# A Study on Combating Emerging Threat of Deepfake Weaponization

**Conference Paper** · October 2020

DOI: 10.1109/I-SMAC49090.2020.9243588



**2 authors**, including:

Anushka Lal
Delhi Technological University
**5** PUBLICATIONS **29** CITATIONS

SEE PROFILE

Proceedings of the Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)
DVD Part Number:CFP20OSV-DVD; ISBN: 978-1-7281-5463-3

# A Study on Combating Emerging Threat of Deepfake Weaponization

Rahul Katarya
*Department of Computer Science and Engineering,*
*Delhi Technological University*
New Delhi, India
rahulkatarya@dtu.ac.in

Anushka Lal
*Department of Computer Science and Engineering*
*Delhi Technological University*
New Delhi, India
anushkalal_2k18co083@dtu.ac.in

**Abstract—A breakthrough in the emerging use of machine learning and deep learning is the concept of autoencoders and GAN (Generative Adversarial Networks), architectures that can generate believable synthetic content called deepfakes. The threat lies when these low-tech doctored images, videos, and audios blur the line between fake and genuine content and are used as weapons to cause damage to an unprecedented degree. This paper presents a survey of the underlying technology of deepfakes and methods proposed for their detection. Based on a detailed study of all the proposed models of detection, this paper presents SSTNet as the best model to date, that uses spatial, temporal, and steganalysis for detection. The threat posed by document and signature forgery, which is yet to be explored by researchers, has also been highlighted in this paper. This paper concludes with the discussion of research directions in this field and the development of more robust techniques to deal with the increasing threats surrounding deepfake technology.**

**Index Terms—Deep Learning, Generative Adversarial Networks, autoencoders, Deepfake detection, Fake image, Fake Video**

## I. INTRODUCTION

Deepfake technology is a new automatic computer graphics tool that portrays entirely unrealistic events as real through digital media manipulation. Deepfake gained its name from the Reddit platform where an anonymous user used this term, a combination of deep learning and fake, for replacing celebrities into adult video clips. As soon as the code was made public, widespread interest spawned in the users about the generation of fake content. ObamaNet [1], was an architecture that featured an impressive use of lip-syncing technology to generate synchronized photo-realistic lip-sync videos. Deepfake technique makes it possible to generate unauthentic videos of people expressing or saying things they have never said before [2].

Deepfake makes use of Artificial Intelligence (AI), machine learning, and deep learning concepts. AI deals with intelligence at the machine level.

Machine learning is an evolving concept of Computer Science where machines can be trained to learn from provided data and accordingly take decisions on their own, just like humans do. Deep learning is a broader aspect of machine learning, where highly complex networks are trained to learn from a massive database of unstructured data.

Today there are various free deepfake applications like the Chinese app, Zao [3] that lets users to easily swap faces with movie stars so they can see their self, playing that role in the movie, DeepNude [4], that can create nonconsensual porn, FakeApp, FaceSwap, and DeepFace Lab. The existence of such open-source software and the availability of devices in the market for fabricating and propagating these falsified information has brought to attention the immediate need for detection and elimination of malicious deepfake content. Deepfakes can act as a powerful weapon to insurgent groups and terrorist organizations, who may depict their adversaries using inflammatory words or engaging in provocative actions, to maximize the galvanizing impact on their target audiences. For instance, a member of the Islamic State (or ISIS), can falsely generate fake content that shows government officials or soldiers discussing bombing attacks at a mosque, to aid their terrorist group's recruitment [5]. States can use this weapon to undermine their non-state opponents. Deepfakes can affect the outcome of an election, hence a threat to democracy. Deepfake is even weaponizing satellite images of Earth by showing the existence of certain objects in landscapes and locations that do not exist in reality, just so they can play with the minds of military analysts and influence their decisions based on these fake images [6][7].

Amidst the threats posed by deepfakes in various sectors, the ability to generate realistic simulations can have a whole new positive impact on humanity. It can create an array of opportunities in fields of education, entertainment, and business. Historical figures can be made to communicate with students. In movies, face-swapping can be achieved for scenes

that cannot be fulfilled by the actors alone. For example, in 2016's Rogue One, late Peter Cushing's appearance as Grand Moff Tarkin was possible through similar technology [8]. Deepfakes can be used in business such that customers can exactly view their appearance in products they wish to buy without applying them in reality. In the medical world, this technology can play a great role in training doctors, nurses, and surgeons to operate on real-life scenarios in a virtual environment [9].

However, the potential of deepfakes to cause a broad spectrum of serious harm to society is a matter of greater concern. They can act as a new weapon to humiliation and destruction, identity theft and exploitation, defamation, and manipulation of legal evidence. Several methods have been proposed to detect deepfakes by focusing on the minute details of the content such as facial texture, head poses, eye-blinking, skin color, lip movements, Spatio-temporal features, and capsule forensics. Most of these rely on the same deep learning techniques that are used for the creation of deepfakes.

In the foreseeable future, deepfakes will continue evolving; thus, it is important to investigate their development and improve the methods of detection accordingly. The main objective of this paper is to present a survey of methods used for the creation and detection of deepfakes. Section II explains the popular underlying principle of deepfake architecture. Section III discusses and compares different proposed methods for deepfake detection. Section IV presents the contribution of this paper to the survey. The research opportunities in this direction are further highlighted in Section V.

## II. DEEPFAKE CREATION

Deepfakes use Deep Neural Networks (DNNs). DNNs consists of a set of interconnected units called neurons. These units together perform some form of computational task and help solve complex problems. Two popular technologies associated with deepfake creation are the autoencoder-decoder model and the GAN architecture that has been discussed below.

### *A. Autoencoders*

Autoencoder was the first technology to be used in deepfake creation. Autoencoder is used to recreate images that it is trained on. The output generated works in three different phases: encoder, latent space, and a decoder [10]. The encoder first compresses the input pixels to a relatively smaller size by encoding special attributes like skin texture, skin color, facial expressions, open eye, closed eye, head pose, and any minute details of the face. This compressed image is sent as input to latent space that is useful for understanding and learning patterns and structural similarities between the data points [11]. Lastly, the decoder decompresses this information to reconstruct an output based on its representation in latent space. The decoder tries to recreate an image that resembles the original as much as possible.

Fig. 1 shows how an autoencoder can be used to swap two faces. Following the path indicated by the red arrows, Face B is reconstructed similar to Face A. The important aspect here is that both the faces have used the same encoder. This will help the encoder to use general features that are common to both faces, and their positioning in latent space will also be similar. This will allow the autoencoders to transform the same picture with faces of the target and the original individual swapped. Here, the latent space of Face A is referred for Decoder of Face B to be able to reconstruct Face B similar to Face A. This technique has its application in various deepfake technologies like DFaker, DeepFaceLab, and TensorFlow-based deepfakes [12].

### B. *Generative Adversarial Networks*

The majority of the current deepfake technologies incorporate the use of GAN. The GAN architecture was first proposed by Ian Goodfellow in 2014 [13]. He introduced a framework that involves the use of two neural networks that work by challenging each other: the first one generates new data while the other discriminates this new data from the original training data set. Contesting the two neural networks tend to improve both the quality of fake data produced as well as the neural network's discrimination ability. If a large number of images are fed to GAN, it can create a unique image on its own [14]. However, it is necessary to attach a filter that can help differentiate these unique outputs as acceptable or not. For this, GANs make use of a discriminative network that checks the generated data with true data. Both are trained to operate together until the discriminator has falsely classified the generated output as authentic, almost 50% of the time. This helps us conclude that the generator model is successfully generating plausible examples. Fig.1 shows the block diagram explaining the workflow of GAN architecture.

The GAN architecture uses the min-max method for training the generator and discriminator [13]. The min (0) represents a fake output while the max (1) represents a genuine output. The goal of the discriminator is to get as close as possible to the max value such that a realistic-looking deepfake is generated which can be further be used for face-swapping in images and videos. GANs are more suitable for generating new data [15]. The main advantage of GANs over autoencoders is that they can be used for a wider range of tasks for example to produce several classes of data, similar to the MNIST dataset [16]. Autoencoders, on the other hand, are more suitable for compressing data to lower dimensions or generating semantic vectors from it.

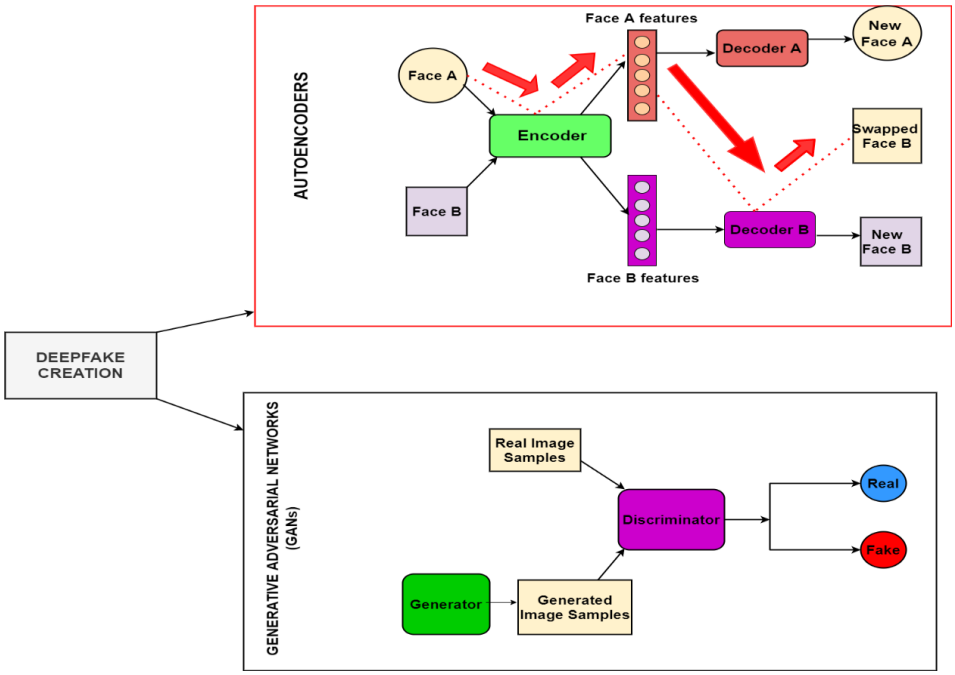


Fig. 1. The workflow of Autoencoders and GAN in the creation of Deepfakes

## B. DEEPFAKE DETECTION

To date, there is much ongoing research in this field. According to data obtained from https://app.dimensions.ai, there have been a total of 931 research publications in this field by the end of August 2020 and is likely to increase further in the coming years. Image, video, and audio are the three major types of deepfakes that are currently being dealt with globally.

Malicious use of deepfakes has become undeniably powerful. It has become a personal weapon of revenge. They are used to manipulate financial markets or to destabilize international relations. Several efforts have been made to prevent against this deceptive and destructive potential of deepfakes. This section presents a survey of deepfake detection methods categorized under the fake image, fake video, and fake audio detection.

### *A. Fake Image Detection*

A new wave of face-swapping technology poses a significant threat to identity and can penetrate systems to get illegitimate access. With many disturbing AI-generated images flooding across various platforms, several detection measures have been proposed in this regard. Most of these fake images are synthesized using high-quality GANs that generate highly realistic content that is very difficult to detect.

Fig.2 shows an example of deepfake image created using a deepfake application called FaceApp. Here U.S President Donald Trump's unhappy face (upper block) has been transformed into a smiling face (lower left) by only using the smile tool in this deepfake application. Next American singer Camila Cabello's face is swapped with that of Donald Trump (lower right) and both these images are challenging to detect as fake.


Fig. 2. Use of Deepfake application FaceApp to generate the fake image

Fake image detection has been treated as a binary classification problem where Convolutional Neural Network (CNN) models have mostly been used for detection [17, 18]. Later adoption of advanced CNN-Xception Network [19] further increased the accuracy of detection. Here they have carried out the extraction of spatial and steganalysis features of the digital content using CNNs. Spatial features involve the identification of visible inconsistencies in the image like facial

blurs, facial texture, artificial smoothness, and contrast difference. Steganalysis help analyze the hidden information in an image through low-level feature extraction. However, with many evolving extensions of GAN, new mechanisms are being adopted. So it is important to improve the generalization ability of detection tools. Instead of focusing on statistical details of a low-level pixel, Xuan et al. [20] proposed the use of Gaussian blur and Gaussian Noise to force classifiers to learn more detailed and meaningful features from improved statistical similarity at the pixel level in images.

Likewise, in [21] a two-phase model is proposed that pairs fake and real images. These pairs then learn from discriminative common fake feature networks (CFFN). The discriminative CFF is used to identify the authenticity of the image. They have highlighted the fact that it is challenging to identify subjects excluded from the training phase using supervised learning. So have introduced a contrastive loss to learn through pairwise learning. The objective of this framework is to detect newly encountered fake images and those generated by a new GAN. Their experimental results have demonstrated to outperform the precision and recall rate of other state-of-the-art methods.

### *B. Fake Video Detection*

When videos are synthesized to swap faces or change facial expressions, the new images do not usually match the lighting conditions or head positioning. This is achieved by geometrically transforming them by rotating, resizing, or otherwise distorting them. This process is guaranteed to leave behind digital artifacts in the resulting image that can make them look doctored. And so various algorithms have been trained to detect these artifacts to check for the authenticity of the content. Li and Lyu [22] propose a method that exposes deepfake videos based on these face-warping artifacts. Instead of self-generating negative data from available datasets like several other models [23,24], which can be a very time consuming and expensive process, they use a simple image processing operation to create negative data. They compare generated face areas with its surrounding using a dedicated CNN model for detection.

Several deepfake algorithms are not able to mimic the normal eye blink rate of a person due to a lack of closed eye images. Wang et al. [25], used contour circle fitting to locate the presence of pupil and thus detect blinks and achieved an accuracy of 96.6%. However, most of these studies were based on datasets created in a laboratory, so their results could not be generalized in a real-world scenario. Li et al. [26], recently proposed the use of CNN along with recurrent neural network (RNN) for detecting eye blinks, to differentiate between a fake and authentic visual. This method outperformed the method proposed in [27], that used landmark detectors to identify the eye corners and eyelid contours for precise detection of eyes and with the use of two different models namely the SVM (Support Vector Machine) and HMM (Hidden Markov Model), detected the rate of eye blinks. This technique was further enhanced to differentiate between a complete and incomplete eye blink. Fogelton and Benesova [28], have designed a model that detects a complete blink, incomplete blink, and no blink, for every frame. This method was implemented on the Researcher's night dataset, and it outperformed all other related work by almost 8%.

Stressing on the fact that several intra-frame inconsistencies and temporal frame inconsistencies can occur across frames in deepfake video generation, Guera and Delp [23] proposed temporal feature analysis of videos to detect fake videos, with the use of a simple convolutional Long-Short - Term- Memory (LSTM) structure. Similar to this, an advanced framework known as SSTNet [29], uses a combination of CNN based spatial feature extraction, steganalysis feature extraction and temporal feature extraction by a single LSTM for fake video detection. This framework achieved a good generalization on the GAN dataset and outperformed previous steganalysis methods when tested on the FaceForensics++ dataset [30]. Fig. 3 shows the working of SSTNet model.

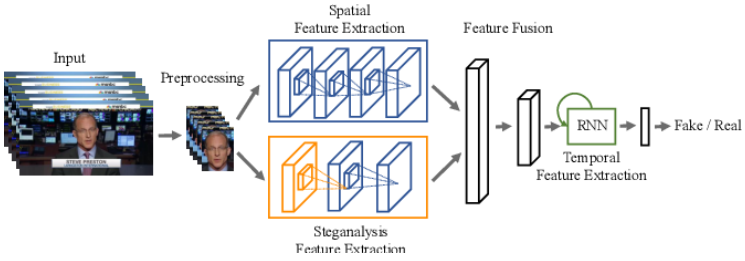


Fig. 3 The proposed framework of SSTNet [29]

### *C. Fake Audio Detection*

Audio manipulation is a combination of AI techniques of deepfakes and general processing like speeding, slowing, cutting, or decontextualizing a video, commonly referred to as Cheapfakes [31]. Resemblyzer is an open-source tool that detects deepfakes by extracting high-level representation of audios that enable developers to compare the two voice samples at any given point of time [32]. Other fake audio detectors focus on the difference in spectrograms - the visual representation of audio, of real and fake audios [33]. However, later an improved CNN based architecture called Wavenet [34] was introduced in various platforms like Text-To-Speech (TTS) and speech recognition. TTS and Voice Conversion (VC) are two types of speech synthesis software. TTS synthesizes human-like speech based on words or phonemes, while VC systems can convert the voice of an utterance into another voice with the same content. DNN models have popularly been used to extract dynamic acoustic

TABLE I
SUMMARY OF PROMINENT DEEPFAKE DETECTION TECHNIQUES

| Method | Type | Dataset | Merits | Demerits |
|---|---|---|---|---|
| CGFace Model[17] | Image | **Real images:** CelebA **Fake images:** generated using PCGAN and BEGAN | automatic extraction of abstract features and use of AdaBoost classifier over softmax | cannot identify hidden features in images |
| Pairwise learning[21] | Images | **Real images**: CelebA **Fake images:** generated from DCGAN, WGAN, and PGGAN | two phase model is proposed that pairs fake and real images and common fake feature network is trained to distinguish the features between the fake and real images using pairwise learning | may not detect new fake features generated by new GAN |
| Preprocessing[20] | Image | **Real images:** CelebA-HQ, **Fake images:** generated from DCGAN, WGANGP and PGGAN. | similar image level preprocessing is done to both real and fake images to destroy low level noise cues so model can learn more intrinsic features for classification | not very large increment in performance compared to other similar approaches |
| Face-warping artifacts[22] | Videos | UADFV, DeepfakeTIMIT | makes use of distinctive artifacts that are obtained during the resolution transformation of images | does not take into consideration temporal inconsistencies across frames |
| Eye Blinking[26] | Videos | **Real:** 49 real interview and presentation videos **Fake**: generated from above-mentioned videos | Highlighted the fact that deepfake videos have inconsistencies in the human blinking rate compared to real videos | only considers lack of blinking for detection and not the frequency of blinking |
| SSTNet: Spatial, temporal and steganalysis[29] | Videos/Image | FaceForensics++ GAN-based Deepfakes | extracts low level, mid-level and high-level artifacts of images in videos and also explores the temporal inconsistencies across successive frames | doesn't take into account the inconsistencies in blinking rate |
| Dynamic acoustic Features[35] | Audio | ASVspoof 2015 database | proposed an HLL scoring method that only makes use of human node outputs for spoof detection | performance is yet to be investigated on replay attacks and spoofing speech produced by WaveNet |

features of audiovisuals and label them as real or fake [35].This proposed methodology has shown to outperform the static feature analysis of Gaussian Mixture Model (GMM) classifiers [36]. The summarized version of the most prominent deepfake detection methods has been listed in Table 1.

## IV. CONTRIBUTIONS

After an in-depth review of several proposed detection methods for deepfakes, the SSTNet model proves to be the most convincing one due to its flexibility and use of a generalized approach in detecting both the fake images and videos. This model achieved an accuracy level of around 90% to 95%, which is much higher compared to other models trained on the same datasets. This method can further be improvised by incorporating other detection methods like the rate of complete eye blinks and investigating for lip-synching evidence, on datasets released by giant tech companies like Google [37] and Facebook [38] to classify fake videos.

Many researchers have failed to recognize that signature forgery and document forgery is also a matter of great concern. Signatures ensure a great deal of authenticity in various sectors of banking, insurance, healthcare, copyrights and governmental regulatory compliance. Forgery of any form could potentially lead to more immense disastrous consequences. Thus, it is important for researchers to equally invest in this sector as well.

## V. CONCLUSION & FUTURE RESEARCH DIRECTIONS

As the deepfake technology approaches towards generating fake content with considerably improved quality, it will likely become impossible to detect them shortly. It is thus, important to immediately respond to the emerging threat posed by deepfakes with great caution. To be able to distinguish between generated and authentic content, organizations can start developing encrypted digital stamps for authentic digital media.

Believability and accessibility have become the most significant drive in deep fake technology. Stopping deepfakes from spreading across massive networks is a considerable challenge and requires social media platforms to step up. They need to develop tools and extensions that can help deal with deepfake content moderation, detection, and prevent their mainstream media coverage. Also confusing deepfakes to generate more flawed output can help detect them easily. This can be achieved with the addition of special noise to digital photos that are uploaded on social media, such that they create a decoy suggesting there is a face when there is none, in reality [39].

To combat deep fake just developing and deploying one or two successful tools is not enough. It will require a constant reinvention of these tools as this technology is evolving at a much faster rate and machine learning plays a crucial role in achieving this. Therefore, the research community should continue their research in developing

countermeasures using machine learning and deep learning to combat the weaponization of deepfakes.

## REFERENCES

[1] R. Kumar, J. Sotelo, K. Kumar, A. de Brebisson, and Y. Bengio, "ObamaNet: Photo-realistic lip-sync from text," pp. 1–4, 2017, [Online]. Available: http://arxiv.org/abs/1801.01442.

[2] D. Yadav and S. Salmani, "Deepfake: A survey on facial forgery technique using generative adversarial network," *2019 Int. Conf. Intell. Comput. Control Syst. ICCS 2019*, pp. 852–857, 2019, doi: 10.1109/ICCS45141.2019.9065881.

[3] "Chinese deepfake app Zao sparks privacy row after going viral | Privacy | The Guardian." https://www.theguardian.com/technology/2019/sep/02/chinese-face-swap-app-zao-triggers-privacy-fears-viral (accessed Aug. 26, 2020).

[4] "AI deepfake app DeepNude transformed photos of women into nudes - Vox." https://www.vox.com/2019/6/27/18761639/ai-deepfake-deepnude-app-nude-women-porn (accessed Aug. 26, 2020).

[5] "Deepfakes and the New Disinformation War | Foreign Affairs." https://www.foreignaffairs.com/articles/world/2018-12-11/deepfakes-and-new-disinformation-war (accessed Aug. 30, 2020).

[6] "The Newest AI-Enabled Weapon: 'Deep-Faking' Photos of the Earth - Defense One." https://www.defenseone.com/technology/2019/03/next-phase-ai-deep-faking-whole-world-and-china-ahead/155944/ (accessed Aug. 24, 2020).

[7] "Deep fakes: AI-manipulated media will be 'WEAPONISED' to trick military | Science | News | Express.co.uk." https://www.express.co.uk/news/science/1109783/deep-fakes-ai-artificial-intelligence-photos-video-weaponised-china (accessed Aug. 24, 2020).

[8] R. Chesney and D. K. Citron, "Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security," *SSRN Electron. J.*, pp. 1753–1820, 2018, doi: 10.2139/ssrn.3213954.

[9] "Don't believe your eyes: Exploring the positives and negatives of deepfakes - AI News." https://artificialintelligence-news.com/2019/08/05/dont-believe-your-eyes-exploring-the-positives-and-negatives-of-deepfakes/ (accessed Aug. 25, 2020).

[10] J. Kietzmann, L. W. Lee, I. P. McCarthy, and T. C. Kietzmann, "Deepfakes: Trick or treat?," *Bus. Horiz.*, vol. 63, no. 2, pp. 135–146, 2020, doi: 10.1016/j.bushor.2019.11.006.

[11] "Understanding Latent Space in Machine Learning | by Ekin Tiu | Towards Data Science." https://towardsdatascience.com/understanding-latent-space-in-machine-learning-de5a7c687d8d (accessed Aug. 28, 2020).

[12] T. T. Nguyen, C. M. Nguyen, D. T. Nguyen, D. T. Nguyen, and S. Nahavandi, "Deep Learning for Deepfakes Creation and Detection: A Survey," pp. 1–12, 2019, [Online]. Available: http://arxiv.org/abs/1909.11573.

[13] I. J. Goodfellow *et al.*, "Generative adversarial nets," *Adv. Neural Inf. Process. Syst.*, vol. 3, no. January, pp. 2672–2680, 2014.

[14] "Generative Adversarial Networks: The Tech Behind DeepFake and FaceApp." https://interestingengineering.com/generative-adversarial-networks-the-tech-behind-deepfake-and-faceapp (accessed Aug. 27, 2020).

[15] M. Mirza and S. Osindero, "Conditional Generative Adversarial Nets," pp. 1–7, 2014, [Online]. Available: http://arxiv.org/abs/1411.1784.

[16] "What is the difference between Generative Adversarial Networks and Autoencoders? - Quora." https://www.quora.com/What-is-the-difference-between-Generative-Adversarial-Networks-and-Autoencoders (accessed Aug. 29, 2020).

[17] L. M. Dang, S. I. Hassan, S. Im, J. Lee, S. Lee, and H. Moon, "applied sciences Identification Using Convolutional Neural Network," 2018, doi: 10.3390/app8122610.

[18] Y. Aslam and N. Santhi, "A Review of Deep Learning Approaches for Image Analysis," *Proc. 2nd Int. Conf. Smart Syst. Inven. Technol. ICSSIT 2019*, no. Icssit, pp. 709–714, 2019, doi: 10.1109/ICSSIT46314.2019.8987922.

[19] C. Google, "Xception : Deep Learning with Depthwise Separable Convolutions," pp. 1251–1258, 2014.

[20] X. Xuan, B. Peng, W. Wang, and J. Dong, "On the Generalization of GAN Image Forensics," *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)*, vol. 11818 LNCS, no. 61502496, pp. 134–141, 2019, doi: 10.1007/978-3-030-31456-9_15.

[21] C. Hsu, Y. Zhuang, and C. Lee, "applied sciences Deep Fake Image Detection Based on Pairwise Learning," 2020, doi: 10.3390/app10010370.

[22] Y. Li and S. Lyu, "Exposing DeepFake Videos By Detecting Face Warping Artifacts."

[23] G. David and E. J. Delp, "Deepfake Video Detection Using Recurrent Neural Networks."

[24] D. Afchar and V. Nozick, "MesoNet : a Compact Facial Video Forgery Detection Network."

[25] M. Wang, L. Guo, and W. Chen, "Blink detection using Adaboost and contour circle for fatigue recognition," *Comput. Electr. Eng.*, vol. 0, pp. 1–11, 2016, doi: 10.1016/j.compeleceng.2016.09.008.

[26] Y. Li, M. Chang, and S. Lyu, "In Ictu Oculi : Exposing AI Generated Fake Face Videos by Detecting Eye Blinking."

[27] M. Perception and C. Technical, "Eye Blink Detection Using Facial Landmarks," 2016.

[28] A. Fogelton and W. Benesova, "Eye blink completeness detection," *Comput. Vis. Image Underst.*, vol. 176–177, pp. 78–85, 2018, doi: 10.1016/j.cviu.2018.09.006.

[29] "SSTNET : DETECTING MANIPULATED FACES THROUGH SPATIAL , STEGANALYSIS AND TEMPORAL FEATURES YuTao Gao Yu Xiao Alibaba Group China," pp. 2952–2956, 2020.

[30] R. Andreas, D. Cozzolino, L. Verdoliva, J. Thies, M. Nießner, and C. Riess, "FaceForensics ++ : Learning to Detect Manipulated Facial Images."

[31] B. Paris and J. Donovan, "Deepfakes and Cheap Fakes," *Data Soc.*, p. 47, 2019, [Online]. Available: https://site.ieee.org/sagroups-7011/%0Ahttps://site.ieee.org/sagroups-7011/blog/%0Ahttps://datasociety.net/library/deepfakes-and-cheap-fakes/.

[32] "Resemble AI launches voice synthesis platform and deepfake detection tool | VentureBeat." https://venturebeat.com/2019/12/17/resemble-ai-launches-voice-synthesis-platform-and-deepfake-detection-tool/ (accessed Aug. 30, 2020).

[33] "Detecting Audio Deepfakes With AI | by Dessa | Dessa News | Medium." https://medium.com/dessa-news/detecting-audio-deepfakes-f2edfd8e2b35 (accessed Aug. 30, 2020).

[34] "WaveNet: A generative model for raw audio | DeepMind." https://deepmind.com/blog/article/wavenet-generative-model-raw-audio (accessed Aug. 30, 2020).

[35] H. Yu, Z. H. Tan, Z. Ma, R. Martin, and J. Guo, "Spoofing Detection in Automatic Speaker Verification Systems Using DNN Classifiers and Dynamic Acoustic Features," *IEEE Trans. Neural Networks Learn. Syst.*, vol. 29, no. 10, pp. 4633–4644, 2018, doi: 10.1109/TNNLS.2017.2771947.

[36] D. Reynolds, "Gaussian Mixture Models," *Encycl. Biometrics*, no. 2, pp. 659–663, 2009, doi: 10.1007/978-0-387-73003-5_196.

[37] "Google has released a giant database of deepfakes to help fight deepfakes | MIT Technology Review." https://www.technologyreview.com/2019/09/25/132884/google-has-released-a-giant-database-of-deepfakes-to-help-fight-deepfakes/ (accessed Aug. 30, 2020).

[38] "Deepfake Detection Challenge Dataset." https://ai.facebook.com/datasets/dfdc/ (accessed Aug. 30, 2020).

[39] "Scientists Are Taking the Fight Against Deepfakes to Another Level | Discover Magazine." https://www.discovermagazine.com/technology/scientists-are-taking-the-fight-against-deepfakes-to-another-level (accessed Aug. 30, 2020).