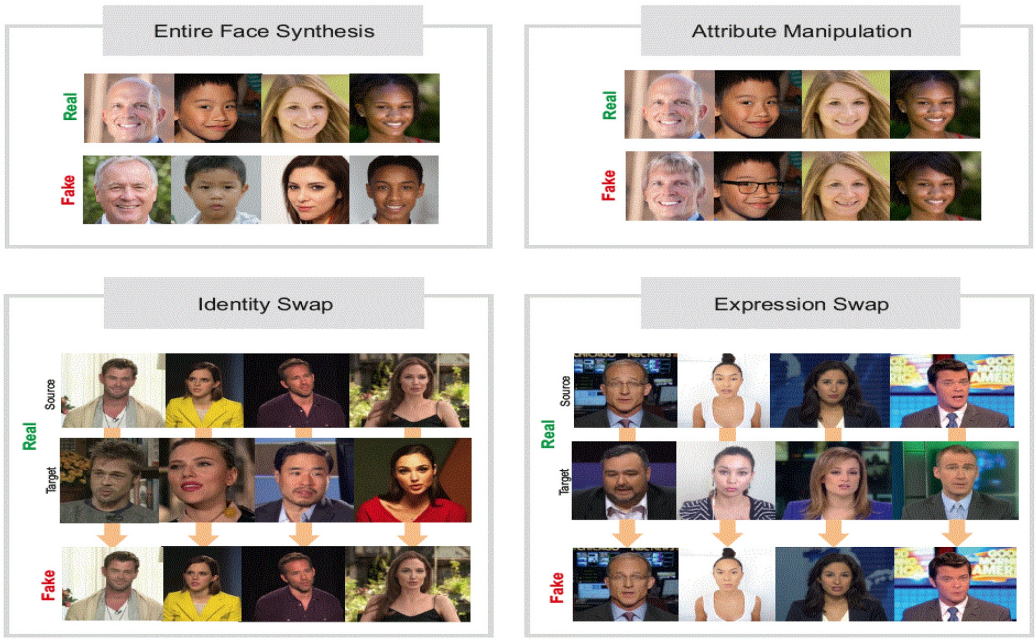

Figure 2: Techniques of facial modification

genuine and fake faces from deep facial recognition systems (e.g., VGG-Face [17], FaceNet [17] and OpenFace [38]), then trained an SVM for final classification. The authors evaluated their proposed techneque using actual faces from the datasets CelebA-HQ [19], FFHQ [20], and synthetic faces produced by InterFaceGAN [42] and styleGAN [4141], obtaining an accuracy of 84.7 percent for dummy identification while employing a model FaceNet. Recently, improved outcomes were reported in [13]. The authors developed a method for detecting counterfeits based on convolutional effects analysis. The expectation-maximization method was used to extract the characteristics [1]. For the final detection, common classifiers such as k-Nearest Neighbors (k-NN), support vector machines (SVM), and linear discriminant analysis (LDA) were employed. Their suggested method was validated using fictitious pictures produced by AttGAN [15], GDWCT [6], StarGAN [7], StyleGAN and StyleGAN2 [21], respectively.

The suggested convolutional neural network with a Y-shaped autoencoder was shown to be successful for both classification and segmentation tasks without the need of a sliding window, as classifiers often do. The exchange of information across the classification, segmentation, and re-construction activities enhanced the network's overall performance, particularly under the mismatch condition for observed attacks. Additionally, the autoencoder may rapidly respond to previously unknown assaults by fine-tuning with a few samples [36].

McCluskey and M. Albright [32] taught GANs to generate synthetic pictures that were almost indistinguishable from actual photos (in certain aspects). They examined the popular GAN application's generation network architecture and showed that the network's color processing is substantially different from that of the real camera in two ways. Additionally, we demonstrate how these two signals can be utilized to discriminate between GAN-generated and camera-generated pictures, showing successful discrimination between GAN images and genuine camera images used for GAN training.

One of the fastest developing phenomena is the well-known Deepfake: the ability to autonomously create and/or alter/swap a person's face in pictures and videos using Deep Learning algorithms. It is feasible to get great outcomes by developing new multi-timedia contents that are difficult for the human eye to distinguish as genuine or false. The term "Deepfake" therefore refers to any audiovisual

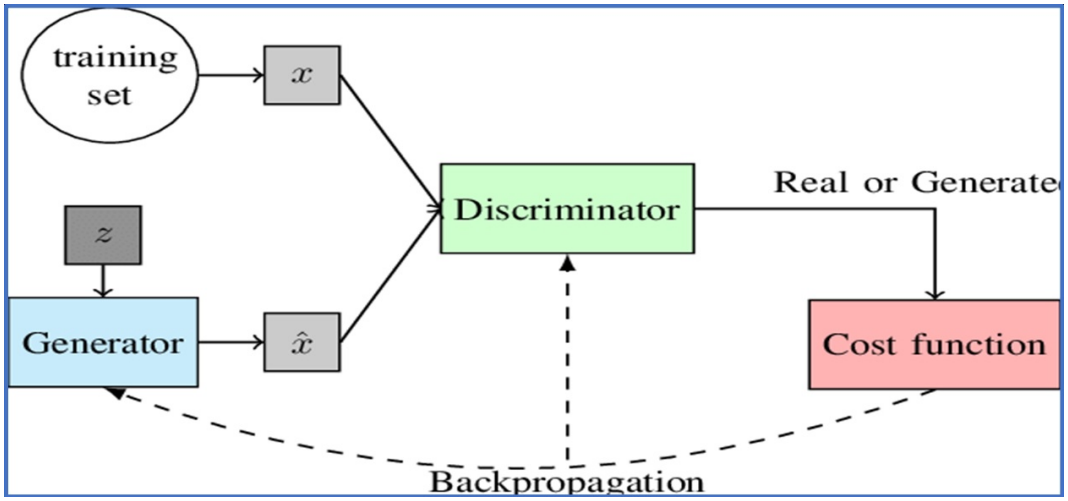

Figure 3: Generative adversarial network diagram

material that has been synthetically changed or produced using machine learning generative models [13].

Generative adversarial networks, also referred to as adversarial generative networks, are a kind of machine learning network developed in 2014 by Ian Goodfellow and colleagues. Two neural networks compete in a game to practice generating fictitious data that closely resembles real data and is difficult for a human or computer observer to distinguish; The mechanism of a generative adversarial network is shown here.

Numerous recent research examined the challenge of determining whether a face is real or digitally produced. Table 1 compares the most relevant methods in this field. We provide information on the technique used, the classifiers used, the highest performance, and the datasets used in each research. The best findings for each public database are shown in bold. It is essential to note that in certain instances, other evaluation scales are used, such as the area under the curve (AUC) or the equal error rate (EER), complicating comparisons across research.

4. Conclusion

Motivated by the continued success of digital face alterations, most notably DeepFakes, this study offers a thorough overview of the subject, including information on the following:

i) face modifications of many kinds

ii) methods of face manipulation,

iii) research databases accessible to the public, and

iv) standards

for detecting each face manipulation group, providing the most representative manipulation detection methods' main findings. By and large, the majority of contemporary facial modifications seem to.

It is very straightforward to discover false detectors in controlled settings, that is, when they are evaluated under the same conditions for which they were trained. This fact was shown in the majority of the benchmarks included in this research, with very low detection error rates. This scenario, however, may not be completely true, since fabricated images and films are often shared on

Table 1: Manipulation Detection Techniques

		COMPARISON OF THE DEEPFAKE DETECTION METHODS AND THE OPTIMUM RESULTS FOR EACH METHOD "EER = EQUAL ERROR RATE", "ACC. = ACCURACY", "TCR = TRUE CLASSIFICATION RATES" , "AUC = AREA UNDER THE CURVE", DLF = Deep- Learning-Features.			
Manipulation Detection Techniques	IDENTITY SWAP	Study	The method used	Dataset	results
		"Marcel and Korshunov "(2018)[23]	Audio-Visual Features	DeepfakeTIMIT (HQ) DeepfakeTIMIT (LQ)	EER = 8.90 % EER = 3.30 %
		"Delp and Güera" (2018) [14]	Temporal Features+ Image	Own	Acc. = 97.10 %
		"Zhou et al." (2018) [52]	Deep Learning Features + Steganalysis Features	FF++ / DFD DeepfakeTIMIT (HQ) DeepfakeTIMIT (LowQwality) Celeb-DF UADFV DFDC Preview	AUC = 70.10 % AUC = 73.50 % AUC = 83.50 % AUC = 53.80 % AUC = 85.10 % AUC = 61.40 %
		"Afchar et al." (2018)" [1]	Mesoscopic Features	Celeb-DF DFDC Preview FF++(RAW, FaceSwap) FF++(HQ, FaceSwap) FF++(LQ , FaceSwap) FF++(RAW ,DeepFake) FF++(HQ, DeepFake) FF++(LQ, DeepFake) Deepfake TIMIT (HQ) Deepfake TIMIT (LQ) UADFV Own	AUC = 54.80 % AUC = 75.30 % Acc. ' 96.00 % Acc. ' 93.00 % Acc. ' 83.00 % Acc. ' 98.00 % Acc. ' 94.00 % Acc. ' 90.00 % AUC = 68.40 % AUC = 87.80 % AUC = 84.30 % Acc. = 98.40 %
		"Li et al. (2018)" [25]	Face Warping Features	Celeb-DF DFDC Preview FF++ / DFD DeepfakeTIMIT DeepfakeTIMIT UADFV	AUC = 64.60 % AUC = 75.50 % AUC = 93.00 % AUC = 99.70 % AUC = 99.90 % AUC = 97.70 %
		"Matern et al." (2019) [31]	Visual Features	Deepfake TIMIT (LQ) DFD/ FF++ Deepfake TIMIT (HQ) DFDC Preview Celeb-DF Own UADFV	AUC = 77.30 % AUC = 66.20 % AUC = 78.00 % AUC = 55.10 % AUC = 85.10 % AUC = 70.20 % AUC = 77.00 %
		"Yang et al." (2019) [49]	Head Pose Features	Celeb-DF DFDC Preview FF++ / DFD DeepfakeTIMIT (HQ) DeepfakeTIMIT (LQ) UADFV	AUC = 54.60 % AUC = 55.90 % AUC = 47.30 % AUC = 53.20 % AUC = 55.10 % AUC = 89.00 %
		"Sabir et al." (2019) [41]	Temporal Features +Image	FF++(LQ , FaceSwap) FF++(LQ, DeepFake)	AUC = 96.30 % AUC = 96.90 %
		"Rossler et al." (2019) [40]	Steganalysis Features Mesoscopic Features Deep Learning Features	FF++(RAW, FaceSwap) FF++(HQ ,FaceSwap) FF++(LQ ,FaceSwap,) FF++(RAW, DeepFake) FF++(HQ, DeepFake) FF++(LQ, DeepFake)	Acc= 99.00 % Acc. =97.00 % Acc. = 93.00 % Acc.= 100.00 % Acc.= 98.00 % Acc. = 94.00 %
		"Nguyen et al." (2019) [36]	DLF	Celeb-DF DFDC Preview FF++(HQ, FaceSwap) DFD / FF++ DeepfakeTIMIT (HQ) DeepfakeTIMIT (LQ) UADFV	AUC = 54.30 % AUC = 53.60 % EER = 15.10 % AUC = 76.30 % AUC = 55.30 % AUC = 62.20 % AUC = 65.80 %
		"Nguyen et al." (2019) [36]	DLF	Celeb-DF DFDC Preview DFD/ FF++ DeepfakeTIMIT (HQ) DeepfakeTIMIT (LQ) UADFV	AUC = 57.50 % AUC = 53.30 % AUC = 96.60 % AUC = 74.40 % AUC = 78.40 % AUC = 61.30 %
		"Dang et al." (2019) [8]	DLF	DFFD	EER = 3.10 % AUC = 99.40 %
		"Dolhansky et al." (2019) [9]	DLF	DFDC Preview	Recall = 8.40 % Precision = 93.00 %
		"Wang and Dantcheva" (2020) [40]		FF++(LQ,FaceSwap) FF++ (LQ,DeepFake)	TCR = 92.25 % TCR = 95.13 %
		"Jung et al." (2020) [18]	Eye Blinking	Own	Acc. = 87.50 %
		"Tolosana et al." (2020) [46]	Facial Regions Features	Celeb-DF DFDC Preview FF++(HQ,FaceSwap)	AUC = 83.60 % AUC = 91. 00 % AUC = 99.40 %

				UADFV	AUC = 100.00 %
	ATTRIBUTE MANIPULATION	"Bharati et al." (2016) [3]	DLF	(ND-IIITD Retouching ,Celebrity Retouching) Own	Acc. = 87.10 % Acc. = 96.20 %
		"Tariq et al." (2018) [43]	DLF	Own (ProGAN,Adobe Photoshop)	AUC = 99.90 % AUC = 74.90 %
		"Wang et al." (2019) [48]	DLF	Own(Adobe Photoshop)	AP = 99.80 %
		"Zhang et al." "(2019) [51]	Spectrum Domain Features	Own (StarGAN/CycleGAN)	Acc. = 100.0 %
		"Jain et al. "(2019) [17]	DLF	Own (ND-IIITD Retouching,StarGAN)	Overall Acc. = 99.60 % Overall Acc. = 99.70 %
		"Wang et al. "(2019) [48]	GAN-Pipeline Features	Own(InterFaceGAN/StyleGAN)	Acc. = 84.70 %
		"Dang et al." (2019) [8]	DLF	DFFD (FaceApp/StarGAN)	AUC = 99.90 % EER = 1.00 %
		"Nataraj et al." (2019) [33]	Steganalysis Features	Own(StarGAN/CycleGAN)	Acc. = 99.40 %
		"Marra et al." (2019) [30]	DLF	Own (Glow/StarGAN)	Acc. = 99.30 %
		"Rathgeb et al." (2020) [39]	PRNU Features	Own(5 Public Apps)	EER = 13.70 %
	EXPRESSION SWAP	"Afchar et al." (2018) [1]	Mesoscopic Features	FF++(RAW, NeuralTextures) FF++(HQ ,NeuralTextures) FF++(LQ, NeuralTextures) FF++(RAW, Face2Face) FF++(HQ, Face2Face) FF++ (Face2Face, LQ)	Acc. = 95.00 % Acc. =85.00% Acc. =75.00% Acc. = 96.80 % Acc. = 93.40 % Acc. = 83.20 %
		"Matern et al." (2019) [31]	Visual Features	FF++(RAW, Face2Face)	AUC = 86.60 %
		"Amerini et al." (2019) [2]	Image + Temporal Features	FF++(Face2Face, -)	Acc. = 81.60 %
		"Rössler et al." (2019) [40]	Steganalysis Features Mesoscopic Features Deep_Learning_ Features	FF++ (RAW, NeuralTextures) FF++ (HQ, NeuralTextures) FF++ (LQ, NeuralTextures) FF++ (RAW, Face2Face) FF++ (HQ, Face2Face) FF++(LQ ,Face2Face)	Acc. =99.00 % Acc. =93.00 % Acc. =81.00 % Acc. =100.00 % Acc. = 98.00 % Acc. =91.00 %
		"Nguyen et al. "(2019) [36]	DLF	FF++ (HQ ,Face2Face) FF++(HQ, NeuralTextures)	EER = 7.10 % EER = 7.80 %
		"Sabir et al." (2019) [41]	Image + Temporal Features	FF++(Face2Face, LQ)	Acc. = 94.30 %
		"Dang et al." (2020) [8]	DLF	FF++(Face2Face, -)	AUC = 99.4% EER = 3.4%
		"Wang and Dantcheva" (2020) [40]	DLF	FF++(LQ, Face2Face) FF++(LQ, NeuralTextures)	TCR = 90.27% TCR = 80.5%
	ENTIRE FACE SYNTHESIS	"McCloskey and Albright" (2018) [32]	GAN-Pipeline-Features	NIST MFC2018	AUC = 70.0%
		"Marra et al." (2019) [30]	DLF	(ProGAN,Glow ,CycleGAN, StyleGAN, StarGAN) Own	Acc. = 99.30 %
		"Wang et al." (2019) [48]	GAN-Pipeline-Features	Own (InterFaceGAN, StyleGAN)	Acc. = 84.70 %
		"Yu et al." (2019) [50]	DLF	Own(ProGAN, SNGAN,CramerGAN, MMDGAN)	Acc. = 99.50 %
		"Nataraj et al." (2019) [33]	Steganalysis-Features	100K-Faces (StyleGAN)	EER = 12.30 %
		"Guarnera et al." (2020) [13]	GAN-Pipeline-Features	(GDWCT ,StarGAN, AttGAN, StyleGAN2, StyleGAN) Own	Acc. = 99.81 %
		"Neves et al. (2020) [34]	DLF	iFakeFaceDB	EER = 0.30 % EER = 4.50 %
		"Dang et al. "(2020) [8]	DLF	DFFD (ProGAN, StyleGAN)	AUC = 100.00 % EER = 0.1 %
		"Hulzebosch et al." (2020) [16]	DLF	Own(StarGAN, Glow,ProGAN, StyleGAN)	Acc. = 99.80 %

social media sites, sometimes with substantial modifications such as like compression ratio, resizing, and noise, among others. Additionally, face modification methods are always evolving. These issues need further study into the false detectors' capacity to generalize to previously unknown situations. This feature has been explored in depth in a variety of publications [34]. Future.

The study may be in line with recent studies [40, 31], since they do not need the use of fictitious films for training, allowing for a greater capacity to generalize to unseen assaults.

The proliferation of misinformation in internet material necessitates the creation of a system for identifying it. Face manipulation in videos is only one facet of a much bigger issue. We demonstrated in this study that combining a recurrent-convolutional model with a face alignment method outperforms the state-of-the-art. Additionally, we investigated several strategies for aligning and merging CNN features through recurrence. We discovered that a landmark-based face alignment combined with bidirectional-recurrent-denset performed the best for detecting face manipulation in videos [41].

References

[1] D. Afchar, V. Nozick, J. Yamagishi and I. Echizen, *MesoNet: A compact facial video forgery detection network*, 10th IEEE Int. Work Inf. Forensics Secur. WIFS 2018, (2019).

[2] I. Amerini, L. Galteri, R. Caldelli and A. Del Bimbo, *Deepfake video detection through optical flow based CNN*, Proc. IEEE/CVF Int. Conf. on Comput. Vis. (ICCV) Workshops, 2019.

[3] A. Bharati, R. Singh, M. Vatsa and K. Bowyer, *Detecting facial retouching using supervised deep learning*, IEEE Transactions on Inf. Forens. Secur. 11(9) (2016) 1903–1913.

[4] BBC Bitesize, *Deepfakes: What Are They and Why Would I Make One?*, www.bbc.co.uk, 2019.

[5] C. Bregler, M. Covell and M. Slaney, *Video rewrite: driving visual speech with audio*, Proc. 24th Annu. Conf. Comput. Graph Interact Tech SIGGRAPH 1997 (1997) 353-360.

[6] W. Cho, S. Choi, D.K. Park, I. Shin and J. Choo, *Image-to-image translation via group-wise deep whitening-and-coloring transformation*, Proceedings-2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR 2019, IEEE Computer Society, 2019-June (2019) 10639–10647.

[7] Y. Choi, M. Choi, M. Kim, J.W. Ha, S. Kim and J. Choo, *StarGAN: unified generative adversarial networks for multi-domain image-to-image translation*, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2018) 8789–8797.

[8] H. Dang, F. Liu, J. Stehouwer, X. Liu and A.K. Jain, *On the detection of digital face manipulation*, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2020) 5780–5789.

[9] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang and C.C. Ferrer, *The deepfake detection challenge (dfdc) dataset*, arXiv preprint arXiv:2006.07397, 2020.

[10] H. Farid, *Image forgery detection*, IEEE Signal Processing Magazine 26(2) (2009) 16–25.

[11] J. Galbally, S. Marcel and J. Fierrez, *Biometric antispoofing methods: A survey in face recognition*, IEEE Access 2 (2014) 1530–1552.

[12] H. Guan, M. Kozak, E. Robertson, Y. Lee, A. Yates, A. Delgado, D. Zhou, T. Kheyrkhah, J. Smith and J. Fiscus, *MFC datasets: large-scale benchmark datasets for media forensic challenge evaluation*, in Proc. IEEE Winter Applications of Computer Vision Workshops, 2019.

[13] L. Guarnera, O. Giudice and S. Battiato, *DeepFake detection by analyzing convolutional traces*, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work, 2020 (2020) 2841–2850.

[14] D. Guera and E.J. Delp, *Deepfake video detection using recurrent neural networks*, Proc. AVSS 2018 - 2018 15th IEEE Int. Conf. Adv. Video Signal-Based Surveill, 2019.

[15] Z. He, W. Zuo, M. Kan, S. Shan and X. Chen, *AttGAN: Facial attribute editing by only changing what you want*, IEEE Trans. Image Process, 28(11) (2019) 5464–5478.

[16] N. Hulzebosch, S. Ibrahimi and M. Worring, *Detecting CNN-generated facial images in real-world scenarios*, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit Work, 2020 (2020) 2729–2738.

[17] A. Jain, R. Singh and M. Vatsa, *On detecting GANs and retouching based synthetic alterations*, 2018 IEEE 9th Int. Conf. Biometrics Theory, Appl. Syst. BTAS 2018 (2018).

[18] T. Jung, S. Kim and K. Kim, *Deepvision: deepfakes detection using human eye blinking pattern*, IEEE Access. 8 (2020) 83144–83154.

[19] T. Karras, T. Aila, S. Laine and J. Lehtinen, *Progressive growing of GANs for improved quality, stability, and variation*, 6th Int. Conf. Learn Represent ICLR 2018 – Conf. Track Proc. (2018) 1–26.

[20] T. Karras, S. Laine and T. Aila, *A style-based generator architecture for generative adversarial networks*, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019 (2019) 4396–4405.
[21] T. Karras, S. Laine, M. Aittala, J. Hellsten, J. Lehtinen and T. Aila, *Analyzing and improving the image quality of stylegan*, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2020 (2020) 8107–8116.
[22] D.P. Kingma and M. Welling, *Auto-encoding variational bayes*, 2nd Int. Conf. Learn Represent ICLR 2014 – Conf. Track Proc. 2014 (2014) 1–14.
[23] P. Korshunov and S. Marcel, *Deepfakes: A new threat to face recognition? Assessment and detection*, arXiv, 2018.
[24] P. Korus, *Digital image integrity – a survey of protection and verification techniques*, Digit. Signal Process, Rev. J. 71 (2017) 1–26.
[25] Y. Li and S. Lyu, *Exposing deepfake videos by detecting face warping artifacts*, arXiv, 2018.
[26] C. Li, K. Xu, J. Zhu and B. Zhang, *Triple generative adversarial nets*, Adv. Neural Inf. Process Syst. 2017 (2017) 4089–4099.
[27] W.S. Lin, S.K. Tjoa, H.V. Zhao and K.J.R. Liu, *Digital image source coder forensics via intrinsic fingerprints*, IEEE Trans. Inf. Forensics Secur. 4(3) (2009) 460–475.
[28] M. Liu, Y. Ding, M. Xia, X. Liu, E. Ding, W. Zuo and S. Wen, *STGAN: A unified selective transfer network for arbitrary image attribute editing*, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. 2019 (2019) 3668–3677.
[29] S. Marcel, M. Nixon, J. Fierrez and N. Evans, *Handbook of Biometric Anti-Spoofing (2nd Edition)*, Springer, 2019.
[30] F. Marra, C. Saltori, G. Boato and L. Verdoliva, *Incremental learning for the detection and classification of GAN-generated images*, 2019 IEEE Int. Work Inf. Forensics Secur. WIFS 2019 (2019).
[31] F. Matern, C. Riess and M. Stamminger, *Exploiting visual artifacts to expose deepfakes and face manipulations*, in Proc. IEEE Winter Applications of Computer Vision Workshops, 2019.
[32] S. McCloskey and M. Albright, *Detecting GAN-generated imagery using color cues*, arXiv, 2018.
[33] L. Nataraj, T.M. Mohammed, B.S. Manjunath, S. Chandrasekaran, A. Flenner, J.H. Bappy and A.K. Roy-Chowdhury, *Detecting GAN generated fake images using co-occurrence matrices*, IS& T Int. Symp. Electron Imaging Sci. Technol. 2019(5) (2019) 1–7.
[34] J.C. Neves, R. Tolosana, R. Vera-Rodriguez, V. Lopes, H. Proença and J. Fierrez, *GANprintR: Improved fakes and evaluation of the state of the art in face manipulation detection*, IEEE J. Sel. Top Signal Process 14(5) (2020) 1038–1048.
[35] H.M. Nguyen and R. Derakhshani, *Eyebrow recognition for identifying deepfake videos*, BIOSIG 2020 – Proc. 19th Int. Conf. Biometrics Spec. Interes. Gr. (2020) 1–6.
[36] H.H. Nguyen, F. Fang, J. Yamagishi and I. Echizen, *Multi-task learning for detecting and segmenting manipulated facial images and videos*, 2019 IEEE 10th Int. Conf. Biometrics Theory, Appl. Syst. BTAS 2019 (2019).
[37] H.H. Nguyen, J. Yamagishi and I. Echizen, *Use of a capsule network to detect fake images and videos*, arXiv, 2019.
[38] O.M. Parkhi, A. Vedaldi and A. Zisserman, *Deep face recognition*, Visual Geometry Group Department of Engineering Science University of Oxford, (2015) 1–12.
[39] C. Rathgeb, A. Botaljov, F. Stockhardt, S. Isadskiy, L. Debiasi, A. Uhl and C. Busch, *PRNU-based detection of facial retouching*, IET Biomet. 9(4) (2020) 154–164.
[40] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies and M. Niessner, *FaceForensics++: Learning to detect manipulated facial images*, Proc. IEEE Int. Conf. Comput. Vis. 2019 (2019) 1–11.
[41] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi and P. Natarajan, *Recurrent convolutional strategies for face manipulation detection in videos*, arXiv, 2019.
[42] F. Schroff, D. Kalenichenko and J. Philbin, *FaceNet: A unified embedding for face recognition and clustering*, Proc. IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. (2015) 815–823.
[43] S. Tariq, S. Lee, H. Kim, Y. Shin and S. Woo, *Detecting both machine and human created fake face images in the wild*, Proc. Int. Workshop Multimedia Priv. Secur.(2018) 81–87.
[44] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt and M. Nießner, *Face2face*, Commun. ACM. 62(1) (2018) 96–104.
[45] J. Thies, M. Zollhöfer, M. Stamminger, C. Theobalt and M. Nießner, *Deferred neural rendering: image synthesis using neural textures*, ACM. Trans. Graph. 38(4) (2019).
[46] R. Tolosana, S. Romero-Tapiador, J. Fierrez and R. Vera-Rodriguez, *Deepfakes evolution: analysis of facial regions and fake detection performance*, Lect. Notes Comput. Sci. (including Subser Lect Notes Artif Intell Lect Notes Bioinformatics), (2021) 12665.

[47] R. Wang, F. Juefei-Xu, L. Ma, X. Xie, Y. Huang, J. Wang and Y. Liu, *Fakespotter: a simple yet robust baseline for spotting ai-synthesized fake faces*, IJCAI Int. J. T. Conf. Artif. Intell. 2020 (2021) 3444–3451.
[48] S.Y. Wang, O. Wang, R. Zhang, A. Owens and A. Efros, *Detecting photoshopped faces by scripting photoshop*, Proc. IEEE Int. Conf. Comput. Vis. 2019 (2019) 10071–10080.
[49] X. Yang, Y. Li and S. Lyu, *Exposing deep fakes using inconsistent head poses*, ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process-Proc. 2019 (2019) 8261–8265.
[50] N. Yu, L. Davis and M. Fritz, *Attributing fake images to GANs: learning and analyzing GAN fingerprints*, Proc. IEEE Int. Conf. Comput. Vis. 2019 (2019) 7555–7565.
[51] X. Zhang, S. Karaman and S.F. Chang, *Detecting and simulating artifacts in GAN fake images*, 2019 IEEE Int. Work Inf. Forensics Secur. WIFS 2019 (2019).
[52] P. Zhou, X. Han, V.I. Morariu and L.S. Davis, *Two-stream neural networks for tampered face detection*, IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit Work 2017 (2017) 1831–1839.