Multimedia Tools and Applications
https://doi.org/10.1007/s11042-021-11733-y

Deepfake generation and detection, a survey

Tao Zhang[1,2,3,4]

Received: 27 August 2020 / Revised: 2 September 2021 / Accepted: 5 November 2021

Abstract
Deepfake refers to realistic, but fake images, sounds, and videos generated by articial intelligence methods. Recent advances in deepfake generation make deepfake more realistic and easier to make. Deepfake has been a signicant threat to national security, democracy, society, and our privacy, which calls for deepfake detection methods to combat potential threats. In the paper, we make a survey on state-ofthe-art deepfake generation methods, detection methods, and existing datasets. Current deepfake generation methods can be classified into face swapping and facial reenactment. Deepfake detection methods are mainly based features and machine learning methods. There are still some challenges for deepfake detection, such as progress on deepfake generation, lack of high quality datasets and benchmark. Future trends on deepfake detection can be efficient, robust and systematical detection methods and high quality datasets.

Keywords Deepfake · Detection · Generation · Survey · Media forensics

1 Introduction

In December 2017, a Reddit user named *Deepfakes* [20] posted fake pornographic videos generated by open-source artificial intelligence (AI) tools that can swap faces in images and videos. Since then, the term *Deepfake* or *Deepfakes* is used to describe the recreation of a human's appearance, face in particular, through AI methods. *Malicious Deep Fake Prohibition Act of 2018* [1] defines *deepfake* as realistic but fake or altered videos and audios which are hard to identify. *DEEP FAKES Accountability Act* [2] defines *deepfake* as videos, audios, and images that can authentically depict one but the person who did not

✉ Tao Zhang
tao.zhang.cn@outlook.com

1 School of Cyber Science and Technology, Beihang University, Beijing, China

2 Guangxi Key Laboratory of Cryptography and Information Security, Guilin University of Electronic Technology, Guilin, China

3 Guangxi Key Laboratory of Trusted Software & Guangxi Key Laboratory of Cryptography and Information Security, Guilin University of Electronic Technology, Guilin, China

4 Key Lab of Film and TV Media Technology of Zhejiang Province, Hangzhou, China

Published online: 08 January 2022

engage in. In summary, *deepfake* refers to seemingly realistic but fake images, audios, videos, and other digital medias produced by AI methods, deep learning in particular.

In September 2019, a Chinese viral face application, ZAO [92], hit the App Store worldwide and climbed to the top lists of Chinese iOS App Store. ZAO is such a deepfake application that can swap faces, change hair color and style, do light and dark makeup, swap genders in specific pictures, and change age with celebrities in a wide selection of different video clips. The face-swapping process is automatic, and what users need to do is only uploading images to ZAO. Before the emergence of ZAO, there are also some similar applications, such as Deepfakes (2017) [22], FaceApp (2019) [27], and DeepFaceLab (2019) [19]. FaceApp is a Russian smartphone application that can generate fake images that appear older than they are. Deepfakes is a software that can help to replace human face with any other person or animal's face. DeepFaceLab is a windows program that enables users to replace faces in videos with machine learning and human image synthesis. Advances in AI, especially deep neural network (DNN) and Generated adversarial network (GAN), make deepfake images and videos much easier, cheaper and simpler to make. With applications like ZAO, FaceApp, Deepfakes, and DeepFaceLab, anyone with a smartphone or a computer can generate highly realistic fake videos and images in a few hours, changing their hairstyle, gender, age, and other face attributes.

These realistic deepfake medias have been a significant threat to privacy, democracy, national security, and society [5, 13]. Realistic deepfake images and videos can be used to bypass facial authentication, create political distress, fake news, and blackmail someone. The spreading of information with fake videos and images may threaten and undermine our trust on online digital content. The Democratic National Committee of the U.S. even sent a security alert during the 2020 presidential campaigns warning not to use the FaceApp developed by Russia for potential democracy threats. Besides, ZAO's privacy policy has caused concerns about identity theft and privacy concerns. According to Deeptrace [5], the number of deepfake videos is rapidly expanding and there are more than 14,678 deepfake videos online.

What is worse, it is hard for humans to identify these realistic deepfake images, audios, and videos. Therefore, distinguishing deepfakes medias from real ones is important, necessary and urgent. In the paper, we briefly introduce the concepts and types of deepfake, then make a survey on state-of-the-art methods deepfake generation and detection. Finally, discuss challenges and future research directions on deepfake detection.

The paper is organized as follows. Section 2 introduces background information on machine learning and deep learning methods. Section 3 and section 4 presents deepfake generation, detection methods, and existing datasets. Discussions on challenges and future directions of deepfake detection are presented in section 5, and section 6 concludes the paper.

2 Background

SVM. The support vector machine (SVM) [38] is a calssical supervised machine learning algorithm that uses separating hyperplanes to classify data instead of straight lines. Compared with traditional classification methods, SVM is with better result and widely used in different classification tasks.

Deep learning. As a branch of machine learning, deep learning [56] is comprised of multiple processing layers, which can learn representations of data with multiple levels

of abstraction. Deep learning has been widely used in pattern recognition and computer vision tasks. Most deep learning methods are largely dependent on labeled data to achieve better performance. However, massive, labeled training data costs a lot, and the training process is also time-consuming.

RNN. Recurrent neural network (RNN) [34] could be regarded as multiple replicates of a neural network connected in a sequence. RNN can memorize long sequences and has the advantage of solving temporal data problems. RNN is ideal for sequential data analysis and widely used in natural language processing tasks, including speech and video data.

LSTM. Long short-term memory (LSTM) [39] network is an improvement of RNN. LSTM is composed of LSTM units, which are composed of cells with input, output, and forget gates. LSTM is mostly used for sequential data analysis, such as speech recognition, language translation, time-series data forecasting. Compared with RNN, LSTM can deal with longer-term dependencies, while RNN can only handle short sequences.

CNN. Convolutional Neural Network (CNN) is among the most popular deep neural networks and composed of convolutional layers, pooling layer, and fully connected layers. CNN is more suitable for classification, image recognition, and computer vision tasks. AlexNet, VGGNet, GoogLeNet, and ResNet are some classic CNN models.

GAN. In 2014, GAN [33] is proposed as a deep learning model composed of a generative and discriminative model. The generative model can generate data randomly, while the discriminative model evaluates if generated data is from training datasets. The competition between generative model and discriminative model can help GAN achieve better performance. GAN is widely used in different types of tasks, such as image classification, image and text generation.

Transfer learning. In real-world scenarios, people can apply knowledge learned from one task to other related tasks. To overcome challenges of isolated learning tasks, transfer learning [73] is proposed to transfer knowledge from one task to another [67], which can solve new problems faster and give better solutions.

3 Deepfake generation

3.1 Types of Deepfake

Deepfake media can be categorized into four types according to media types, as video with audio, video without audio, audio, and image. WaveNet [72] is an example of a fake audio generation method, which can generate speech that mimics any human voice with fully CNN. Speeches generated with WaveNets sound more natural than the best existing text-to-speech systems. Current deepfake researches and applications mainly focus on human face image and short video generation, which can generate highly realistic fake face videos and images with changed face hairstyle, gender, age, and other attributes. In the paper, we focus on deepfake face generation and detection.

3.2 Face-based generation methods

Current deepfake generation methods can be categorized into two types as face swapping and facial reenactment. Table 1 is a summary of deepfake generation methods.

Table 1 A summary of deepfake generation methods

Methods	Categories	Features used	Techniques	Videos/images
Deepfakes [22]	Face swapping	Face-ROI	GAN	Videos
Fast Face-Swap [52]	Face swapping	Original pose, facial expression, light	CNN	Images
RSGAN [67]	Face swapping	Face and hair	GAN	Images
Face2Face [85]	Face swapping	3D facial expressions	GAN	Videos
HeadOn [83]	Face reenactment	Facial expressions, eye gaze, rigid head pose, and motions of the upper body	–	Videos
Deep Video Portraits [51]	Face reenactment	Illumination, identity, pose, expression, eyes	Neural network	Videos
NeuralTextures [84]	Face reenactment	–		Images
StarGAN [16]	Face reenactment	Face attribute and facial expression	GAN	Images
Geng et al. [31]	Face reenactment	3D face fitting	Deep generative network	Images
GANimation [76]	Face reenactment	–	GAN	Images
FSGAN [71]	Face swapping	–	GAN	Images
Chan et al. [10]	Facial reenactment	Pose	GAN, RNN	Videos

3.2.1 Face swapping

Face swapping methods can replace the face in the reference image with the same facial shape and features as the input face [12]. Deep learning methods are used to extract facial features from the input face and then transfer to the generated face. Face in a video face may be replaced by another person, while the remaining original scene content and the original facial expression are preserved. Both Deepfakes and the ZAO mobile application are face-swapping deepfake techniques. Deep neural networks, especially GAN, have shown great power in generating deepfake face images and videos on quality and efficiency. Chen et al. [52] use CNN to capture the target's appearance features from a set of his/her photographs. Then frame the face-swapping problems in terms of style transfer. Chen et al. [12] propose a face swapping algorithm which can swap the face in the reference image with the same facial shape and features from the input image. Natsume et al. [67] present RSGAN as a deep neural network based method which can generate and edit face images through face swapping, attribute-based editing, and face synthesis. Besides, RSGAN can handle face and hair appearance independently. Nirkin et al. [71] propose Face Swapping GAN (FSGAN) for face swapping and reenactment which does not rely on training faces. In FSGAN, a RNN based approach is adopted for face reenactment, which can control the expressions of the face appearing in the source image.

3.2.2 Facial reenactment

Another research branch of deepfake face generation is facial reenactment, which can manipulate certain facial attributes or reenact faces with deep learning methods. Thies et al. [85] present Face2Face as a real-time facial reenactment system that can generate realistic videos with facial expressions from the source actor. Face2Face can construct a 3D face model with only an imput image, and use corresponding 3D geometry to render the generated fake face. HeadOn [83] is a real-time interactive reenactment system for human portrait videos. With a controllable 3D actor model, HeadOn can capture facial expressions, eye gaze, rigid head pose, and motions from a source actor and transfer them to a target actor in real-time. Kim et al. [51] propose a face reenactment approach that enables video re-animation using only an input video. 3D head model is also used in the method and 3D head position, head rotation, face expression, eye gaze, and eye blinking are captured first, then transfer to a target actor. Combined with traditional graphics pipeline and learnable components, NeuralTextures [84] can generate synthetic face images in real-time.

GANs are widely used in deepfake facial reenactment methods. Isola et al. [43] show that GAN is effective at generating high-quality and large-resolution synthetic images. StarGAN [16] can transfer facial attributes from one image to another and synthesize facial expression synthesis. Geng et al. [31] present a fine-grained face manipulation method that can synthesize faces with arbitrary expressions. In the method, a conditional generative network is used to change the appearance, a fully connected network is used to predict the accurate shapes, and the available depth data is for supervision. GANimation [76] can animate a given image and render novel expressions with a GAN conditioning scheme based on Action Units (A.U.) annotations, which can also change both background and lighting conditions. StyleGAN [46] is a style-based generator architecture with GAN that can gradually generate an artificial image, starting from a very low resolution and continuing to a high resolution. By modifying input of each level of GAN, all the pose, face shape, hair

color, and other features of the generated synthesized images can be controlled. PGGAN [45] is also GAN based and can generate high-quality images (1024×1024 pixels) with some layers to fine details during the training process.

These deepfake face generation methods can achieve both image-to-image and video-to-video face swapping and facial reenactment, which can transfer face, corresponding facial attributes, and expressions to the target face. Besides, there are also some deepfake generation methods, which can generate realistic but non-existing faces, such as [10] and [93]. Chan et al. [10] present a video-to-video face generation method which can transfer performance from one target to another in only a few minutes. Zakharov et al. [93] propose an approach that can generate realistic talking head models with a video dataset.

4 Deepfake detection

4.1 Detection methods

Realistic deepfake videos and images have potential threats to national security, democracy, and privacy. Therefore, it is urgent and necessary to combat malicious deepfake medias, such as developing detection methods. State-of-the-art deepfake detection methods can be classified into features based methods and machine learning based methods. There are some features in deepfake medias that can be used to identify deepfake medias. And machine learning methods, deep learning in particular, are widely used to extract these features and distinguish deepfake medias. Table 2 is a summary of state of the art deepfake detection methods.

4.1.1 Features based detection methods

From the nature of how deepfake images and videos are generated, deepfake medias have some unique features that can be distinguished from real images and videos. These features can be summarized as biometric features, model features, and media features.

Biometric features. In videos and images, there are some human biometric features that can be used to distinguish deepfake medias, such as eye blinking, lip-sync, facial and head movements, head pose, color, texture, shape cues. Li et al. [58] find that eye blinking in deepfake videos is different from humans, then an eye blinking detection methods is proposed to detect deepfake videos. DeepVision [44] also uses eye blinking patterns as features to detect GAN generated deepfake videos, and eye blinking count, blink elapsed time, blink period time are measured as eye blinking patterns. The accuracy rate of DeepVision is 0.875. In [69], the eyebrow is used as features, and four deep learning methods, LightCNN, Resnet, DenseNet, and SqueezeNet, are applied to detect deepfake videos. The highest AUC (Area Under Curve) on UADFV and Celeb-DF datasets can reach 0.984 and 0.712. In FakeET [37], eye movement and EEG signal features are combined to distinguish deepfake videos.

Yang et al. [90] present a 3D head pose estimation based method to detect deepfake videos. Wang et al. [87] propose a Face-Aware Liquify (FAL) tool to detect deepfake face. In FAL, the width of the nose, eye distance, chin height, and other 16 parameters are used. In [62], global consistency, illumination estimation, and geometry estimation are used to detect deepfake faces in images. In [86], the consistent facial geometry across videos is

Table 2 A summary of the state of the art deepfake detection methods

Method	Features	Machine learning method	Dealing with	Datasets	Accuracy
Li et al. [58]	Eye blinking	DNN (LRCN)	Images	Deepfake	0.99
Nguyen and Derakhshani [69]	Eyebrow	CNN	Videos	UADFV, Celeb-DF	UADFV(0.984), Celeb-DF(0.712)
Yang et al. [90]	3D head pose estimation	SVM	Videos, images		0.89
Zhou et al. [96]	Tampering artifacts	Two-Stream Neural Networks	Images	2010 tampered images	0.99
Matern et al. [62]	Face global consistency, Illumination Estimation, Geometry Estimation	–	Images, videos	Dataset of generated faces	0.866
Agarwal et al. [4]	Facial and head movements	SVM	Videos	From youtube	0.99
DeepFD [40]	–	Deep neural network	Images	LSGAN, DCGAN, WGAN, WGAN-GP, PGGAN	0.947
Sabir et al. [82]	Temporal information	RCN	Videos	FaceForensics++	Baseline+0.455
Li and Lyu [61]	Face Warping Artifacts	CNN	Videos	DeepfakeTIMIT, UADFV	0.994
Gera and Delp [32]	Temporal awareness inconsistency	CNN+LSTM	Videos	600 videos collected online	0.97
VTD-Net [89]	Inconsistency across the temporal domain	CNN	Videos	Celeb-DF	0.9189
Zhao et al. [95]	–	CNN	Videos	Faceforensics++	0.9993
SDHF [57]	Frame level, clip-level, and video-level features	CNN	Videos	UADFV, DFDC, Celeb-DF, FaceForensics	0.982 (FaceForensics)
XceptionNet* [15]	Visual frames, edge maps, and dense optical flow maps	CNN	Videos	FaceForensics++	1
OC-FakeDect [47]	–	VAE	Videos, images	FaceForensics++	0.975
CGFace [18]	–	CNN	Images	PCGAN, BEGAN	0.98
DeepfakeStack [78]	–	CNN	Videos	FaceForensics++	0.9965
Nataraj [66]	–	CNN	Images	Images generated using CycleGAN, StarGAN	0.99

Table 2 (continued)

Method	Features	Machine learning method	Dealing with	Datasets	Accuracy
Mo et al. [64]	–	CNN	Images	PGGAN	0.994
Xie et al. [88]	–	CNN	Videos, images	UADFV, FaceForensics++, Celeb-DF dataset	UADFV(0.9873), FaceForensics++(0.9132), Celeb-DF(0.9885)
NA-VGG [11]	Image noise and image augmentation	VGG/CNN	Videos	Celeb-DF	0.857
ADDNet [99]	Attention mask	CNN	Images, videos	FaceForensics++	0.9982 (2D), 0.9830 (3D)
MesoNet [3]	–	Deep neural network	Videos	Deepfake, Face2Face	0.98(Deepfake), 0.95(Face2Face)
Ding et al. [23]	–	Transfer learning	Images	Collected online	0.96
Farid [28]	–	CNN/DNN	Images	103250 unique paintings	0.9

used to detect deepfake videos. When one speaks, there should be distinct facial expressions and movements, and Agarwal et al. [4] use distinct and consistent facial expressions to detect deepfake videos. In [54], lip-syncing and dubbing are used as features to distinguish deepfake medias. In [6], inconsistencies among adjacent frames have been used as features to distinguish manipulated videos. In [94], differences in facial expressions between consecutive frames in a video are used to detect deepfake videos. And accuracy of detection on FaceForensics++ is 0.981. In [96], low-level inconsistencies between image patches are used as features to detect tampered faces.

Montserra et al. [65] propose a visual and temporal features based method to detect deepfake videos. To extract these features, both CNN and RNN are used. Accuracy of this method on the DFDC dataset is 0.9261. In [29], the heart rate of people in videos is used to distinguish between original and fake videos. In [63], affective cues related to perceived emotion are used to detect deepfakes. AUC on Deepfake-TIMIT and DFDC dataset is 0.966 and 0.844 respectively. Ciftci et al. [17] extract 32 photoplethysmography signals from the face, windows of frames, and video of windows, then generate signatures with these biological signal spaces as unique model features to detect deepfake videos. The accuracy of the method on the FaceForensics++ dataset is 0.9339.

Model features. As shown in Table 2, current deepfake generation methods are mainly based on deep learning methods, GAN in particular, to generate realistic deepfake images and videos. And there will be model features left in deepfake medias. Yu et al. [91] show that there are specific model fingerprints in GAN and GAN generated medias. So GAN fingerprints can be used to detect fake images and videos generated with GAN. Pu et al. [75] use unique patterns in the noise space of images generated by GAN to detect deepfake images generated by GAN-based methods. The average F1 score of the test on different GAN-generated image datasets is 0.9968. Besides, there will be convolutional traces of fingerprints generated during the generative process of GAN. L. Guarnera et al. [36] propose a deepfake image detection method based on convolutional trace. In the method, an expectation-maximization algorithm is presented to extract convolutional traces in GAN-generated images. In DeepFD [40], contrastive loss is used to obtain typical features from GAN-generated synthesized images. However, these model features are mainly from GAN models and GAN fingerprints based deepfake detection methods are proved to be not robust. For example, GANprintR [68] can remove GAN fingerprints in the deepfake videos successfully, these GAN fingerprint based detection methods will not work.

Media features. Apart from biometric features and model features, there are some media features in deepfake medias that can be used to distinguish deepfake medias, such as temporal information, inconsistency between frames, and noise artifacts. Sabir et al. [82] use the temporal information present in media streams to detect face manipulation in videos. While Li et al. [61] use face warping artifacts to detect fake videos. Due to the inconsistent choice of illuminants between scenes with frames in the fake videos, Güera et al. [32] propose a temporal-aware pipeline to automatically detect deepfake videos. In [89], inconsistencies across the temporal domain are used as features to detect deepfake medias. In [7], CNN classifiers are used to learn inter-frame dissimilarities. In [48], inconsistencies between deepfake faces and the rest of the background, such as differences in lighting conditions, are used to distinguish deepfake medias.

Zhao et al. [95] propose a two-stream CNN to learn semantic anomalies and noise artifacts features to detect deepfake medias. Zhuang et al. [97] find that there are some common fake features in fake images, which can be used to detect fake images. SDHF

[57] is a multilevel deepfake video detection framework that uses CNN to extract frame-level features, clip-level features, and video-level features. Then combine these features to distinguish deepfake videos. Test results show that the highest accuracy of SDHF is 0.982 on the FaceForensics++ dataset.

Li et al. [59] use the output of the last convolutional layer in VGG-16 as features to detect fake images generated by GAN. In [42], image EXIF metadata can be used to determine whether an image is self-consistent. In [15], visual frames, edge maps, and dense optical flow maps are used to detect deepfake media. In [26], underlying patterns from original contents and frequency domain representation are used to detect deepfake images and videos. The highest test result is 1 on the DF-TIMIT dataset. In [49], the probability distribution of intensity values in each pixel of images extracted by PixelRNN are used as general-purpose features to distinguish deepfake images. The average accuracy of this detection method on four different datasets is 0.9573.

4.1.2 Machine learning-based detection methods

Traditional machine learning. Traditional machine learning methods applied in deepfake detection methods includes SVM and random forest. Trans-DF [74] is a random forest-based method to detect deepfake videos, detection accuracy of Trans-DF is 0.902. Agarwal et al. [4] present a one-class SVM based detection method to distinguish deepfake. Yang et al. [90] propose a SVM classifier based method which uses differences between head poses as features to differentiate deepfakes images and videos.

Deep learning methods. Compared with traditional machine learning models, deep learning methods can extract deepfake features automatically and are widely used in deepfake detection, such as GAN, CNN, and RNN. Furthermore, deep learning based detection methods can achieve better detection accuracy.

Bayar et al. [8] present a forensic approach based on deep learning to detect potential manipulation. Rahmouni et al. [77] propose a deep learning method that combines feature extraction and a CNN framework to distinguish computer-generated fake images. Sabir et al. [82] use recurrent convolutional models to detect face manipulation. Li et al. [61] use a dedicated CNN model to detect face warping artifacts in deepfake videos. Dang et al. [18] present CGFace which use a customized convolutional neural network to detect deepfake images. DeepfakeStack [78] is a two-level detection architecture with base learners and 2nd level classifier. CNN is adopted in the 2nd level classification and accuracy of the method is 0.9965.

Gera et al. [32] use CNN to extract frame-level features, then RNN and features extracted by CNN are used to classify if a video is fake or manipulated. Nataraj et al. [66] use a deep CNN framework to train a model to detect GAN-generated fake images. Mo et al. [64] propose a CNN-based method to identify fake face images generated by PGGAN. Xie et al. [88] propose a light weight deepfake detection method based on modified AlexNet model. Accuracy of test on UADFV dataset, FaceForensics++ dataset, and Celeb-DF dataset are 0.9873, 0.9132, and 0.9885 respectively. NA-VGG [11] improves VGG and uses image noise and image augmentation as features to distinguish original and deepfake medias. The average AUC performance of NA-VGG on the Celeb-DF dataset is 0.857.

As an attention mask is used to generate deepfake images and videos, Zhou et al. [96] present a two-stream network to detect face tampering. Zi et al. [99] propose ADDNet as an attention-mask-based deepfake detection method. There are two versions of

ADDNet, 2D ADDNet and 3D ADDNet. An attention mask combined with 2D CNN and 3D CNN are used to achieve image-level and sequence-level detection. The best detection accuracy of 2D ADDNet and 3D ADDNet are 0.9830 and 0.9982 on FaceForensics++ (HD) dataset.

XcepTemporal, XceptionNet* and VTD-Net are all XceptionNet based deepfake detection methods. XcepTemporal [14] combines XceptionNet and a bidirectional LSTM module to extract latent features, which can be used to discriminate deepfake faces. XceptionNet* [15] is an improved version of XcepTemporal, and visual frames, edge maps, and dense optical flow maps are combined to detect deepfake videos. Moreover, XceptionNet* has better detection accuracy than XcepTemporal. VTD-Net [89] is a frame level deepfake video detection method with CNN, Xception and LSTM. In VTD-Net, multitask cascaded CNN is used to extract faces from video frames, then Xception is used to learn the crucial features of real and fake faces. Finally, LSTM is used to learn the inconsistency across the temporal domain between frames, which can determine whether each frame is tampered or not. The detection accuracy is 0.8256 at frame-level, and 0.9189 at video-level. Amerini et al. [6] implement a detection approach based on inconsistencies among consecutive frames with CNN and LSTM, test accuracy of CNN and LSTM on FaceForensics++ is 0.9172 and 0.9429.

Li et al. [59] use the discriminator of GAN to detect fake images generated by GAN. Afchar et al. [3] present a deep neural network method to automatically and efficiently detect face tampering in videos generated by Deepfake and Face2Face. Yu et al. [91] propose a neural network classifier to classify an image as real or GAN generated by the GAN fingerprints. Hsu et al. [40] present DeepFD, a deep forgery discriminator, to efficiently and effectively detect computer-generated images. DeepFD uses contrastive loss to find the typical features of the synthesized images generated by GAN, then uses a classifier to detect those features.

Besides, capsule network and transfer learning are also used. Nguyen et al. [70] use a capsule network to detect deepfake images and videos. Ding et al. [23] use deep transfer learning to detect face-swapping deepfake media. Pranjal Ranjan et al. [79] use transfer learning to improve the accuracy and generalizability of deepfake detection methods. Huh et al. [42] propose a learning algorithm for detecting visual image manipulations trained only using a large dataset of real photographs. Zhuang et al. [97] propose a pairwise learning based approach for deepfake detection, which can learn common fake features from deepfake medias.

Deep learning based deepfake detection methods are heavily depending on training data. To overcome this problem, some detection methods training on real face datasets are proposed. In OC-FakeDect [47], one-class Variational Autoencoder (VAE) is used and only real face datasets are used and there is no need to train on real face datasets. Test accuracy of OC-FakeDect on FaceForensics++ is 0.975.

4.2 Datasets

Both deepfake generation and detection methods are mainly based on deep learning methods, and large training and test datasets are important and necessary to detect deepfake efficiently. However, it is challenging to collect a large deepfake dataset, especially datasets with human faces, for privacy concerns. Fortunately, there are also some deepfake related datasets released since 2017, including UADFV [58], FaceForensics [80],

Table 3 Deepfake related datasets

Datasets	Type	Number of images	Number of Videos	Year
Flickr-Faces-HQ (FFHQ) [46]	Images	70,000 (fake)	–	2019
CelebA [45]	Images	30,000 (1024×1024)	-	2017
UADFV [58]	Videos	–	98 (49 real + 49 fake)	2018
WildDeepfake [99]	Videos	–	707	2020
Ding et al. [23]	Images	-	420,053	2019
FaceForensics [80]	Images, videos	1,500,000	1004	2019
FaceForensics++ [81]	Images, videos	1,800,000	3000	2019
DeepfakeTIMIT datasets [53]	Videos	–	320	2018
Celeb-DF [60]	Videos	–	1203 (408 real + 795 fake)	2020
MFC Datasets [35]	Videos, images	35,000,000(100,000 manipulated)	300,000 (4000 manipulated)	2019
FFW [50]	Images	53,000	150	2018
VidTIMIT [55]	Videos	–	620	2019
DFDC Preview [24]	Videos	–	5214	2019
DFDC [25]	Videos	–	4113	2020
DeepfakeDetection [21]	Videos	–	3363	2019

FaceForensics++ [81], DeepfakeTIMIT [53], and Celeb-DF [60]. Table 3 is a summary of deepfake related datasets.

Flickr-Faces-HQ (FFHQ) [46] is a human face dataset of 70,000 high-quality images (1024×1024 pixels). FaceForensics [80] dataset is a deepfake face forensic dataset with 1004 videos generated by Deepfakes and Face2Face. FaceForensics++ [81] is an extension of FaceForensics dataset, which has 1000 videos and 1.8 million images. Similar to FaceForensics, FaceForensics++ dataset is generated by popular deepfake generation methods, including FaceSwap, Deepfakes, Face2Face, and NeuralTexures. FaceForensics++ dataset is of diversity, with different gender and resolutions. And videos in FaceForensics++ are of different quality levels as VGA(480p), HD (720p), and FHD (1080p).

MFC Datasets [35] is a large-scale media forensic benchmark dataset with 35 million images and 300,000 videos. Of which, there are three subsets generated with GAN as GAN full sets (134 images), GAN crop sets (1000 images), and GAN video sets(118 videos). DeepfakeDetection [21] dataset contains 977 videos collected from Youtube and over 3000 manipulated videos generated by Deepfakes. The UADFV [58] dataset has 98 videos in total, and 49 of which are fake videos generated by FakeAPP.

Celeb-DF [60] dataset is a deepfake video dataset with 6229 videos and more than 2 million frames in total. In Celeb-DF, there are 590 real videos and 5639 deepfake videos generated by Deepfake. These 590 real videos are 59 celebrities of different ages, genders, and races. The average length of videos in Celeb-DF is 13 seconds, and the frame rate is 30 FPS (frame per second). Compared with other deepfake datasets, deepfake videos in Celeb-DF are with higher resolution and visual quality (256×256 pixels).

Ding et al. [23] present a dataset with 420,053 images of 86 celebrities, and all these images are still images rather than video frames. DeepfakeTIMIT dataset [53] consists of 620 deepfake videos generated by face swapping algorithms. And there are two kinds of

videos in DeepfakeTIMIT dataset as low quality (64×64) and high quality (128×128). Fake Face in the Wild (FFW) [50] dataset has more than 53,000 fake and tampered images from 150 videos. Pavel Korshunov and Sebastien Marcel [55] present VidTIMIT dataset with 620 deepfake videos generated with GAN based methods. In Oct. 2019, Facebook released the Deepfakes Detection Challenge (DFDC) Preview dataset with 5024 videos generated by two face-swapping algorithms. In 2020, the final version of Deepfakes Detection Challenge (DFDC) dataset was released, which is composed of 4113 videos from 66 actors with different genders, ages, and ethnic groups. WildDeepfake [99] dataset has 707 videos, 7314 face sequences (3805 is real and 3509 is fake), and 1,180,099 face images. All 707 videos are collected online and of high quality, as low-quality videos and images are dropped. WildDeepfake dataset has more diversity and diverse scenes and is more realistic.

As shown in Table 3, most deepfake related datasets are collected online, mainly from youtube.com, such as Celeb-DF, FaceForensics, FaceForensics++. While the source of the DeepfakeDetection, DFDC Preview, DFDC, DeepFakeDetection datasets are from actors, and these datasets have more diversities and scenarios, which is much closer to real world scenarios.

5 Discussions

Deepfake has been regarded as an emerging threat that calls for urgent responses. Under this circumstance, deepfake detection is still a big challenge. Existing detection methods mainly use machine learning and deep learning methods and features extracted from deepfake images and videos. However, accuracy and robustness of deepfake detection methods are still low. In the section, we discuss challenges and future directions on deepfake detection.

5.1 Challenges

Evolving technologies. Both deepfake generation and detection techniques keep evolving. However, accuracy of current detection methods is still low. What's worse, deepfake generation technology is constantly evolving. As new deepfake generation methods emerged, both the accuracy and efficiency of current detection methods will decrease, making deepfake harder to detect. Besides, state of the art deepfake detection methods are not robust. Carlini et al. [9] show that current deepfake detection methods are vulnerable to both black-box and white-box attacks, including distortion-minimizing attack, loss-maximizing attack, and adversarial patch attack. These attacks on deepfake detection methods can reduce detection accuracy to 0. Gandhi et al. [30] show that adversarial perturbations added to deepfake images have a significant influence on detection result. Huang et al. [41] use shallow reconstruction to diminish artifact patterns, which can significantly reduce the detection accuracy. In summary, deepfake detection is still a great challenge, and robust detection methods that can detect deepfake images and videos effectively are necessary.

Lack of high quality datasets. Most current deepfake detection methods are based on deep learning methods, which need large scale training datasets. The more high quality datasets, the better detection result can achieve. Large scale datasets for training and testing are necessary for efficient deepfake detection. However, the number of high quality deepfake images and videos is still limited. Current deepfake related datasets are generated with prevailing deepfake generation algorithms and applications, which are mainly on

human face datasets and collected from online websites. These datasets often lack diversity and enough scenarios, which are not enough for deepfake detection in real-world scenarios. Besides, it is challenging to construct a large scale dataset with diversity for deepfake detection for potential privacy concerns.

Lack of benchmark. Although there are some assessments on current deepfake detection methods, there is no standardized and uniform deepfake detection benchmark to follow. For example, current deepfake datasets are of different resolutions (for images and videos), different length (for video only), lack of diversity, so standard benchmark datasets are necessary for deepfake detection. Furthermore, there should be automated benchmark test methods to assess both deepfake generation and detection methods.

5.2 Future directions

With the development of deepfake generation methods, biometric features like eye blinking may have little help to the deepfake detection. Future directions on deepfake detection involve development of systematical methods which should combine multi-modal cues from fake images and videos and can achieve better performance and robustness. With the application of new deep learning algorithms in deepfake generation, the quality of deepfake images and videos is greatly improved. Traditional deepfake detection techniques are not sufficiently reliable or efficient for deepfake detection, especially for deepfake videos. There should be robust deepfake detection systems that are resistant to adversarial attacks. Zhu et al. [98] embed regularization into deepfake detection methods that can resist adversarial examples and improve the robustness and generalization of deepfake detection method method. Another direction on deepfake detection should be datasets. Compared with deepfake image detection, deepfake video detection is more challenging and more attention should be paid. Large scale datasets are necessary for detection methods based on deep learning models. Only with enough training data can prominent features extracted from deepfake medias and achieve better detection accuracy. Besides, both training datasets and benchmark datasets should be closer to real world scenarios and with diversity in gender, age, ethnic groups and scenarios.

6 Conclusion

Advances on AI, deep learning and GAN in particular, make it easy to make deepfake images and videos. Everyone with a smartphone can make realistic deepfake images and videos. Deepfake has been regarded as an emerging threat, which calls for novel detection methods. In the paper, we make a survey on state-of-the-art deepfake generation, detection methods and related datasets, which shows that GAN and other deep learning methods have a good performance on deepfake detection. However, here are still some challenges for deepfake detection, including evolution of deepfake generation technology, lack of high quality datasets, lack of benchmark assessment methods and datasets. To effectively combate deepfake threats, future directions on deepfake detection involve systematical methods and large scale high quality deepfake datasets. Systematical methods should take advantages of current detection methods and combine multi-modal features. Compared with deepfake image detection, there should be more attention on deepfake videos detection in the future.

Acknowledgements This work was supported by Guangxi Key Laboratory of Cryptography and Information Security (GCIS201806), Guangxi Key Laboratory of Trusted Software (No. kx202016), Key Lab of Film and TV Media Technology of Zhejiang Province (No.2020E10015), Guangxi Key Laboratory of Hybrid Computation and IC Design Analysis, Guangxi University for Nationalities(GXIC20-03), Key Laboratory of Oceanographic Big Data Mining & Application of Zhejiang Province(obdma202001). We'd like to thank Zelei Cheng from Purdue University and Yingjie Wang from Virginia Tech for writing assistance, language editing, and proofreading.

Declarations

Conflicts of interest The authors declare that they have no conflict of interest.

References

1. (2018) Malicious deep fake prohibition act of 2018. https://www.congress.gov/bill/115th-congress/senate-bill/3805/text
2. (2019) Defending each and every person from false appearances by keeping exploitation subject to accountability Act of 2019. https://www.congress.gov/bill/116th-congress/house-bill/3230/text
3. Afchar D, et al (2018) Mesonet: a compact facial video forgery detection network. In: 2018 IEEE international workshop on information forensics and security (WIFS). IEEE
4. Agarwal S, et al (2019) Protecting world leaders against deep fakes. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops
5. Ajder H, Cavalli F, Patrini G, Cullen L (2019) The state of Deepfakes: Landscape, threats, and impact
6. Amerini I, Caldelli R (2020) Exploiting prediction error inconsistencies through LSTM-based classifiers to detect deepfake videos. In: Proceedings of the 2020 ACM workshop on information hiding and multimedia security. Association for Computing Machinery: Denver, CO, USA, pp 97–102
7. Amerini I, et al (2019) Deepfake video detection through optical flow based CNN. In: Proceedings of the IEEE international conference on computer vision workshops
8. Bayar B, Stamm MC (2016) A deep learning approach to universal image manipulation detection using a new convolutional layer, pp 5–10
9. Carlini N, Farid H (2020) Evading deepfake-image detectors with white- and black-box attacks. In: 2020 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)
10. Chan C, et al (2018) Everybody dance now. arXiv:1808.07371
11. Chang X, et al (2020) Deepfake face image detection based on improved VGG convolutional neural network. In: 2020 39th chinese control conference (CCC)
12. Chen D et al (2019) Face swapping: realistic image synthesis based on facial landmarks alignment. Mathematical Problems in Engineering 2019:1–11
13. Chesney R, Citron DK (2018) Deep fakes: a looming challenge for privacy, democracy, and national security
14. Chintha A et al (2020) Recurrent convolutional structures for audio spoof and video deepfake detection. IEEE Journal of Selected Topics in Signal Processing 14(5):1024–1037
15. Chintha A, et al (2020) Leveraging edges and optical flow on faces for Deepfake detection. In: 2020 IEEE international joint conference on biometrics (IJCB)
16. Choi Y, et al (2018) Stargan: Unified generative adversarial networks for multi-domain image-to-image translation. In: Proceedings of the IEEE conference on computer vision and pattern recognition
17. Ciftci UA, Demir I, Yin L (2020) How do the hearts of deep fakes beat? Deep fake source detection via interpreting residuals with biological signals. In: 2020 IEEE international joint conference on biometrics (IJCB)
18. Dang L et al (2018) Deep learning based computer generated face identification using convolutional neural network. Applied Sciences 8(12):2610
19. DeepFaceLab. Available from: https://github.com/iperov/DeepFaceLab
20. Deepfake reddit. Available from: https://www.reddit.com/user/Deepfakes/
21. DeepFakeDetection data. https://github.com/ondyari/FaceForensics/blob/master/dataset/README.md
22. Deepfakes. Available from: https://github.com/Deepfakes/

23. Ding X, et al (2019) Swapped face detection using deep learning and subjective assessment. arXiv:1909.04217
24. Dolhansky B, et al (2019) The deepfake detection challenge (DFDC) preview dataset. arXiv:1910.08854
25. Dolhansky B, et al (2020) The deepfake detection challenge (DFDC) dataset. arXiv:2006.07397
26. Du CXT, et al (2020) Efficient-frequency: a hybrid visual forensic framework for facial forgery detection. In: 2020 IEEE symposium series on computational intelligence (SSCI)
27. Faceapp. Available from: http://www.faceapp.com/
28. Farid H (2009) Image forgery detection. IEEE Signal Processing Magazine 26(2):16–25
29. Fernandes S, et al (2019) Predicting heart rate variations of deepfake videos using neural ODE. In: Proceedings of the IEEE international conference on computer vision workshops
30. Gandhi A, Jain S (2020) Adversarial perturbations fool deepfake detectors. In: 2020 international joint conference on neural networks (IJCNN)
31. Geng Z, Cao C, Tulyakov S (2019) 3d guided fine-grained face manipulation. In: Proceedings of the IEEE conference on computer vision and pattern recognition
32. Gera D, Delp EJ (2018) Deepfake video detection using recurrent neural networks. In: 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS)
33. Goodfellow I, et al (2014) Generative adversarial nets. In: Advances in neural information processing systems
34. Gouhara K, Watanabe T, Uchikawa Y (1991) Learning process of recurrent neural networks. In: [Proceedings] 1991 IEEE international joint conference on neural networks
35. Guan H, et al (2019) MFC datasets: Large-scale benchmark datasets for media forensic challenge evaluation. In: 2019 IEEE winter applications of computer vision workshops (WACVW). IEEE
36. Guarnera L, Giudice O, Battiato S (2020) Deepfake detection by analyzing convolutional traces. In: 2020 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)
37. Gupta P, et al (2020) The eyes know it: FakeET-an eye-tracking database to understand Deepfake perception, In: Proceedings of the 2020 international conference on multimodal interaction, Association for Computing Machinery: Virtual Event, Netherlands, pp 519–527
38. Hearst MA et al (1998) Support vector machines. IEEE Intelligent Systems and their Applications 13(4):18–28
39. Hochreiter S, Schmidhuber J (1997) Long short-term memory. Neural Computation 9(8):1735–1780
40. Hsu C-C, Lee C-Y, Zhuang Y-X (2018) Learning to detect fake face images in the wild. In: 2018 international symposium on computer, consumer and control (IS3C). IEEE
41. Huang Y, et al (2020) FakePolisher: making deepfakes more detection-evasive by shallow reconstruction. In: Proceedings of the 28th ACM international conference on multimedia. Association for Computing Machinery, pp 1217–1226
42. Huh M, et al (2018) Fighting fake news: Image splice detection via learned self-consistency. In: Proceedings of the European conference on computer vision (ECCV)
43. Isola P, et al (2017) Image-to-image translation with conditional adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition
44. Jung T, Kim S, Kim K (2020) DeepVision: Deepfakes detection using human eye blinking pattern. IEEE Access 8:83144–83154
45. Karras T, et al (2017) Progressive growing of gans for improved quality, stability, and variation. arXiv:1710.10196
46. Karras T, Laine S, Aila T (2019) A style-based generator architecture for generative adversarial networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition
47. Khalid H, Woo SS (2020) OC-FakeDect: classifying deepfakes using one-class variational autoencoder. In: 2020 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)
48. Kharbat FF, et al (2019) Image feature detectors for deepfake video detection. In: 2019 IEEE/ACS 16th international conference on computer systems and applications (AICCSA)
49. Khodabakhsh A, Busch C (2020) A generalizable deepfake detector based on neural conditional distribution modelling. In: 2020 international conference of the biometrics special interest group (BIOSIG)
50. Khodabakhsh A, et al (2018) Fake face detection methods: can they be generalized? In: 2018 international conference of the biometrics special interest group (BIOSIG). IEEE
51. Kim H et al (2018) Deep video portraits. ACM Transactions on Graphics (TOG) 37(4):163
52. Korshunova I, et al (2017) Fast face-swap using convolutional neural networks. In: Proceedings of the IEEE international conference on computer vision
53. Korshunov P, Marcel S (2018) Deepfakes: a new threat to face recognition? assessment and detection. arXiv:1812.08685

54. Korshunov P, Marcel S (2018) Speaker inconsistency detection in tampered video. In: 2018 26th european signal processing conference (EUSIPCO). IEEE
55. Korshunov P, Marcel S (2019) Vulnerability assessment and detection of Deepfake videos. In: The 12th IAPR international conference on biometrics (ICB)
56. LeCun Y, Bengio Y, Hinton G (2015) Deep learning. Nature 521(7553):436–444
57. Liang T, et al (2020) SDHF: spotting deepfakes with hierarchical features. In: 2020 IEEE 32nd international conference on tools with artificial intelligence (ICTAI)
58. Li Y, Chang M-C, Lyu S (2018) In ictu oculi: Exposing ai created fake videos by detecting eye blinking. In: 2018 IEEE international workshop on information forensics and security (WIFS). IEEE
59. Li H, et al (2018) Can forensic detectors identify GAN generated images? In: 2018 Asia-pacific signal and information processing association annual summit and conference (APSIPA ASC)
60. Li Y, et al (2020) Celeb-DF: A large-scale challenging dataset for deepfake forensics. In: 2020 IEEE/CVF conference on computer vision and pattern recognition (CVPR)
61. Li Y, Lyu S (2018) Exposing Deepfake videos by detecting face warping artifacts. arXiv:1811.00656
62. Matern F, Riess C, Stamminger M (2019) Exploiting visual artifacts to expose Deepfakes and face manipulations. In: 2019 IEEE winter applications of computer vision workshops (WACVW)
63. Mittal T, et al (2020) Emotions Don't Lie: An audio-visual deepfake detection method using affective cues. In: Proceedings of the 28th ACM international conference on multimedia. Association for Computing Machinery, pp 2823–2832
64. Mo H, Chen B, Luo W (2018) Fake faces identification via convolutional neural network. In: Proceedings of the 6th ACM workshop on information hiding and multimedia security. ACM
65. Montserrat DM, et al (2020) Deepfakes detection with automatic face weighting. In: 2020 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)
66. Nataraj L, et al (2019) Detecting GAN generated fake images using co-occurrence matrices. arXiv:1903.06836
67. Natsume R, Yatagawa T, Morishima S (2018) RSGAN: face swapping and editing using face and hair representation in latent spaces. arXiv:1804.03447
68. Neves JC et al (2020) GANprintR: improved fakes and evaluation of the state of the art in face manipulation detection. IEEE Journal of Selected Topics in Signal Processing 14(5):1038–1048
69. Nguyen HM, Derakhshani R (2020) Eyebrow recognition for identifying Deepfake videos. In: 2020 international conference of the biometrics special interest group (BIOSIG)
70. Nguyen HH, Yamagishi J, Echizen I (2019) Capsule-forensics: using capsule networks to detect forged images and videos. In: ICASSP 2019-2019 IEEE international conference on acoustics, speech and signal processing (ICASSP). IEEE
71. Nirkin Y, Keller Y, Hassner T (2019) FSGAN: Subject agnostic face swapping and reenactment. arXiv:1908.05932
72. Oord Avd, et al (2016) Wavenet: A generative model for raw audio. arXiv:1609.03499
73. Pan SJ, Yang Q (2009) A survey on transfer learning. IEEE Transactions on Knowledge and Data Engineering 22(10):1345–1359
74. Patel M, et al (2020) Trans-DF: a transfer learning-based end-to-end deepfake detector. In: 2020 IEEE 5th international conference on computing communication and automation (ICCCA)
75. Pu J, et al (2020) NoiseScope: detecting deepfake images in a blind setting. In: Annual computer security applications conference. Association for Computing Machinery, Austin, pp 913–927
76. Pumarola A, et al (2018) Ganimation: Anatomically-aware facial animation from a single image. In: Proceedings of the european conference on computer vision (ECCV)
77. Rahmouni N, et al (2017) Distinguishing computer graphics from natural images using convolution neural networks. In: 2017 IEEE workshop on information forensics and security (WIFS). IEEE
78. Rana MS, Sung AH (2020) DeepfakeStack: a deep ensemble-based learning technique for deepfake detection. In: 2020 7th IEEE international conference on cyber security and cloud computing (CSCloud)/2020 6th IEEE international conference on edge computing and scalable cloud (EdgeCom)
79. Ranjan P, Patil S, Kazi F (2020) Improved generalizability of deep-fakes detection using transfer learning based CNN framework. In: 2020 3rd international conference on information and computer technologies (ICICT)
80. Rssler A, et al (2018) Faceforensics: A large-scale video dataset for forgery detection in human faces. arXiv:1803.09179
81. Rssler A, et al (2019) FaceForensics++: learning to detect manipulated facial images. In: 2019 IEEE/CVF international conference on computer vision (ICCV)
82. Sabir E et al (2019) Recurrent convolutional strategies for face manipulation detection in videos. Interfaces (GUI) 3:1

83. Thies J et al (2018) Headon: Real-time reenactment of human portrait videos. ACM Transactions on Graphics (TOG) 37(4):164
84. Thies J, Zollhfer M, Niener M (2019) Deferred neural rendering: image synthesis using neural textures. ACM Trans. Graph 38(4):66
85. Thies J, et al (2016) Face2face: Real-time face capture and reenactment of rgb videos. In: Proceedings of the IEEE conference on computer vision and pattern recognition
86. Tursman E, et al (2020) Towards untrusted social video verification to combat deepfakes via face geometry consistency. In: 2020 IEEE/CVF conference on computer vision and pattern recognition workshops (CVPRW)
87. Wang S-Y, et al (2019) Detecting photoshopped faces by scripting photoshop. arXiv:1906.05856
88. Xie D, et al (2020) Deepfake detection on publicly available datasets using modified AlexNet. In: 2020 IEEE symposium series on computational intelligence (SSCI)
89. Yang T, et al (2020) VTD-Net: depth face forgery oriented video tampering detection based on convolutional neural network. In: 2020 39th chinese control conference (CCC)
90. Yang X, Li Y, Lyu S (2019) Exposing deep fakes using inconsistent head poses. In: 2019 IEEE international conference on acoustics, speech and signal processing (ICASSP)
91. Yu N, Davis L, Fritz M (2018) Attributing fake images to GANs: Analyzing fingerprints in generated images. arXiv:1811.08180
92. ZAO. Available from: https://www.zaoapp.net
93. Zakharov E, et al (2019) Few-shot adversarial learning of realistic neural talking head models. arXiv:1905.08233
94. Zhao Y et al (2020) Capturing the persistence of facial expression features for deepfake video detection. Springer International Publishing, Cham
95. Zhao Z, Wang P, Lu W (2020) Detecting deepfake video by learning two-level features with two-stream convolutional neural network. In: Proceedings of the 2020 6th international conference on computing and artificial intelligence. Association for Computing Machinery, pp 291–297
96. Zhou P, et al (2017) Two-stream neural networks for tampered face detection. In: 2017 IEEE conference on computer vision and pattern recognition workshops (CVPRW). IEEE
97. Zhuang Y-X, Hsu C-C (2019) Detecting generated image based on a coupled network with two-step pairwise learning. In: 2019 IEEE international conference on image processing (ICIP). IEEE
98. Zhu K, Wu B, Wang B (2020) Deepfake detection with clustering-based embedding regularization. In: 2020 IEEE fifth international conference on data science in cyberspace (DSC)
99. Zi B, et al (2020) WildDeepfake: a challenging real-world dataset for deepfake detection. In: Proceedings of the 28th ACM international conference on multimedia. Association for Computing Machinery, pp 2382–2390