Deep fake Detection using deep learning techniques: A Literature Review

Amala Mary*, Anitha Edison†
†Computer Vision Lab, College of Engineering Trivandrum, Kerala.
†Affiliated to APJ Abdul Kalam Technological University, Trivandrum, Kerala, India.
Email: *am77909183@gmail.com, †anithaedison@cet.ac.in

2023 International Conference on Control, Communication and Computing (ICCC) | 979-8-3503-3412-8/23/$31.00 ©2023 IEEE | DOI: 10.1109/ICCC57789.2023.10164881

***Abstract*—Deep learning is a sophisticated and adaptable technique that has found widespread use in fields such as natural language processing, machine learning, and computer vision. It is one of the most recent deep learning-powered applications to emerge. Deep fakes are altered, high-quality, realistic videos/images that have lately gained popularity. Many incredible uses of this technology are being investigated. Malicious uses of fake videos, such as fake news, celebrity pornographic videos, financial scams, and revenge porn are currently on the rise in the digital world. As a result, celebrities, politicians, and other well-known persons are particularly vulnerable to the Deep fake detection challenge. Numerous research has been undertaken in recent years to understand how deep fakes function and many deep learning-based algorithms to detect deep fake videos or pictures have been presented.**

This study comprehensively evaluates deep fake production and detection technologies based on several deep learning algorithms. In addition, the limits of current approaches and the availability of databases in society will be discussed. A deep fake detection system that is both precise and automatic. Given the ease with which deep fake videos/images may be generated and shared, the lack of an effective deep fake detection system creates a serious problem for the world. However, there have been various attempts to address this issue, and deep learning-related solutions outperform traditional approaches.

***Index Terms*—-Deep Fakes, Deep Learning, Fake Generation, Fake Detection, Machine Learning**

I. Introduction

The face is the most unique feature of human beings. With the rapid advancement of face synthesis technology, the security risk provided by face alteration is becoming increasingly significant. Deep fake is one of the artificial intelligence technology in which one person's face is superimposed on top of another person's face without his/her permission.

Deep learning is a powerful and valuable technology that has been applied in many fields, including machine learning, computer vision, and natural language processing. As a result of advancements in deep learning, modifying digital material, and creating synthetic content has become quite simple. Generative adversarial networks (GANs) [1] and deep learning algorithms are used to create fake images and videos that are difficult for humans to distinguish from the real ones. These are produced using enormous datasets, then those models are used to train on a dataset and create fictitious videos and pictures. In reality, the widespread availability of videos/images on social media might aid people in creating plausible rumors and false information that could lead to creating a negative impact on society.

According to recent research, deep fake videos and images are widely disseminated on social media. Deep fake video/image detection has therefore become increasingly crucial and important. Many deep learning approaches such as Recurrent Neural Network (RNN) [2], Convolutional Neural Network (CNN) [3], and Long short-term memory (LSTM) [4], [5] are proposed to detect deep fake videos/images. And this will bring up more research on this area.

This study focuses on deep fake detection algorithms that have previously been deployed. It primarily covers classic detection methods as well as deep Learning based methods such as CNN, RNN, and LSTM. The first section of the study provides a quick overview of deep fakes and their societal consequences. The evaluated overview of relevant studies is mentioned in Section II. Then Section III discusses several detection approaches and strategies, with an emphasis on conventional deep learning. Section IV provides an overview of recently published datasets. And Section V contains the conclusion part.

II. Related Works

A. Deep Learning

It is a machine learning approach similar to neural networks [6] [7] and refers to the use of several hidden units in a network. Its basic architecture, influenced by artificial networks, employs an unlimited number of hidden units of bound size. This is done to evoke additional information from the input data. The complexity of the trained data determines the number of hidden layers [6]. More complicated data need more hidden layers to generate accurate findings. In recent years, it has been successfully employed in a wide range of domains and it will continue to be used.

1) Convolutional Neural Network: CNN is the deep neural network architecture most commonly used. It has an input layer, an output layer, and one or more hidden layers, just like other neural networks. In CNN [3], the hidden layers first read the inputs from the first layer and then execute a convolution mathematical operation on the input values. In addition to matrix multiplication, CNN employs non-linearity activation methods such as Rectified Linear Units (RELU) and extra convolutional approaches such as pooling layers. To minimize the complexity of the data, pooling layers provide outputs using methods such as average pooling.

2) Recurrent Neural Network: It is another artificial neural network application that can learn characteristics from sequence data [2]. Basically, RNN is built from a variety of hidden layers, each with its own bias and weight. The connection between the nodes in an RNN-based direct cycle graph runs sequentially. By offering a recurrent hidden state that encapsulates time-scale dependencies, it can handle a temporal sequence.

3) Long Short-Term Memory: It is a sort of artificial RNN that manages long-term dependencies [4] [5]. The full data sequence may be learned using the feedback connections in LSTM. The input gate, forget gate and output gate make up the basic LSTM architecture. The cell state remembers the values from prior intervals and stores them. The input gate first selects the values that ought to be written into the cell state. The forget gate may logically select which information has to be forgotten by employing a sigmoid function. In which the information from the present moment should be taken into account in the following phase is decided by the output gate.

III. Deep fake Generation and Detection

The deep fake approach creates fake videos/images using GANs techniques. In this part, first, provide a summary of the current software and resources for producing deep-fake images/videos. And then go through different deep-learning detection methods to address this problem.

A. Deep fake Generation

GANs are deep neural network that is frequently used to create deep fakes. GANs have the benefit of learning from a group of training data sets and producing a sample of information with the same characteristics and qualities. For example, GANs might be used to substitute a "genuine" picture or video of a person with an "altered one" [8]. GAN architecture consists of two neural network components: an encoder and a decoder. The model uses the encoder to train on a vast data set in order to generate fictitious data. The decoder is then used to discriminate between real and fake data. This model requires a significant amount of input in order to create realistic-looking faces (images and videos). The GAN architecture is depicted in Figure 1. To produce a fake sample, the encoder is first fed random input seeds, as seen in the image. The decoder is trained using these fake samples. This decoder is a binary classifier that takes in both real and fake data and then uses a Softmax function to differentiate between them.

Numerous deep fake applications have been used for a long time. The first technique that has been widely utilized for deep fake production is Fake App. This app uses an auto encoder-decoder pairing structure created by a Reddit user to swap faces in movies [9] [10]. Similar to GANs, Fake App employs an autoencoder to extract latent features from images of human faces and a decoder to re-extract those features from the same images. This method is effective because it can create fake videos that are surprisingly real and challenging to distinguish apart from the real thing. Another well-known deep fake method is based on a generative adversarial network and that is called VGGFace (GAN). The structure of VGGFace [11] is enhanced by the addition of two layers known as adversarial loss and perceptual loss. In order to capture hidden features of facial images these layers are added to the auto encoder-decoder. This produces more convincing and realistic fake images.

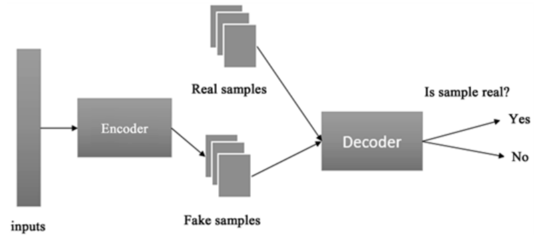

Fig. 1. Basic architecture of GAN [1]

B. Deep fake Detection

Deep learning has shown considerable achievement in the identification of deep fakes. In the section below first go through the deep learning-based image detection models then continue on to the video detection models.

1) Image Detection Methods: Several techniques have been explored to recognize GAN-generated pictures using deep networks. Tariq *et al*. [12] proposed using neural networks to detect false GAN videos. This methodology analyzes image statistical components and enhances recognition of artificially produced fake facial photos. Nhu *et al*. [13] contribute another strategy for identifying fraudulent images generated by GANs that is based on a deep CNN. This technique starts by utilizing a deep learning network to evoke facial attributes from face identification networks.

Xuan *et al*. [14] employed image preprocessing techniques, such as Gaussian blur and Gaussian noise. This improves the mathematical similarity between authentic photos and imitations at the pixel level, allowing the scientific classifier to pick up more intrinsic features, and improves generalization capacity than earlier techniques for picture forensics [15] [16].

Zhao *et al*. [17] recently introduced a methodology for deep fake detection utilizing the self-consistency of local source features, which are spatially-local, content-independent details of pictures. A CNN model employs a unique representation learning approach to extract these source features, which are represented as down-sampled feature maps referred to as pairwise self-consistency learning. This aims to punish feature vector pairings that correspond to areas in the same picture with poor cosine similarity scores. When dealing with false pictures created by technologies that output the entire image directly and whose source features are constant throughout each point inside each image, it could have a disadvantage.

2) Video Detection Methods: Due to the huge loss of frame content during video compression, existing deep learning algorithms for image identification cannot effectively detect bogus videos.

The severe deterioration of the frame data following video compression prevents the majority of image recognition techniques from being employed for videos [18]. Additionally, videos provide a problem for techniques intended to identify only still fake images since their temporal features vary across sets of frames. Based on the discovery that temporal coherence is not properly preserved in the synthesis process of deep fakes, Sabir *et al.* [19] used spatiotemporal characteristics of video streams to detect deep fakes. Frame-by-frame editing is used in video editing. A framework on which low-level face manipulation defects are expected to further appear as temporal distortions with irregularities between the frames.

However, deep learning algorithms frequently employ face photos from the internet that typically display people with wide eyes; fewer pictures of persons with closed eyes may be seen online. As a result, deep fake algorithms are unable to generate fake faces that blink often in the absence of photographs of actual people doing so. Deep fakes, in other words, have far lower blink rates than regular videos.

Li *et al.* [20] trim eye regions from the films and distribute them to long-term recurrent convolutional networks (LRCN) [21] for dynamic state prediction in order to distinguish between authentic and fake videos.

A deep learning technique that is used to detect deep fakes was presented in [22]. The UADFV and Deep fakeTIMIT deep fake datasets are used to assess the proposed approach. The total number of frames in the UADFV database [23] is 32,752. It consists of 49 authentic videos and 49 fraudulent videos. The suggested strategy eliminates the necessity for deep fake videos to be created as negative examples before training the detection algorithms. Instead, the negative instances are created dynamically by deleting the face region from the original picture, applying Gaussian blur to a scaled image of a random choice, and then stretching back to the original image after numerous scale alignments. Compared to previous approaches that call for the creation of deep fakes in advance, this requires a significant reduction in time and computing resources.

IV. Different types of Datasets

Next, we discuss the databases used for the identification of deep fakes.

A. Fake Face Dataset (DFFD)

It includes 100,000 and 200,000 false photos that taken from ProGAN and StyleGAN models. The majority of the samples in the collection are between the ages of 21 and 50, with around 47.7% of the photos being male and 52.3% being female.

B. VGGFace2

The large-scale face dataset is known as VGGFace2 consists of 3 million face photographs of nine thousand unique individuals, with an average of more than 300 photos per subject. The Google search engine provided images, which have a lot of information like age, race, lighting, and occupation.

C. Flickr-Faces-HQ (FFHQ)

This database contains human face information. And the database FFHQ contains 70,000 face images with a high-quality resolution created by GAN. The author claims that the dataset underwent pre-processing to reduce the size of the collection and remove picture noise.

D. 100K-Faces

A well-known publicly accessible dataset called 100K-Faces contains 100,000 original human photos created with StyleGAN. StyleGAN was used to create photographs with a flat backdrop from a big dataset of more than 29,000 images collected from 69 distinct models.

V. Commonly used Evaluation Parameters

The effectiveness of various deep fake detection approaches is evaluated using the metrics listed below. They are:

1) Accuracy:
The most basic metric for assessing a classification model's performance is accuracy. According to the given equation, classification accuracy is calculated by dividing the number of true results by the total number of results.

$$Accuracy = \frac{T_p + T_n}{T_p + T_n + F_p + F_n} \times 100 \quad (1)$$

where T_p, T_n, F_p, and F_n represent the true positive, true negative, false positive, and false negative rates.

2) Area Under Curve:
It gives a total performance evaluation across all potential classification thresholds.

3) Precision:
The following equation, which reflects the ratio of the total number of positive items divided by the number of true positives, is used to calculate accuracy during the classification step.

$$Precision = \frac{T_p}{T_p + F_n} \times 100 \quad (2)$$

4) Receiver Operating Characteristics(ROC):
The recall values are represented on the y-axis, while the specificity values are plotted on the x-axis is called ROC.

TABLE I
OVERVIEW OF POPULAR DEEP FAKE DETECTION METHODS

SL.No.	Authors	Methodology	Techniques	Key Features	Databases Used
1	Li, Yuezun, Ming-Ching Chang, and Siwei Lyu. [20]	Eye blinking	Long term recurrent CNN	Use LRCN to understand the temporal patterns of eye blinking.	Consist of 49 interview & presentation videos, & their corresponding generated deepfakes.
2	Afchar, Darius *et al*. [18]	MesoNet	CNN	Two deep networks Mes0-4 & Mesoinception-4 are introduced to examine deep false videos at the mesoscopic level of analysis.	Two databases: Deep fake one constituted from online videos & Face Forensics one created by face2face approach
3	Sabir, Ekraam *et al*. [19]	Spatio-temporal features with RCN	RCN	RCN, which combines the convolutional network DenseNet with the gated recurrent unit cells, is used to investigate temporal differences between frames.	FaceForensics++ dataset, including 1000 videos
4	Chintha, Akash *et al*. [9]	Spatio-temporal features with LSTM	convolutional bidirectional recurrent LSTM network	For face feature extraction, an XceptionNet CNN is employed, and audio embeddings are generated by stacking numerous convolution modules.	FaceForensics++, Celeb-DF & ASVSpoof 2019 logical Access audio dataset
5	Fernandes, Steven *et al*. [24]	Using attribution based confidence (ABC) metric	ResNet50 model pretrained on VGGFace2	Without access to training data, deep fake videos are detected using the ABC measure.	VidTIMIT, COHFACE
6	Ciftci, Umur Aybars, Ilke Demir, and Lijun Yin. [25]	FakeCatcher	CNN	Biological signals are not spatially and temporally well maintained in deep fakes, they are extracted from portrait videos and used as an implicit descriptor of authenticity.	UADFV, FaceForensics, FaceForensics++, Celeb-DF
7	Yang, Xin, Yuezun Li, and Siwei Lyu. [23]	Head poses	Support Vector Machine	Features are evoked using 68 landmarks of the face region.	UADFV is made up of 49 false videos and their actual videos, as well as 241 real photos & 252 fake images from the DARPA MediFor GAN Video/image Challenge.
8	Xuan, Xinsheng *et al*. [14]	Preprocessing combined with deep network	DCGAN, WGAN-GP & PGGAN.	Improve deep learning models' generalisation capacity to recognize GAN produced images.	Real dataset: CelebA-HQ, Fake datasets: generated by DCGAN, WGANGP and PGGAN
9	Güera, David, and Edward J. Delp. [26]	Passed through CNN for feature extraction	CNN	A model which automatically detect deep fake videos using recurrent neural network	HOHA dataset & deep fake videos from multiple hosting websites
10	Wang, Sheng-Yu *et al*. [27]	Using common artifacts of CNN-generated images	ResNet-50 pre-trained with ImageNet	Using a large number of fictitious pictures produced by a high-performing unconditional GAN model, i.e., PGGAN, train the classifier and assess how well it generalizes to other CNN-generated images.	A new dataset of CNN-generated images, namely ForenSynths, consisting of synthesized images from 11 models such as StyleGAN, super-resolution methods & FaceForensics++
11	Gandhi, Apurva, and Shomik Jain [28]	Defenses against adversarial perturbations in deep fakes	VGG & ResNet	Introduce adversarial perturbations to enhance deep fakes and fool deep fake detectors.	5,000 real images from CelebA & 5,000 fake images generated by the "Few-Shot Face Translation GAN" method
12	Li, Lingzhi *et al*. [29]	Face X-ray	CNN	Instead of collecting the synthesized artefacts of certain operations, try to find the border between the target and original faces.	FaceForensics++, DeepfakeDetection (DFD), DFDC & Celeb-DF

SL.No.	Authors	Methods	Classifiers/Techniques	Key Features	Datasets Used
13	Zhang, Ying, Lilei Zheng, and Vrizlynn LL Thing. [30]	Bag of words & shallow classifiers	SVM, RF, MLP	Using the bag of words approach, extract discriminant characteristics and input these into SVM, RF, and MLP for binary classification: real versus fake.	The popular LFW face database, which has 13,223 photos at a resolution of 250x250.
14	Chen, Zehao, and Hua Yang. [31]	Manipulted Face detector	CNN	Based on the Multilevel Facial Conceptual Extraction and Cascade Attention Mechanism, face identification methodology is modified.	FF, Celeb-A, FF++
15	Durall, Ricard *et al*. [32]	Unmasking Deepfakes	SVM, LR, k-MN	Method is based on a traditional frequency domain analysis, which is followed by a basic classifier.	Celeb-A, FF++
16	Li, Xiaodan *et al*. [33]	Sharp Multi-Instance Learning	MIL	In contrast to the classic MIL, which creates a straight mapping from instance embeddings to bag prediction before moving on to instance prediction, a sharp MIL (S-MIL) is presented.	Celeb-DF, FF, FF+, DFDC
17	Rana, Md Shohel, and Andrew H. Sung. [34]	DeepFake Stack	CNN	For identifying such altered videos, use the DeepfakeStack deep ensemble learning approach.	Celeb-DF, FF++
18	Do, Nhu-Tai, In-Seop Na, and Soo-Hyung Kim. [13]	Forensics Face detection	CNN	A deep CNN to detect forensics face.	Celeb-A
19	Nguyen, Huy H *et al*. [35]	Multi-task Learning	CNN	CNN that use the multi-task learning strategy to find the altered areas for each query while also simultaneously detecting altered pictures and videos.	FF, FF++
20	Guo, Zhiqing *et al*. [36]	Using deep features extracted by CNN	A new CNN model, namely SC net	High-level forensics characteristics may be automatically learned by CNN-based SCnet from visual input due to a hierarchical feature extraction block constructed by stacking four convolutional layers.	Glow model was used to produce a collection of 321,378 face photos from the CelebA face image dataset.

5) F1-score:
It represents the arithmetic mean of accuracy and recall.

$$F1 = 2 \times \frac{Precision \times Recall}{Precision + Recall} \times 100 \qquad (3)$$

VI. Conclusion

Various researchers have created a number of deep-learning approaches for deep fake images and videos. Due to the extensive availability of photographs and videos in social media material, deep fakes had grown in popularity. This is especially crucial in social networking sites that make it simple for users to spread and share such fake information. Numerous deep learning-based approaches have recently been put out to deal with this problem and effectively identify fake images and videos. The first section discussed the existing programs and technologies that are extensively used to make fake photos and videos. And in the second section discuss the different type of techniques that are used for deep fake images and videos. Also, provide details of available datasets and evaluation metrics that are used for deep fake detection.

Despite the fact that deep learning has done well in detecting deep fakes, the quality of deep fakes has been increasing. In order to recognize fake videos & photos properly must be enhanced current deep learning approaches. Furthermore, given present deep learning approaches, it is unknown how to identify the number of layers necessary and the appropriate architecture for deep fake detection. To improve their capacity to cope with the ubiquitous impacts of deep fakes and mitigate their consequences, social media companies are integrating deep fake detection tools.

References

[1] M. Mirza and S. Osindero, "Conditional generative adversarial nets," *arXiv preprint arXiv:1411.1784*, 2014.

[2] Y. Bengio, P. Simard, and P. Frasconi, "Long short-term memory," *IEEE Trans. Neural Netw*, vol. 5, pp. 157–166, 1994.

[3] I. Goodfellow, Y. Bengio, and A. Courville, *Deep learning*. MIT press, 2016.

[4] S. Hochreiter, "Ja1 4 rgen schmidhuber (1997)."long short-term memory"," *Neural Computation*, vol. 9, no. 8.

[5] M. Schuster and K. Paliwal, "Networks bidirectional reccurent neural," *IEEE Trans Signal Proces*, vol. 45, pp. 2673–2681, 1997.

[6] J. Hopfield *et al.*, "Rigorous bounds on the storage capacity of the dilute hopfield model," *Proceedings of the National Academy of Sciences*, vol. 79, pp. 2554–2558, 1982.

[7] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, *et al.*, "Google's neural machine translation system: Bridging the gap between human and machine translation," *arXiv preprint arXiv:1609.08144*, 2016.

[8] L. Nataraj, T. M. Mohammed, B. Manjunath, S. Chandrasekaran, A. Flenner, J. H. Bappy, and A. K. Roy-Chowdhury, "Detecting gan generated fake images using co-occurrence matrices," *Electronic Imaging*, vol. 2019, no. 5, pp. 532–1, 2019.

[9] B. Zi, M. Chang, J. Chen, X. Ma, and Y.-G. Jiang, "Wilddeepfake: A challenging real-world dataset for deepfake detection," in *Proceedings of the 28th ACM international conference on multimedia*, 2020, pp. 2382–2390.

[10] H. A. Khalil and S. A. Maged, "Deepfakes creation and detection using deep learning," in *2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC)*. IEEE, 2021, pp. 1–4.

[11] J. Luttrell, Z. Zhou, Y. Zhang, C. Zhang, P. Gong, B. Yang, and R. Li, "A deep transfer learning approach to fine-tuning facial recognition models," in *2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. IEEE, 2018, pp. 2671–2676.

[12] S. Tariq, S. Lee, H. Kim, Y. Shin, and S. S. Woo, "Detecting both machine and human created fake face images in the wild," in *Proceedings of the 2nd international workshop on multimedia privacy and security*, 2018, pp. 81–87.

[13] N.-T. Do, I.-S. Na, and S.-H. Kim, "Forensics face detection from gans using convolutional neural network," *ISITC*, vol. 2018, pp. 376–379, 2018.

[14] X. Xuan, B. Peng, W. Wang, and J. Dong, "On the generalization of gan image forensics," in *Chinese conference on biometric recognition*. Springer, 2019, pp. 134–141.

[15] P. Yang, R. Ni, and Y. Zhao, "Recapture image forensics based on laplacian convolutional neural networks," in *International Workshop on Digital Watermarking*. Springer, 2016, pp. 119–128.

[16] B. Bayar and M. 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*, 2016, pp. 5–10.

[17] T. Zhao, X. Xu, M. Xu, H. Ding, Y. Xiong, and W. Xia, "Learning self-consistency for deepfake detection," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 15 023–15 033.

[18] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: a compact facial video forgery detection network," in *2018 IEEE international workshop on information forensics and security (WIFS)*. IEEE, 2018, pp. 1–7.

[19] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan, "Recurrent convolutional strategies for face manipulation detection in videos," *Interfaces (GUI)*, vol. 3, no. 1, pp. 80–87, 2019.

[20] Y. Li, M.-C. Chang, and S. Lyu, "In ictu oculi: Exposing ai created fake videos by detecting eye blinking," in *2018 IEEE International workshop on information forensics and security (WIFS)*. IEEE, 2018, pp. 1–7.

[21] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 2625–2634.

[22] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," *arXiv preprint arXiv:1811.00656*, 2018.

[23] X. Yang, Y. Li, and S. Lyu, "Exposing deep fakes using inconsistent head poses," in *ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2019, pp. 8261–8265.

[24] S. Fernandes, S. Raj, R. Ewetz, J. S. Pannu, S. K. Jha, E. Ortiz, I. Vintila, and M. Salter, "Detecting deepfake videos using attribution-based confidence metric," in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2020, pp. 308–309.

[25] U. A. Ciftci, I. Demir, and L. Yin, "Fakecatcher: Detection of synthetic portrait videos using biological signals," *IEEE transactions on pattern analysis and machine intelligence*, 2020.

[26] D. Güera and E. J. Delp, "Deepfake video detection using recurrent neural networks," in *2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS)*. IEEE, 2018, pp. 1–6.

[27] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "Cnn-generated images are surprisingly easy to spot... for now," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 8695–8704.

[28] A. Gandhi and S. Jain, "Adversarial perturbations fool deepfake detectors," in *2020 International joint conference on neural networks (IJCNN)*. IEEE, 2020, pp. 1–8.

[29] L. Li, J. Bao, T. Zhang, H. Yang, D. Chen, F. Wen, and B. Guo, "Face x-ray for more general face forgery detection," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2020, pp. 5001–5010.

[30] H. Cui, A. Liu, X. Zhang, X. Chen, K. Wang, and X. Chen, "EEG-based emotion recognition using an end-to-end regional-asymmetric convolutional neural network," *Knowledge-Based Systems*, vol. 205, p. 106243, 2020.

[31] Z. Chen and H. Yang, "Attentive semantic exploring for manipulated face detection," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 1985–1989.

[32] R. Durall, M. Keuper, F.-J. Pfreundt, and J. Keuper, "Unmasking deepfakes with simple features," *arXiv preprint arXiv:1911.00686*, 2019.

[33] X. Li, Y. Lang, Y. Chen, X. Mao, Y. He, S. Wang, H. Xue, and Q. Lu, "Sharp multiple instance learning for deepfake video detection," in *Proceedings of the 28th ACM international conference on multimedia*, 2020, pp. 1864–1872.

[34] M. S. Rana and A. H. Sung, "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)*. IEEE, 2020, pp. 70–75.

[35] H. H. Nguyen, F. Fang, J. Yamagishi, and I. Echizen, "Multi-task learning for detecting and segmenting manipulated facial images and videos," in *2019 IEEE 10th International Conference on Biometrics Theory, Applications and Systems (BTAS)*. IEEE, 2019, pp. 1–8.

[36] Z. Guo, L. Hu, M. Xia, and G. Yang, "Blind detection of glow-based facial forgery," *Multimedia Tools and Applications*, vol. 80, no. 5, pp. 7687–7710, 2021.