TIJER || ISSN 2349-9249 || © February 2024, Volume 11, Issue 2 || www.tijer.org

Detecting Deepfake: Audio, Video & Image using Deep Learning Technique

1st Nirav K. Shah, 2nd Arpit B. Parekh

[1]Assistant Professor, Department of Computer Science, Shree Swaminarayan College of Computer Science, Bhavnagar, India

[2]Assistant Professor, Department of Computer Science, Shree Swaminarayan College of Computer Science, Bhavnagar, India

Abstract - Deepfake is a rapidly developing area of artificial intelligence technology that is widely used on social media. It involves superimposing one person's photos, sounds, and videos over those of another. The key components of Deepfake are machine learning (ML) and deep learning (DL), which can identify them and produce Deepfake audio, video, and images much more quickly and cheaply. The term "Deepfake" has bad connotations, although the technology is being used more frequently both individually and commercially. Politicians and celebrities are the main targets in an effort to harm their reputations. In this paper we are explore the techniques to detect Deepfake audio, video and images.

Index Terms – Deepfake, Deep synthetic, machine learning, artificial intelligence, deep learning, speech synthesis, synthetic voice.

I. INTRODUCTION

Recently, technologies synthesized using AI have been developed that can produce voices that are believable. Though these technologies were intended to be helpful, they have also been used maliciously to use audio to broadcast misinformation around the world, which has created concern of the "Audio Deepfake." These days, anyone can easily obtain audio Deepfakes (also known as audio manipulations) with basic mobile devices or personal computers. As a result, public cybersecurity worries about the adverse impacts of AD use have spread throughout the globe. It can be applied to logical-access voice spoofing, which is a tactic used to sway public opinion for terrorist, propaganda, and defamatory purposes. Every day, enormous volumes of voice recordings are shared online, and it can be difficult to distinguish fakes among them. Nevertheless, AD attackers have also attacked governments and politicians in addition to people and organizations. In 2019, scammers mimicked a CEO's voice over the phone using AI-based software. To prevent the spread of false information, we must authenticate any distributed audio recordings. Thus, the scholarly community has been quite interested in this subject recently. Many detection techniques have been created in relation to Deepfake in order to distinguish fake audio recordings from real speech. Many machine learning and deep learning models have been created that employ various techniques to identify false audio. The overall AD detection procedure is described by the strategies listed below, as seen in Figure 1.

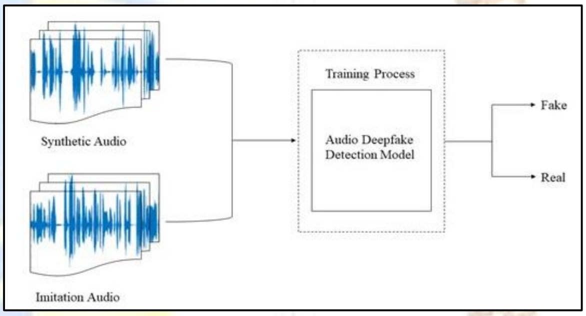


Fig.1 Audio Detection Procedure

Each audio clip should first undergo preprocessing to convert it into appropriate audio features, like Mel-spectrograms. The detection model receives these features and uses them to carry out the necessary functions, such training. If the job involves nonlinearity, the output can be fed into any fully connected layer that has an activation function to get a prediction likelihood of class 0 being fake or class 1 being real. But there's a trade-off between computing complexity and precision.In recent years, the issue of face-manipulated movies has drawn a lot of attention, particularly with the introduction of Deepfake technology, which uses deep learning techniques to alter videos. Using generative adversarial networks or autoencoders, the Deepfake algorithm can swap out faces in the source video with faces in the target video. Face-manipulated videos can be produced quite easily with this technology, provided that a lot of data can be accessed. Deepfake technology is still widely exploited for malevolent purposes, even if it has potential applications in fields like virtual reality and filmmaking. According to Figure 2.

Fig.2 Face Deepfake

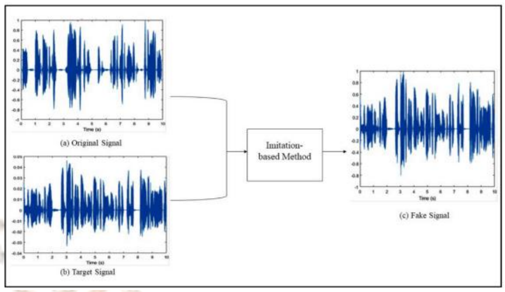
On the Internet, a vast array of phone videos have been shared, the majority of them are directed towards celebrities and politicians. Since Deepfake technology was first developed, it is inevitable that it will be used maliciously. The first example of Deepfake content was a celebrity pornographic film made in 2017 by a Reddit user going by the handle Deepfakes. FakeApp, FaceSwap, and other Deepfake-based apps quickly started to show up on a regular basis. A clever undressing software called Deepnude was even released in June 2019, which caused a global panic. In addition to violating privacy, these applications' produced videos are being used more frequently to sway public opinion and political campaigns. Identifying Deepfake content has become as a critical concern for people, organizations, and governments worldwide. The greatest distinguishing characteristic of humans is their face. The security risk posed by face modification is growing in importance due to the rapid advancements in face synthesis technologies. Thanks to the plethora of deep learning algorithms, people's faces can frequently be substituted with those of others that seem real. Deepfake technology can use people's images, sounds, and voices to produce political, satirical, or sexual content about them without their permission. Anyone can create any fake content that is undetectable from the real thing because of how simple it is to use different applications.

Cyberbullying is affecting a lot of young people. In the worst situation, a great number of victims take their own lives.

II. AUDIO DEEPFAKE ATTACK.

The use of AD technology, which was developed recently, users may now produce audio clips that mimic the speech of real people. Originally, this technique was created for a range of uses aimed at enhancing human life, like audiobooks, where it might be applied to simulate calming voices. Three primary categories of audio fakeness—imitation-based, synthetic-based, and replay-based Deepfakes—are identified by the AD literature.

Based on imitation In order to preserve the anonymity of the secret audio, Deepfakes are "a method of transforming speech (secret audio) so that it sounds like another speech (target audio)". There are several techniques to mimic a voice, such as using people whose voices are similar to the actual speaker. Nevertheless, masking algorithms have been developed to mimic audio and Deepfake speech, such as Efficient Wavelet Mask (EWM). Specifically, comparable qualities will be recorded for the original and target sounds. The signal of the original audio, Figure 2a, will then be converted, as indicated in Figure 3, to say the speech in the target audio, Figure 2b, using an imitation generation technique that will produce a new speech, which is the false one, shown in Figure 2c. Because of this, it is challenging for people to distinguish between the real and artificial sounds produced by this technology.



Three modules make up Synthetic-based or Text-To-voice (TTS): a text analysis model, an acoustic model, and a vocoder. TTS attempts to convert text into understandable and natural voice in real time [9]. There are two essential procedures to take in order to produce synthetic Deepfake audio. First, clear, well-organized raw audio should be gathered together with an audio speech transcript. Second, in order to create a model for creating synthetic sounds, the TTS model needs to be trained using the information gathered. The most natural-sounding audio can be produced using well-known model generation approaches as Tactoran 2, Deep Voice 3, and FastSpeech 2. Tactoran 2 uses a modified WaveNet vocoder to produce Mel-spectrograms. A position-augmented attention method is employed by the neural text-to-speech model Deep Voice 3 to provide an attention-based decoder. The fastest training time is achieved with high-quality results with FastSpeech 2. The generation model in the synthetic technique will be given the transcript text containing the target speaker's voice. After processing the received text, the text analysis module transforms it into linguistic features. Based on the linguistic features produced by the text analysis module, the acoustic module then extracts the target speaker's parameters from the dataset. The final step involves teaching the vocoder how to produce speech waveforms using the parameters of the acoustic feature. This will result in the generation of the final audio file, which contains the synthetic false audio in a waveform format. The procedure of creating synthetic voice is shown in Figure 3.

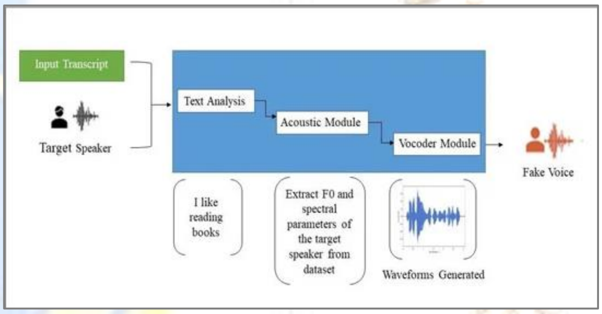


Fig.3 Synthetic Voice Creating Procedure

One kind of malicious operation that attempts to replay a recording of the target speaker's voice is called a replay-based Deepfake. Cut-and-paste detection and far-field detection are the two varieties. A test segment of the victim's microphone recording is played on a phone handset equipped with a loudspeaker in the far-field detection process. Cutting and pasting, on the other hand, entails fabricating the sentence that a text-dependent system demands. Rather than focusing on techniques that use modified recordings, this article will concentrate on Deepfake methods that mimic actual voices. The detection techniques for distinguishing between synthetic and imitation Deepfakes will therefore be covered in this review; replay-based attacks will be deemed outside of its purview.

III. TECHNIQUE USED IN IMAGE AND VIDEO DEEPFAKE DETECTION.

Generally speaking, Deepfake detection is considered a binary classification problem in which altered and real movies must be distinguished using classifiers. To train classification algorithms, this type of approach needs a sizable collection of authentic and fraudulent movies. Although there are more and more false videos online, there are still not enough of them to establish a standard by which different detection techniques may be evaluated. In order to solve this problem, Korshunov and Marcel [1] used the open source code Faceswap-GAN [2] to create a noteworthy Deepfake dataset of 620 movies based on the GAN model. The VidTIMIT database, which is accessible to the public, was utilized to create Deepfake films of varying quality. These videos can accurately replicate facial emotions, lip movements, and eye blinking. Then, different Deepfake detection techniques were tested using these films. According to test results, popular face recognition algorithms based on VGG [3] and Facenet are not capable of accurately identifying Deepfake. When used to identify Deepfake videos from this recently created dataset, other techniques like lip-syncing approaches and picture quality measurements with support vector machines (SVM) [4] result in extremely high error rates. This raises questions regarding the urgent need for more reliable techniques to be developed in the future in order to distinguish real Deepfake from fakes.

A. Fake Image Detection Using Handcrafted Features based Method.

Despite the fact that GAN is constantly being developed and that numerous new extensions are often released, the majority of studies on the detection of GAN-generated images do not take the generalization ability of the detection models into account. Gaussian blur and Gaussian noise were two picture preprocessing techniques employed by Xuan et al. [5]. to eliminate GAN picture low level high frequency cues. Compared to earlier image forensics techniques [6,7] or image steganalysis networks [8], this improves the statistical similarity at the pixel level between real and fake images and enables the forensic classifier to learn more intrinsic and meaningful features.

Zhang et al. [9] extracted a collection of compact features using the bag of words method, which they then put into other classifiers including SVM, random forest (RF). using multi-layer perceptrons (MLP) to distinguish between real and swapped face images. The synthesised images produced by GAN models are likely the hardest to identify among deep learning-generated images since they are realistic and of a high caliber due to GAN's capacity to learn the distribution of complex input data and produce fresh outputs with comparable input distribution.

However, Agarwal and Varshney [10] presented the GAN-based deepfake detection as a challenge for hypothesis testing, and they used the information-theoretic research of authentication to offer a statistical framework. The oracle error, which measures the minimal distance between distributions of authentic images and images produced by a specific GAN, is defined. The

According to analytical findings, this gap grows when the GAN's accuracy decreases, making it simpler to identify deepfakes in this situation. When dealing with high-resolution image inputs, a very precise GAN is needed to produce bogus images, which are challenging to identify using this technique.

B. Fake Image Detection Using Deep Features based Method

Face swapping is quite useful for replacing faces in photos with stock photos, which makes it ideal for identity protection as well as video compositing and transfiguration in portraits. pictures. It is also one of the methods used by cybercriminals to get past authentication or identity systems and obtain unauthorized access, though. Since deep learning algorithms like CNN and GAN can retain the lighting, posture, and facial expression of photos, they have increased the difficulty of dealing with swapped face images for forensics models [11].

A two-phase deep learning technique for the identification of deepfake images was presented by Hsu et al. [12]. A feature extractor based on the popular fake is used in the first six phases. application of the Siamese network design described in [13] in a feature network (CFFN). Several dense units, each containing a varying number of dense blocks, are included in the CFFN [14] in order to enhance the representational ability for the input images. Using pairwise information—the label of each pair of two input photos—the CFFN learning process extracts discriminative features between the fake and real images.

The pairwise label is 1 if the two photos are of the same kind, that is, fake-fake or real-real. On the other hand, the pairwise label is 0 if they are of different categories, such as fake-real. The CFFN-oriented A neural network classifier is then fed discriminative information to separate real photos from fake ones. The suggested approach has been verified for the detection of both bogus faces and bogus general images. 202,599 aligned face photos in a variety of positions with background clutter and 10,177 identities make up the first part of the face dataset, which is taken from CelebA [15]. Five GAN. A number of variations, such as deep convolutional GAN (DCGAN) [16], Wasserstein GAN (WGAN) [17], WGAN with gradient penalty (WGAN-GP), least squares GAN, and PGGAN, are employed to create fictitious images with a 64x64 size. To validate the suggested strategy, a total of 385,198 training photos and 10,000 test images—both real and fake—are acquired. Conversely, the overall dataset is taken from the ILSVRC12. Fake 128x128 images are produced using the large scale GAN training model for high fidelity natural image synthesis (BIGGAN) [18], self-attention GAN and spectral normalization GAN. 10,000 of both actual and false photos make up the test set, whereas 600,000 of both kinds make up the training set. The suggested strategy performs better than competing strategies like those introduced according to experimental results.

C. Fake Video Detection Using Temporal Features across Video Frame Method

Temporal Elements in Different Video Frames Sabir et al. [19] used spatiotemporal characteristics of video streams to identify deepfakes, based on the finding that temporal coherence is not adequately enforced in the synthesis process of deepfakes. Frame-by-frame video modification is done in order to further display low level artifacts resulting from face manipulations as temporal artifacts that exhibit frame-to-frame discrepancies. To take advantage of temporal differences between frames, a recurrent convolutional model (RCN) was developed by integrating the convolutional network DeneNet [14] with the gated recurrent unit cells (see Fig. 6). Tested on the 1,000-video FaceForensics++ dataset [20], the suggested approach yields encouraging results.

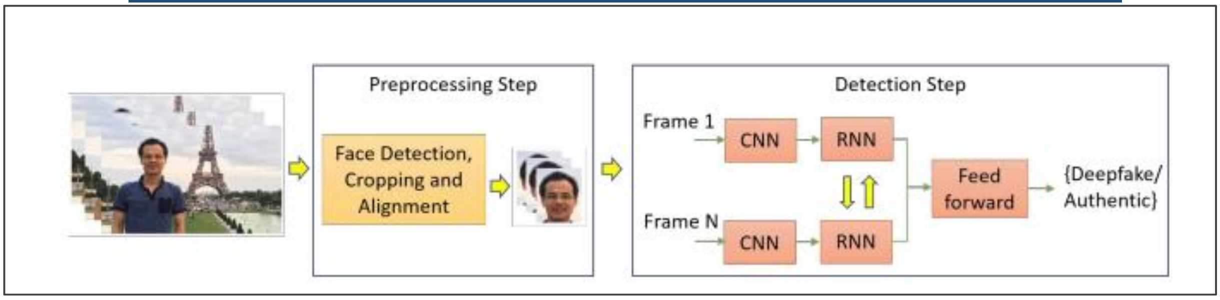


Fig.4 A two step process for face manipulation detection where the preprocessing step aims to detect, crop and align faces on a sequence of frames and second step distinguish manipulated and authentic face image by combing convolution neural network(CNN) and recurrent neural network(RNN).

Similarly, Guera and Delp [21] pointed out that there are temporal and intra-frame anomalies in deepfake videos. The temporal-aware pipeline approach, which employs long short term memory (LSTM) and CNN to identify deepfake films, was then put out. Frame-level features are extracted using CNN and put into the LSTM to produce a temporal sequence descriptor. In the end, a fully-connected network is employed, as shown in Fig. 5, to distinguish between actual and doctored videos using the sequence descriptor. Using a dataset of 600 films—300 deepfake videos gathered from various videohosting websites and 300 clean videos chosen at random from the Hollywood human activities dataset in [22]—an accuracy of more than 97% was achieved.

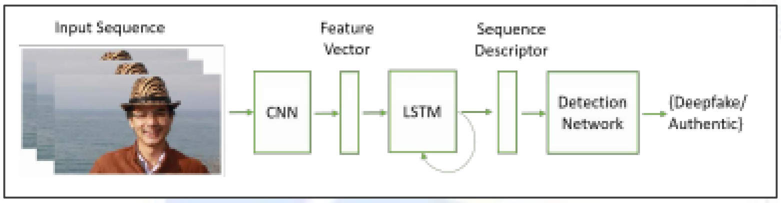


Fig.4 deepfake detection method using convolutional neural network(CNN) and long short term memory(LSTM) to extract temporal feature of a given video sequence which represented via the sequence descriptor. The detection network consisting of fully connected layer is employed to take the sequence descriptor as input and calculate probabilities of the frame sequence belonging to either authentic or Deepfake class.

However, Li et al. [23] suggested using eye blinking as a physiological indicator to identify deepfakes because they show that the individual in a deepfake blinks far less frequently than in a video that hasn't been altered. An adult individual in good health would typically blink every 2 to 10 seconds, taking between 0.1 and 0.4 seconds for each blink. However, deepfake algorithms frequently use face photos that may be found online for training, most of which feature people with open eyes—that is, very few images found online include persons with closed eyes. Thus, deepfake algorithms are unable to produce fake faces that can blink regularly in the absence of photos of real individuals blinking. Put differently, blinking rates in deepfake are far lower than in regular videos. Li et al. [23] crop eye regions in the movies and distribute them into long-term recurrent convolutional networks (LRCN) [24] for dynamic state prediction in order to distinguish between authentic and false videos. The CNN-based feature extractor, the long short-term memory (LSTM)-based sequence learning, and the fully connected layer-based state prediction, which forecasts the likelihood of an eye-open or closed state, comprise the LRCN. Strong temporal relationships can be seen in the blinking of the eyes, and LSTM implementation facilitates the efficient capturing of these temporal patterns.

D. Fake Video Detection Using Visual Artifacts within Video Frame.

As was said in the previous section, deep recurrent network models are the main foundation for the techniques that use temporal patterns across video frames to identify deepfake films. The alternative method, which often breaks down films into frames and looks for visual artifacts within each frame to extract discriminant features, is examined in this subsection. Then, in order to distinguish between real and false videos, these attributes are fed into a deep or shallow classifier. As a result, we categorize techniques in this part according to the kind of classifiers—that is, deep or shallow.

Deep classifiers. Since deepfake films are typically produced at low resolutions, affine face warping—that is, scaling, rotation, and shearing—is necessary. Gotta conform to the original ones' layout. CNN models like VGG16 [25], ResNet50, ResNet101, and ResNet152 [26] can identify artifacts from this technique because to the resolution discrepancy between the warped face area and the surrounding environment. In a deep learning technique for identifying deepfakes was presented. It relies on artifacts seen during the face warping stage of the deepfake generating algorithms. The UADFV and DeepfakeTIMIT deepfake datasets are used to assess the suggested approach. With a total of 32,752 frames, the UADFV dataset includes 49 authentic videos and 49 fraudulent ones.

The DeepfakeTIMIT dataset consists of 10,537 low-quality films with a 64 × 64 size and high-quality videos with a 128 x 128 size. For every quality set, 34,023 fake photos and perfect photographs were taken from 320 videos. The suggested method's performance is contrasted with that of other widely used techniques, including HeadPose, twostream NN and two deepfake detection MesoNet approaches, namely Meso-4 and MesoInception-4. The suggested strategy has the advantage of not requiring the creation of deepfake movies as negative examples prior to training the detection models. Rather, the negative instances are created dynamically by taking the original image's face region and scaling it to fit several scales, then warping and adding Gaussian blur to a scaled version of the random select image returning to the initial picture. As opposed to other approaches that call for the generation of deepfakes beforehand, this requires a significant reduction in both time and computational resources.

shallow classifiers. Deepfake detection techniques mostly rely on intrinsic feature abnormalities or discrepancies between actual and fake photos or videos. By comparing the variations in 3D head poses, which include head orientation and position and are computed using 68 facial landmarks in the central face region, Yang et al. [27] suggested a detection approach. Due to a flaw in the deepfake face creation pipeline, the 3D head positions are analyzed. To get the detection results, the retrieved features are passed into an SVM classifier. Tests conducted on two datasets demonstrate how well the suggested strategy outperforms its rivals. 49 deepfake videos and the corresponding genuine videos make up the first dataset, called UADFV [27]. The second dataset, which is a subset of data utilized in the DARPA MediFor GAN, consists of 241 genuine photos and 252 deep fake images. Challenge Image/Video. Similarly examined a technique to take use of deepfake and face-manipulation artifacts based on eye, tooth, and facial contour visual cues. The absence of global consistency, inaccurate or imprecise calculation of the incident illumination, or imprecise estimation of the underlying geometry are the causes of the visual artifacts. In order to detect deepfakes, textural features derived from the facial region based on facial landmarks are utilized, along with absent reflections and details in the areas surrounding the eyes and teeth. Consequently, features taken from the full-face crop and the eye and teeth feature vectors are used.

IV. CONCLUSIONS

People's confidence in media content has started to decline as a result of deepfakes because believing in them is no longer correlated with seeing them. In addition to intensifying hate speech and disinformation, they may also incite political unrest, agitate the populace, spark violence, or even start a war. These days, this is especially important because deepfake technologies are becoming more accessible and social media sites can swiftly disseminate the fake content. This review offers a comprehensive assessment of the difficulties, possible trends, and future possibilities in the field of deepfake production and detection. It also gives a contemporary summary of these techniques. Therefore, this study will be helpful to the artificial intelligence research community in creating practical strategies to combat deepfakes.

V. REFERENCES

[1] Pavel Korshunov and Sebastien Marcel. Vulnerability assessment and detection of deepfake videos. In 2019 *International Conference on Biometrics (ICB)*, pages 1–6. IEEE, 2019.
[2] Faceswap-GAN. https:/ /github.com/shaoanlu/faceswap-GAN, .
[3] Omkar M Parkhi, Andrea Vedaldi, and Andrew Zisserman. Deep face recognition. *In Proceedings of the British Machine 15 Vision Conference (BMVC)*, pages 41.1–41.12, 2015.
[4] Javier Galbally and Sebastien Marcel. Face anti-spoofing based on general image quality assessment. In The *22nd International Conference on Pattern Recognition*, pages 1173–1178. IEEE, 2014.
[5] Xinsheng Xuan, Bo Peng, Wei Wang, and Jing Dong. On the generalization of GAN image forensics. In *Chinese Conference on Biometric Recognition*, pages 134–141. Springer,2019.
[6] Pengpeng Yang, Rongrong Ni, and Yao Zhao. Recapture image forensics based on laplacian convolutional neural networks. In *International Workshop on Digital Watermarking*, pages 119–128. Springer, 2016.
[7] Belhassen Bayar and Matthew C Stamm. A deep learning approach to universal image manipulation detection using a new convolutional layer. In Proceedings of the 4th ACM *Workshop on Information Hiding and Multimedia Security*, pages 5–10, 2016
[8] Yinlong Qian, Jing Dong, Wei Wang, and Tieniu Tan. Deep learning for steganalysis via convolutional neural networks. *In Media Watermarking, Security, and Forensics*, volume 9409, page 94090J, 2015.
[9] Ying Zhang, Lilei Zheng, and Vrizlynn LL Thing. Automated face swapping and its detection. In The 2nd *International Conference on Signal and Image Processing (ICSIP)*, pages 15–19. IEEE, 2017.
[10] Sakshi Agarwal and Lav R Varshney. Limits of Deepfake detection: A robust estimation viewpoint. arXiv preprint arXiv:1905.03493, 2019.
[11] Iryna Korshunova, Wenzhe Shi, Joni Dambre, and Lucas Theis. Fast face-swap using convolutional neural networks. *In Proceedings of the IEEE International Conference on Computer Vision*, pages 3677–3685, 2017.
[12] Vision, pages 3677–3685, 2017. [85] Chih-Chung Hsu, Yi-Xiu Zhuang, and Chia-Yen Lee. Deep fake image detection based on pairwise learning. *Applied Sciences*, 10(1):370, 2020
[13] Sumit Chopra, Raia Hadsell, and Yann LeCun. Learning a similarity metric discriminatively, with application to face verification. In *IEEE Computer Society Conference on Computer Vision and Pattern Recognition* (CVPR'05), volume 1, pages 539–546. IEEE, 2005.
[14] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks.
In Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition*, pages 4700–4708, 2017.
[15] Ziwei Liu, Ping Luo, Xiaogang Wang, and Xiaoou Tang. Deep learning face attributes in the wild. In *Proceedings of the IEEE International Conference on Computer* Vision, pages 3730–3738, 2015.
[16] Alec Radford, Luke Metz, and Soumith Chintala. Unsupervised representation learning with deep convolutional generative adversarial networks. *arXiv preprint arXiv*:1511.06434, 2015.
[17] Martin Arjovsky, Soumith Chintala, and Leon Bottou. Wasser-stein generative adversarial networks. In *International Conference on Machine Learning*, pages 214–223. PMLR, 2017.
[18] Andrew Brock, Jeff Donahue, and Karen Simonyan. Large scale GAN training for high fidelity natural image synthesis. *arXiv preprint arXiv*:1809.11096, 2018
[19] Ekraam Sabir, Jiaxin Cheng, Ayush Jaiswal, Wael AbdAlmageed, Iacopo Masi, and Prem Natarajan. Recurrent convolutional strategies for face manipulation detection in videos. Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 3(1):80–87, 2019.

[20] Andreas Rossler, Davide Cozzolino, Luisa Verdoliva, Christian Riess, Justus Thies, and Matthias Nießner. FaceForensics++: Learning to detect manipulated facial images. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 1–11, 2019.
[21] David Guera and Edward J Delp. Deepfake video detection using recurrent neural networks. In 15th *IEEE International Conference on Advanced Video and Signal based Surveillance (AVSS)*, pages 1–6. IEEE, 2018.
[22] Ivan Laptev, Marcin Marszalek, Cordelia Schmid, and Benjamin Rozenfeld. Learning realistic human actions from movies. In *IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2008.
[23] Yuezun Li, Ming-Ching Chang, and Siwei Lyu. In ictu oculi: Exposing ai created fake videos by detecting eye blinking. In 2018 *IEEE International Workshop on Information Forensics and Security* (WIFS), pages 1–7. IEEE, 2018.
[24] Jeffrey Donahue, Lisa Anne Hendricks, Sergio Guadarrama, Marcus Rohrbach, Subhashini Venugopalan, Kate Saenko, and Trevor Darrell. Long-term recurrent convolutional networks for visual recognition and description. In Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition*, pages 2625–2634, 2015.
[25] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014.
[26] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In Proceedings of the *IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016.
[27] Xin Yang, Yuezun Li, and Siwei Lyu. Exposing deep fakes using inconsistent head poses. In *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 8261–8265. IEEE, 2019.