2023 2nd International Conference on Vision Towards Emerging Trends in Communication and Networking Technologies (ViTECoN) | 979-8-3503-4798-2/23/$31.00 ©2023 IEEE | DOI: 10.1109/ViTECON58111.2023.10157962

Deepfake Creation Using Gans and Autoencoder and Deepfake detection

Manoj kumar Das
dept.of computer engineering
delhi technological university
delhi,india
manoj5das3@gmail.com

Manav kumar
dept.of computer engineering
delhi technological university
delhi,india
kumarmanav245@gmail.com

Ishank kumar kapil
dept.of computer engineering
delhi technological university
delhi,india
ikapil15@gmail.com

Dr Rajesh kumar yadav
dept.of computer engineering
delhi technological university
delhi,india
rkyadav@dtu.ac.in

***Abstract*— Deep learning techniques is useful in a myriad of applications such as computer vision processing, natural language processing, as well as deepfake detection. The development of deep learning algorithms for imaging detection has resulted in the development of deepfakes. These fakes employ advanced algorithms for deep learning to generate fake images that are extremely difficult to differentiate from authentic images. In a fake video, the face, expression or speech is replaced with an image of another's face, with a distinct speech or emotion with the help of the technology of deep learning. These videos are typically so sophisticated that the traces of manipulation are hard to spot. Social media are among the most frequently targeted and most serious because they are vulnerable platforms that are susceptible to blackmailing or making a person look bad. There are several existing efforts to detect fake images, but there have been very few efforts developed for video content on social media.**

Keywords— deep-fake, deep learning, GANS, autoencoder.

I. INTRODUCTION

Machine vision is expanding every day across a broad range of fields, ranging from basic computer software that recognizes images to automated robotics. One of the numerous applications based on machine vision is known as Deep-fake[1]. Deep-fake makes use of deep-learning algorithms to build counterfeit images by switching faces of an original image to make an image of another. This results in fake images that are difficult to distinguish. The reason behind this is due to the use of decoders as well as encoders that are based on deep-learning which are extensively used in the area of machine-vision [2].Deepfakes are created by employing machine learning algorithms to change or swap out certain parts of an original video or image, like a person's face.Recent development in this field allows to create deepfake with just a motionless image[3]. Deep fake detection refers to the difficulty of spotting false movies or photographs made with deep learning methods. To detect these modifications and distinguish them from real movies or photographs, deepfake detection is used. We propose to use our technology to the detection of false faces in photographs at a macroscopic degree of analysis. The human vision struggles to distinguish phony images at higher semantic levels, especially when they include a person's face or video. We used a deep neural network with a defined set of layers as a middle ground approach . The architecture that come next had the decent classification results out of all of our tests since they have a low degree of representation and a lower degree of parameters. They are based on powerful image classification networks that switch between convolutional layer and pooling to extract features and a dense network to categorize the images. [4]. Major tech companies are seeking ways to identify fake and false news to end the growing amount of fake news being published on the internet. Recently many big companies like Meta, Amazon.com, Inc , Microsoft corporation are investing in deepfake detection contest to promote more exploration in this field [5] .The enthusiasm displayed by big names such as Google and Microsoft highlights the significance of the Deep-fake problem.The paper highlights the possible options of deepfake generation. Motivation behind the research is the rising risk of manipulated images and videos that poses the threat on the privacy of the people .There is a increasing threat to world peace as deepfake can be used manipulate speeches and videos of global leaders[10-12].The advantage of the proposed methodology is that it is efficient in classifying fake and real images .The Fig. 1 depicts the genuine fake and real images created using advanced technology and a network structure that has the capability to use a lot of data for developing the network. .

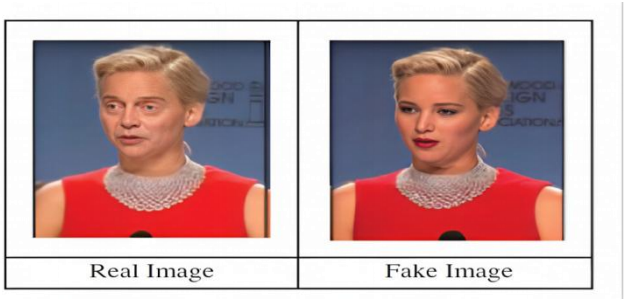

Fig. 1. Real and Deep-fake image.

979-8-3503-4798-2/23/$31.00 ©2023 IEEE

II. LITERATURE REVIEWS

The paper article[6]: Profound phony Video Location Utilizing Intermittent Brain Organization**,** David Guera and Edward J Delp propose the idea of a fleeting cognizant framework which can perceive counterfeit Profound phony video. To distinguish counterfeit recordings, first we should comprehend the system of their creation and assist us with perceiving the imperfections in Profound phony creation to take advantage of the shortcomings, a profound phony recognition can be made. The strategy talked about in this article, outline level irregularity is the primary perspective which is utilized to recognize fakes. If the encoder does not know about the skin's or different points of interest of the scene, there could be limit impacts because of unconstrained converging of the face and the casing, which could be an alternate fault. One more issue that is utilized in this situation is that it makes a few irregularities, and can cause the presence of flashing in the facial region. This is an element that is normal that is available in by far most of deceitful recordings. It's difficult to distinguish outwardly, it is feasible to record utilizing a Pixel-level CNN include extraction. The dataset utilized in this study contains 3000 recordings from the HOHA dataset. The pre-handling steps are made sense of in this article. The proposed framework is made out of a LSTM-based convolution structure used to handle outlines. CNN to eliminate highlights from outlines with LSTM to concentrate on the transient succession are two key components that make up the convolution LSTM. In case of an unseen test arrangement, the highlights of each edge are made by CNN. From that point forward, highlights from various casings are consolidated and afterward shipped off LSTM to be dissected. LSTM then, at that point, gives a gauge of whether the succession is phony or has not been modified at all. In under 2 seconds, the framework can distinguish in the event that the video is being examined is phony film or not. The exactness is more noteworthy than 97%.

In the paper[7] : Powerful and Quick Profound phony discovery strategy in view of Haarwavelet Changeby Mohammed Akram Younus and TahaMohammed Hasan, the creators present an elective technique for recognizing counterfeit recordings by utilizing the wavelet change. The procedure portrayed in this paper exploits the way that when the Profound phony calculation produces counterfeit video it can create counterfeit countenances that is features with specific size and goal. To ensure that they match the plan of the face that was initially made on the real video, there is a haze highlight is incorporated into the phony countenances. This change makes a particular haze irregularity between the created face and the result of the foundation counterfeit recordings. The calculation can identify this irregularity by contrasting the obscured locales of the created return for money invested in contrast with the encompassing region by utilizing a specific Haar Wavelet change highlight. The chief advantage of utilizing the Haar Wavelet change is that it at first separates the various types of edges and afterward kills the obscured picture's lucidity. Extremely productive and quick as the foundation isn't changing countenances in the pictures won't be impacted and there's no necessity to copy this obscure lattice capability. To survey the haze's size, two strategies like immediate and deviant can be utilized. Direct strategy can be utilized to decide the size of the haze by looking at unmistakable highlights in the image. For example, edge highlights. The backhanded strategy depends on obscure reproduction in case of a H network which isn't distinguished (the H-framework is described by obscure's appraisals , as well as obscure acknowledgment). Dirac Construction step structures, Dirac structure and the rooftop are three particular sorts of edges that show up in a picture. The haze reaches out by taking the nature of the G design of the moves toward be thought about. Sharpness on edges is estimated in the number (0! 2!) that, if more noteworthy, proposes that the picture is more fine. At the point when we think about the haze term of the return for capital invested as well as the haze size of the remainder of the picture, we can decide if obviously the images(frames inside video) were changed or have not been modified. The UADFV dataset that includes 49 non manipulated also with 49 recordings that have been adjusted, is utilized to play out this. The recordings are isolated into outlines. In each edge, the face is taken . Then, a Profound phony location calculation utilizing haar wavelet change used. The calculation is made sense of top to bottom in this article. The calculation is proposed to have an exactness that is 90.5 percent.

The paper [8] OC Counterfeit Dect: Ordering Profound phony s utilizing a One Class Variational Autoencoder composed by Hasam Khalid, and Simon S. Charm, the model depends on genuine pictures utilized for preparing. Since new strategies for making misleading recordings are being developed because of mechanical advances to assist a model with distinguishing fake recordings. information that contain counterfeit recordings are not normal for preparing. This effects on the accuracy of the model. The model introduced in this paper involves just genuine video film for preparing to beat the constraints of information lack. The Face Forensic++ dataset is the dataset utilized in this paper. It is made out of genuine pictures and five phony sets The Face-Trade dataset F2Fdataset (F2F) Profound phony dataset(DF) Brain Text-tures dataset (NT) Profound phony recognition Dataset(DFD) Subsequent to gathering the video information, the information is changed into outlines and the face discovery and arrangement process is finished utilizing MTCNN. One-class variational encoders are utilized in this occurrence. It comprises of an autoencoder. In the encoder's case, pictures are utilized as information. Scaling is finished through convolution layers alongside the mean and change are determined. The outcomes are then used for input into the decoder. The RMSE esteem is determined, which is lower for real pictures, however higher when phony pictures are used. Two strategies are examined in this article: OCFakeDect1andOCFakeDect2. In OCFakeDect1 with result and information pictures the score of remaking is determined progressively. OCFakeDect2 likewise has an extra encoder structure that computes scores of their making in view of the information and result dormant data. It can arrive at 97.5 rate precision most productive variant is just for NT and DFD datasets.

In the paper [9] Computerized Legal sciences and Examination of Profound phony Recordings by Mousa Tayseer Jafar Muhammed AbabnehMuhammad Al-Zoube, Ammar Elhassan proposed a technique to perceive fakes by utilizing elements of the mouth. These days, counterfeit recordings can actually hurt society, and can harm the validity of a person. "Profound phony" is a term used to depict a phony "profound phony" alludes to a video made to cause it to seem like somebody is saying or accomplishing something that are not really what they have said or done.

For this reason there is an expansion popular for techniques to detect fakes. The model utilizes elements of the mouth can be utilized to distinguish counterfeit recordings. A profound phony identification model with mouth features(DFT-MF),using profound learning way to deal with recognize Profound phony recordings by segregating dissecting and checking lip/mouth development is planned and carried out here. The information here is a combination of genuine and counterfeit recordings. A pre-handling process is completed before directing an examination. From that point forward, the mouth segment is taken out from the picture. The proper directions that are utilized for face. While working with a picture outline facial milestone identifier is utilized to decide the specific area in those 68 (X,Y)coordinates. Following that, eliminate any countenances with shut mouths. Then, at that point, the main individual that has an open mouth is delegated having teeth of satisfactory lucidity.CNN decides if recordings are certified or counterfeit by deciding an edge number that is phony with regards to outlines.It involvs two elements for each sentence: the discourse rate and casing rate. Assuming the misleading casings is more prominent than 60, the video is tagged as genuine or fake.

III. PROPOSED METHODOLOGIES

A. Goals and Objectives

Our research aims to uncover the lies behind the real lies. Our plan will cut down on the falsehoods and misunderstandings of the public at large on the internet as a whole. Our project will identify and categorize the image as deepfake or the original.

B. Dataset

1.celebrity face dataset is used to generate deepfake images .

2.Real vs FAKE dataset is used for detection to classify images as fake or real. dataset consists of 140 thousand images which is toned down as per our needs.

TABLE I.

S.no	Images	Number of images	Label
1	Real	4257	1
2	Fake	2987	0

C. Proposed flowchart for image generation

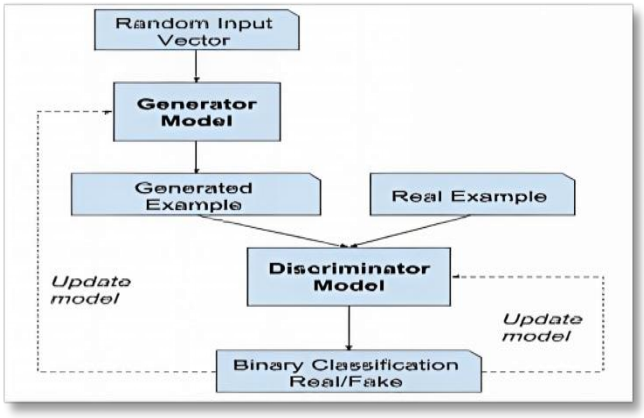

Fig. 2. The experiment's generators and discriminators.

In the network, the generative adversarial network (GANs) employs two directly competing neural networks -generator and discriminator. Generator creates new image based on the latent vector provided and on which the network is trained . discriminator discriminates whether the images is authentic or fake. The network will train and the generator will be trained to create images close to real ones so as to deceive the discriminator. The approach is to produce images that can misclassified by the discriminator [16].

Discriminator Network

The discriminator uses an image as input and then attempts to categorize the image either "real" or "generated". In this way it's similar to every other network. It's the convolution neural network (CNN) that outputs one number for each image.

Generator Network

Generator input is usually an image array of numbers or vectors (referred to as latent tensor) that is used as a seed in the creation of an image. The generator can convert an unstructured latent tensor (128 1, 1, 1) into an image-based tensor of shape 3 x 28 x 28.

Discriminator loss

This method measures the extent to which the discriminator can discern authentic and fake images. It examines the discriminator's forecasting for authentic images with an list of 1, and the discriminator's forecasting on false (fake) pictures to an list of zeros.

```
Def discriminator_loss(authentic_output,forged_output):
authentic_loss=cross_entropy(ones_like(authentic_output), authentic_output)
forged_loss=cross_entropy(zeros_like(forged_output), forged_output)
Sum_total_loss = authentic_loss + forged_loss
return sum_total_loss
```

Generator loss

The generator's misfortune is a proportion of the degree to which deceiving the discriminator was capable. By and large, in the event that the generator is functioning admirably and the discriminator can perceive those phony pictures as credible (or one).

```
Def generator_loss(forged_output):
return
cross_entropy(ones_like(forged_output),forged_output)
```

Initiation capability utilized

Beside the standard ReLU capability Cracked ReLU grants the death of a tiny slope sign to make up for negative

qualities. This upgrades the manner in which the slopes produced by the discriminator stream more grounded to the generator. Rather than passing a point (slant) of 0 during the back-prop the generator passes a minuscule negative inclination.

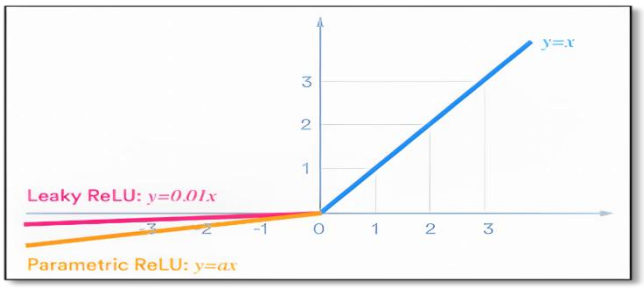

Fig. 3. different from leaky ReLU and parameter ReLU.

D. Deepfake Using Autoencoder

To make a Deepfake with the help of auto-encoder and decoder, two trained pair of auto-encoder and decoder is used such and after training the decoder are swapped so that latent feature of each encoder-decoder pair is used to generate deepfake images [13-14].

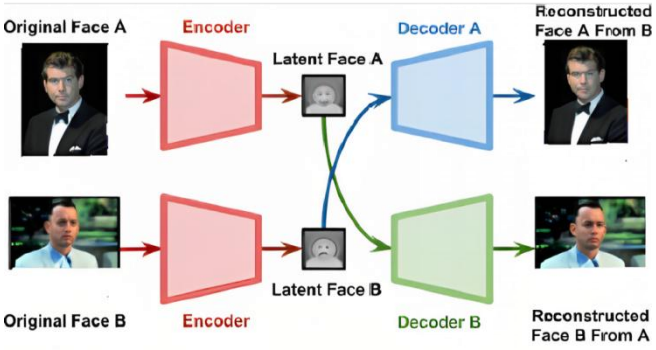

Fig. 4. encoder-decoder architecture.

At the time of training autoencoder is served with a pack of images, the encoder- decoder layer will tune the parameter to generate an output identical to input image as close as possible . The images used to train autoencoder is sourced from celebrity face dataset.

E. Image Detection Model

Many techniques have been explored to recognize GAN-generated images using deep networks. The use of pre-processing techniques in neural network-based technology improves the detection of fraudulent face images created by individuals and analyses the statistical characteristics of images. This also offers an alternative technique for identifying fake images created by deep convolutional neural networks that are based on GANs.The face features are extracted from face recognition network, for this the model used a deep learning network. Subsequently, facial features is adjusted by fine tuning phase so that they are acceptable for for real/fake image recognition.

A dense network with only one hidden layer follows a sequence of four convolutional and pooling layers in this network. Convolutional layers use the Batch Normalization and ReLU activation functions to regularize their output and prevent the vanishing gradient effect in order to improve conceptualization.Dropout is used to regularize and boost robustness in fully connected layers.An elective advancement contains in superseding the hidden two convolutional layers of the Model by an assortment of the start module.

The capability space in which the model can be improved can be expanded by stacking the results of a few convolutional layers with various piece shapes using the module.

We propose replacing the 5x5 convolutions in the original module with 3x3 dilated convolutions to avoid high semantics[15]. The concept of dealing with multi-scale information by employing dilated convolutions with the inception module can be found in. We have, however, added 1x1 convolutions prior to the dilated convolutions and 1x1 convolutions in parallel to serve as a skip-connection between subsequent modules in order to reduce the dimension.

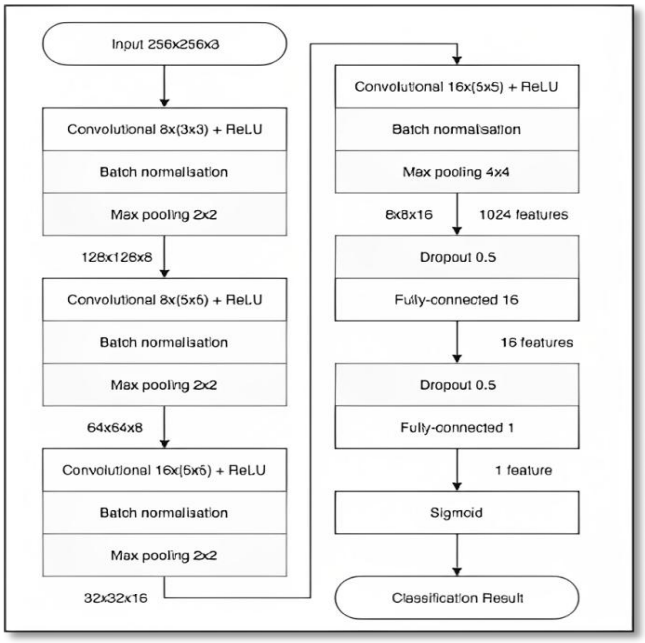

Fig. 5. The network architecture.

The architecture has been implemented with the help of python 3.10 and keras 2.11.0 module. Optimization of weights is done by batch of 75 images using ADAM [16]. Learning rate is taken as 10^{-3} which is reduced down to 10^{-6} by dividing the learning rate by 10 for successive 1000 iteration.

IV. Results Analysis

A. correct real prediction

Fig. 6. Real images from the dataset.

B. Misclassified real prediction

Fig. 7. Real images from the dataset.

C. correct deepfake prediction

Fig. 8. Fake images from the dataset.

D. Misclassified deepfake prediction

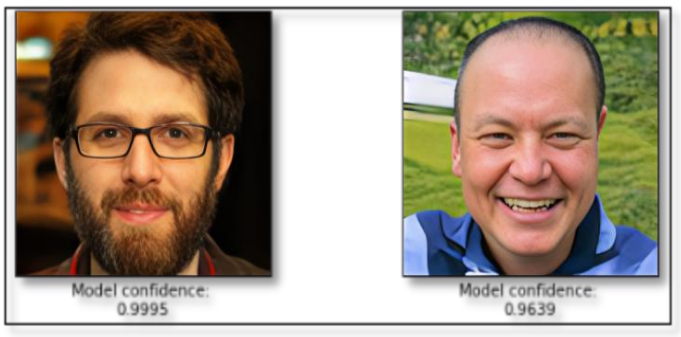

Fig. 9. Fake images from the dataset.

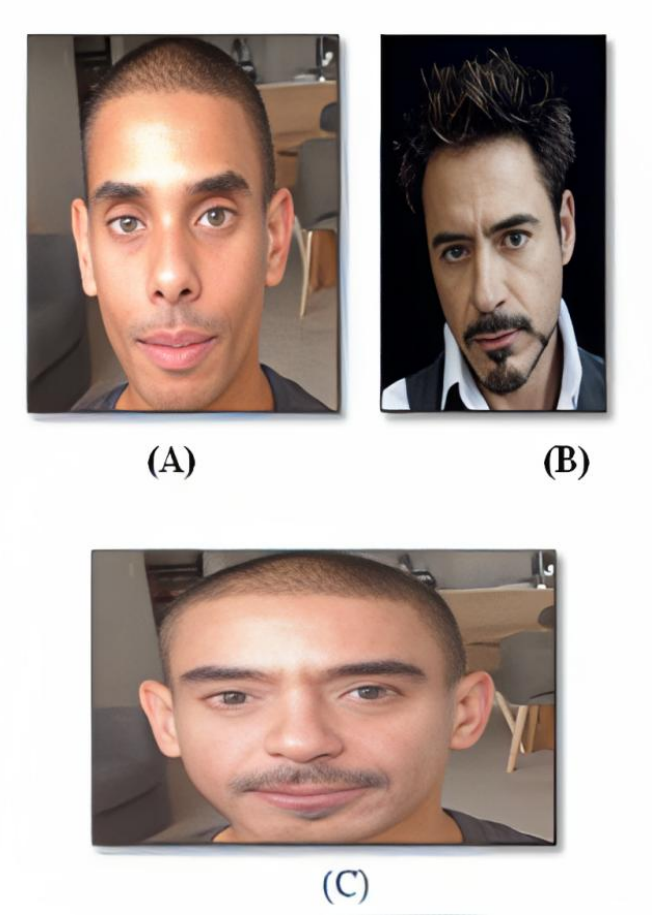

Fig. 10. Image (A) and (B) is real but (C) is deep-fake.

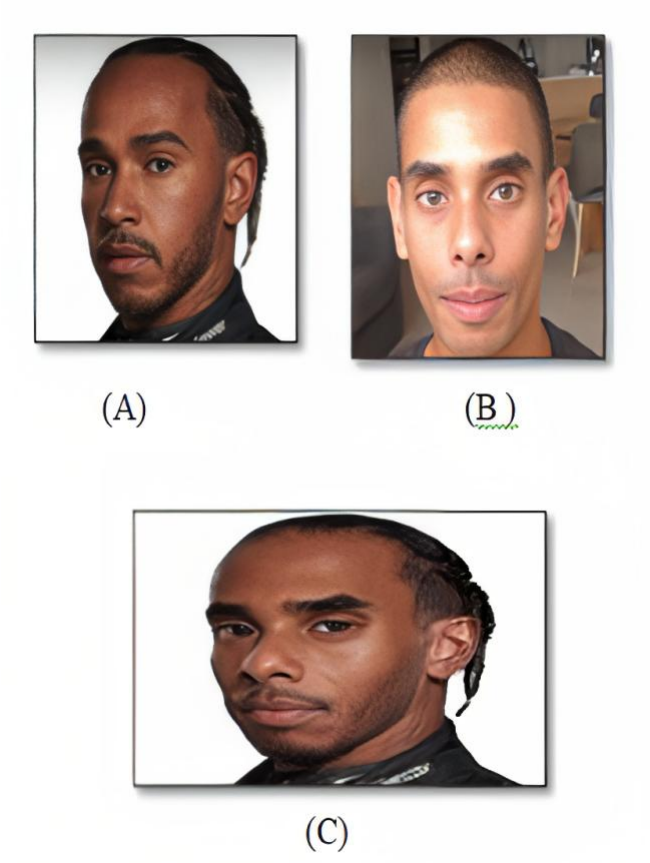

Fig. 11. Image (A) and (B) is real and (C) is deep-fake.

V. CONCLUSION

Because the viewing of them no longer corresponds with believing fake news, deep-fakes are beginning to undermine people's trust in media content. They can cause distress to those who are being targeted as well as amplify the spread of hate speech and disinformation or even increase the level of political discontent and incite public protests as well as violence or conflict. This is crucial today as social media platforms have the ability to quickly spread fake technology and news, and they are becoming more readily available.

This research offers a comprehensive review of the challenges that are affecting the future, as well as prospective trends and future perspectives in this field as in a comprehensive review of deep-fake production as well as detection methods. Artificial intelligence researchers can benefit greatly from the research in order to develop effective strategies to combat deep-fakes.

REFERENCES

[1] Ayush Tewari, Michael Zollhoefer, Florian Bernard, Pablo Garrido, Hyeongwoo Kim, Patrick Perez, and Christian Theobalt. High-fidelity monocular face reconstruction based on an unsupervised model-based face autoencoder. IEEE Transactions on Pattern Analysis and Machine Intelligence, 42 (2):357–370, 2018.

[2] Jiacheng Lin, Yang Li, and Guanci Yang. FPGAN: Face deidentification method with generative adversarial networks for social robots. Neural Networks, 133:132–147, 2021.

[3] Egor Zakharov, Aliaksandra Shysheya, Egor Burkov, and Victor Lempitsky. Few-shot adversarial learning of realistic neural talking head models. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 9459–9468, 2019.

[4] Bloomberg. How faking videos became easy and why that's so scary. https://fortune.com/2018/09/11/deepfakes-obama- video/, September 2018.

[5] Robert Chesney and Danielle Citron. Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. Foreign Affairs, 98:147, 2019.

[6] D. Guera¨ and E. J. Delp, "Deep-fake video detection using recurrent neural networks," in 201815th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2018, pp. 1–6.

[7] M. A. Younus and T. M. Hasan, "Effective and fast Deep-fake detection method based on haar wavelet transform," in 2020 International Conference on Computer Science and Software Engineering (CSASE), 2020, pp. 186–190.

[8] H. Khalid and S. S. Woo, "Oc-fakedect: Classifying Deep-fake s using one-class variational autoencoder," in 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 2020, pp. 2794–2803.

[9] U. Ciftci, I. Demir, and L. Yin, "How do the hearts of deep fakes beat? deep fake source detection via interpreting residuals with biological signals," 08 2020.

[10] Bloomberg. How faking videos became easy and why that's so scary. https://fortune.com/2018/09/11/deepfakes-obama-video/, September 2018.

[11] Robert Chesney and Danielle Citron. Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. Foreign Affairs, 98:147, 2019.

[12] T. Hwang. Deepfakes: A grounded threat assessment. Technical report, Centre for Security and Emerging Technologies, Georgetown University, 2020.

[13] Faceswap: Deepfakes software for all. https://github. com/deepfakes/faceswap.

[14] FakeApp 2.2.0. https://www.malavida.com/en/soft/ fakeapp/.

[15] F. Yu and V. Koltun. Multi-scale context aggregation by dilated convolutions. arXiv preprint arXiv:1511.07122, 2015.

[16] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. Advances in Neural Information Processing Systems, 27:2672–2680, 2014.