### Abstract

Deepfake videos, which enable one to accurately manipulate faces using advanced deep learning algorithms have attracted much attention in recent years. While the abundance of deepfake videos on-line (especially those in relation to celebrities and politicians) is troubling. Those doctored videos are weaponized to harm reputations and manipulate public opinion, severely undermining social harmony. But while the deepfake tech itself is value-neutral, its deployment as a weapon of malintent has gotten all too common.

In ticking both these boxes, deepfakes are therefore the subject of many research efforts to try and prevent their societal risks. This segment encompasses the instantiation of detection techniques and initiates in gigantic benchmarks. In this paper, we survey the work in recent years on detecting deepfake videos across different works including audio/visual-based counter-forensics and adversarial detection defense approaches proposed towards obtaining new benchmarks to speed up research development. It was observed that available detection approaches are still not suitable enough for deployment in the real. Therefore, future research should focus on enhancing the generalization and robustness of these detection methods.

1. **Introduction**

The concern regarding morphed videos has raised concern over the years especially when deepfake technology which makes use of deep learning tools modifies images and pictures used since then. Deepfake algorithms are capable of inserting faces seen in other video clips or movies into given video clips through procedures like autoencoders or generative adversarial networks (GANs). It thereby simplifies the entire process involved when constructing substituted face videos as long as being

While deepfake technology has potential applications in areas such as filmmaking and virtual reality, its abuse for malicious purposes has become widespread as shown in Figure 1, with many fake videos circulate online targeting mainly politicians and celebrities. The first deepfake cases surfaced in 2017 when a Reddit user named deepfakes made a celebrity porn video. This is the beginning of the inevitable manipulation of technology. After that, applications like FakeApp and FaceSwap appeared, making it easier to create deepfakes. In June 2019, an app called DeepNude created real-world clothing, causing global concern. In addition to violating individual privacy, these tools are widely used to run political campaigns and sway public opinion. As a result, deepfake content detection has become an important issue for individuals, companies and governments around the world.



****FIGURE 1****

The growing interest in deepfake technology, has led to an increase in related research efforts. The past two years have seen tremendous progress in the development of new detection techniques. Initially, the number of video datasets available for deepfake detection tasks has been expanded. From small datasets like DeepFake-TIMIT and UADFV to large datasets like FaceForensic++, Celeb-DF, DFDC, DeeperForensic, the resources available for training and recognition models have grown exponentially. Recently, companies such as Amazon, Facebook and Microsoft collaborated to create the DeepFake Detection Challenge (DFDC), which aims to advance new technologies in deepfake video detection as well as the DeeperForensics Challenge This was organized by the These initiatives has led to the development of many effective detection methods, which have proven effective in fraud detection industries Despite these advances, there are still significant issues in the depth field of vision. As deepfake techniques continue to evolve, the number of automated videos increases, potentially making traditional detection methods inadequate to new deepfake algorithms so it is important to investigate future improvements to deepfake research and develop corresponding methods of discovery. This review will focus on existing video detection systems developed for deepfake video, with the aim of promoting further development of deepfake video detection.

1. **Related Work**

This section targets to discover and analyze one of a kind techniques hired for detecting deepfakes. Various techniques have leveraged deep learning as their center era. These strategies both attention on identifying inconsistencies in video frames or in person frames. Image analysis techniques goal several elements such as face warping artifacts, eye blinking prices, and head moves. In 2018, "MesoNet" was delivered, making use of the Inception model to become aware of faults on the mesoscopic level. Convolutional Neural Networks (CNNs) have confirmed super function extraction capabilities, that are essential for deepfake detection.

Other techniques have incorporated CNNs with extra learning fashions like Recurrent Neural Networks (RNNs), Long Short-Term Memory Networks (LSTMs), and Capsule Networks to decorate accuracy by way of identifying temporal inconsistencies. These approaches have shown promising consequences on datasets containing videos generated by way of FaceSwap and other deepfake techniques.

Despite the improvements made, developing greater sturdy models to come across decrease high-quality deepfakes remains a task. As deepfake era techniques preserve to adapt, retaining excessive detection accuracy is an ongoing warfare.

## **DEEPFAKE VIDEO GENERATION**

Since the first release of deepfake videos, many new manipulation algorithms based primarily on generative networks have emerged. These deepfake techniques can generate false positives, posing a serious threat to individual privacy and social security. This section reviews the development of deepfake algorithms and divides them into two main types.

### 3.1 Development of Deepfake Technologies

The use of faces is not a recent innovation. One of the first cases on record is the US. A stunning 1865 portrait of President Abraham Lincoln, worked on to some extent. Advances in computer graphics have made digital image processing easier and more sophisticated. However, recent advances in deep learning have changed the use of faces in depth, creating the practice of modern deep inflammation. Deepfake algorithms can be broadly classified into two types based on their purpose: face swapping and face reconstruction.

### 3.2 Face Swapping Algorithms

The facial swapping videos of the individuals in the two videos have received a lot of attention in recent years. Research on this topic has been ongoing since 2017. In a study by Korshunova et al, convolutional neural networks (CNNs) were trained to capture the method of target detection from a collection of unstructured images, enabling the creation of an optimal angle-altering image but this method did not measure in time development, and limits its application to high-quality video generation.

In the same year, Olszewski et al. Introduced a novel method for creating videos using RGB images and source video sequences. A deep generative network accounted for the per-frame texture distortions of target detection by using source textures and a target texture. This method allowed the newly compressed faces to be associated with the resulting video, instead of the original faces as described in .

The first face-swapping video using deepfake technology was posted by a Reddit user in December 2017, which had a huge global impact. The inspiration for the deepfake algorithm is widely believed to have come from the work of Korshunova et al. , where CNN was used to generate face-altering images. This has led to face swap videos all over the world, for both good and bad purposes.

An improved original deeppfake algorithm, Faceswap-GAN, was proposed by . Including enemy loss and sensitivity loss to provide more realistic faces, which enabled the auto-encoder of VGGFace [19 and DeepFaceLab , an open source deepfake generation system, developed for pipelines ease-of-use performance was greater for individuals without a professio

#### **3.1.2 Face reenactment**

Unlike facial-repetition technologies, facial repetition algorithms aim to alter individual faces in videos, allowing them to create videos in which individuals appear to Origins of facial repetition algorithms 2006 perform actions they did not actually perform can be seen as the first transaction of the year .Vlasic et al. proposed facial reconstruction based on modified facial templates under different exponential parameters. Subsequent studies have augmented parametric models to capture facial images, achieving higher realism but often no temporal consistency

In recent years, advances in computing power have led to new advances in facial reconstruction techniques. Thies and so on. Introduced Face2Face, a real-time monocular facial reconstruction. In this approach, a new global non-rigid model-based bundling method was used to reconstruct the facial features of target and source actors. In addition, the subspace distortion placement technique facilitated expression placement between actors. Despite significant performance improvements over previous methods, Face2Face has limitations in terms of consistent head movements and easy detection of deep facial regions

Recognizing these challenges, Suvajanakorn et al. addressed the similarity between facial expression and voice. A technique for visualizing audio sequences on videos was developed, allowing actors to synchronize lip movements with utterances. Fried and so on. Further enhanced video head language editing by applying neural facial rendering techniques to speech content processing.

To obtain photorealistic image and video reanimation, Kim et al proposed a new method using a generative neural network with spatio-temporal structure. This method not only provides facial expressions but also head posture, gaze, and eye blinks, thus improving the limits of Face2Face. Thies and so on. also contributed to optimizations in Face2Face, introducing neural texture methods to increase texture clarity in the face area.

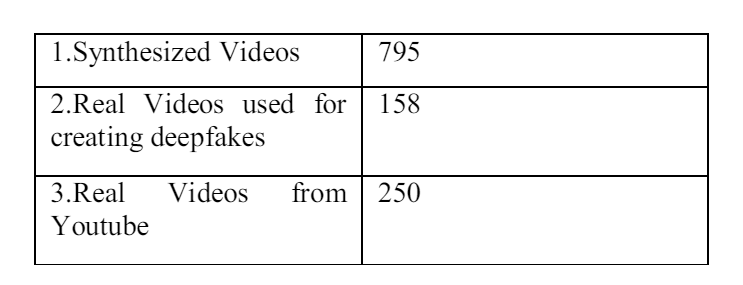
These developments highlight the ongoing efforts to overcome the technical challenges in facial reconstruction driven by advances in deep learning techniques and computational capabilities

## **DEEPFAKE VIDEO DETECTION**

Deepfake movies are more and more diagnosed as a widespread threat to personal privateness and societal security. Detecting these manipulated videos has spurred a whole lot of research efforts, categorised into wonderful techniques primarily based at the features they utilize. Early processes in general centered on figuring out inconsistencies springing up from the face synthesis process. In assessment, modern-day detection strategies embody a broader variety of techniques. Initially, trendy neural networks emerged as a common tool, treating deepfake detection as a conventional category venture. Subsequent techniques have honed in on temporal consistency functions, aiming to come across anomalies between successive frames inside faux movies. Researchers have additionally exploited visual artifacts generated in the course of the mixing process to determine manipulated content extra successfully. Recent advancements underscore a shift closer to leveraging essential features. Innovations in camera fingerprinting and biological signal-based totally approaches keep promise in improving detection accuracy significantly. These strategies capitalize on precise identifiers inherent in video pictures, providing strong defenses against increasingly state-of-the-art deepfake technology. In the drawing close sections, we are able to delve deeper into every of these detection methodologies, highlighting their strengths, barriers, and implications for fighting the proliferation of deepfake movies.

## **5. DATASETS**

The Celeb-DF dataset contains 408 real videos and 795 artificial videos, generated using the enhanced DeepFake generation method. Videos averaged 13 seconds in length and recorded at 30 frames per second (fps). Unlike previous datasets, which are composed of high-quality videos with a lot of visual content, the videos produced in this dataset exhibit relatively few visual elements, causing their quality to fly effective This makes the dataset particularly robust because it contains few high-quality deepfakes objects.



**6. Methodology**

The CNN model architecture is implemented using TensorFlow's Keras API, featuring several carefully designed layers to optimize image classification for detecting deepfakes. The model begins with an input layer that accepts images of shape (64, 64, 3). To ensure compatibility with TensorFlow's operations, a lambda layer is utilized to cast the input data to the float32 format. Following this, a normalization layer standardizes the input data, which is crucial for stabilizing the learning process and improving the model's performance.

The core of the architecture consists of two convolutional layers equipped with ReLU activation functions, designed to extract spatial features from the images. These layers are followed by a max-pooling layer, which down-samples the feature maps, reducing their spatial dimensions and enabling the model to concentrate on the most prominent features. To mitigate the risk of overfitting, dropout layers are incorporated, which randomly set a fraction of input units to zero during training, thereby improving the model's generalization capabilities.

After the convolutional and pooling layers, the feature maps are flattened into a one-dimensional vector to prepare them for the subsequent dense layer. This dense layer, equipped with a single neuron, uses a sigmoid activation function to output a probability score, facilitating binary classification of the images as either real or deepfake. The entire model architecture is summarized, providing a detailed overview of each layer and the corresponding number of parameters.

The training process involves feeding the reshaped and fused training data into the model, which then learns to classify images based on subtle differences that may indicate manipulation. The architecture is specifically designed to detect these subtle cues, enhancing the model's ability to distinguish between real and deepfake images effectively.

1. **REFERENCES**

### Research Paper 1: Very Deep Convolutional Networks for Large-Scale Image Recognition

**What they do:** Developed very deep convolutional networks for large-scale image recognition. **Model they use:**

* **Technology:** Convolutional Neural Network (CNN)
* **Models/Algorithms:** VGGNet

**Reference:** Simonyan, K., & Zisserman, A. (2015). Very deep convolutional networks for large-scale image recognition. International Conference on Learning Representations (ICLR) 2015. arXiv:1409.1556v6 [cs.CV].

### Research Paper 2: Deep Fakes: A Looming Challenge for Privacy, Democracy, and National Security

**What they do:** Explored the implications of deep fakes on privacy, democracy, and national security. **Model they use:**

* **Technology:** Legal and social analysis

**Reference:** Chesney, R., & Citron, D. K. (2018). Deep fakes: A looming challenge for privacy, democracy, and national security. California Law Review, 107, Forthcoming. U of Texas Law, Public Law Research Paper No. 692; U of Maryland Legal Studies Research Paper No. 2018-21.

### Research Paper 3: Exposing AI Generated Fake Face Videos by Detecting Eye Blinking

**What they do:** Exposed AI-generated fake face videos by detecting the absence of eye blinking. **Model they use:**

* **Technology:** Computer Vision
* **Models/Algorithms:** Eye Blinking Detection

**Reference:** Li, Y., Chang, M.-C., & Lyu, S. (2018). In Ictu oculi: Exposing AI generated fake face videos by detecting eye blinking. arXiv:1806.02877v2 [cs.CV].

### Research Paper 4: Exposing DeepFake Videos By Detecting Face Warping Artifacts

**What they do:** Exposed deepfake videos by detecting face warping artifacts. **Model they use:**

* **Technology:** Computer Vision
* **Models/Algorithms:** Face Warping Artifact Detection

**Reference:** Li, Y., & Lyu, S. (2019). Exposing deepfake videos by detecting face warping artifacts. arXiv:1811.00656v3 [cs.CV]. [https://doi.org/10.1016/S0969-4765(19)30137-7](https://doi.org/10.1016/S0969-4765(19)30137-7" \t "_new).

### Research Paper 5: MesoNet: A Compact Facial Video Forgery Detection Network

**What they do:** Developed a compact network for detecting facial video forgeries. **Model they use:**

* **Technology:** Neural Networks
* **Models/Algorithms:** MesoNet

**Reference:** Afchar, D., Nozick, V., Yamagishi, J., & Echizen, I. (2018). MesoNet: A compact facial video forgery detection network. arXiv:1809.00888v1 [cs.CV]. [https://doi.org/10.1109/WIFS.2018.8630761](https://doi.org/10.1109/WIFS.2018.8630761" \t "_new).

### Research Paper 6: Exposing Deep Fakes Using Inconsistent Head Poses

**What they do:** Exposed deep fakes by detecting inconsistent head poses. **Model they use:**

* **Technology:** Computer Vision
* **Models/Algorithms:** Head Pose Detection

**Reference:** Yang, X., Li, Y., & Lyu, S. (2019). Exposing deep fakes using inconsistent head poses. In ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (pp. 14-18). [https://doi.org/10.1109/ICASSP.2019.8682999](https://doi.org/10.1109/ICASSP.2019.8682999" \t "_new).

### Research Paper 7: Use of a Capsule Network to Detect Fake Images and Videos

**What they do:** Used a capsule network to detect fake images and videos. **Model they use:**

* **Technology:** Capsule Networks
* **Models/Algorithms:** Capsule Network

**Reference:** Nguyen, H. H., Yamagishi, J., & Echizen, I. (2019). Use of a capsule network to detect fake images and videos. arXiv:1910.12467v2 [cs.CV].

### Research Paper 8: Exploiting Visual Artifacts to Expose Deepfakes and Face Manipulations

**What they do:** Exploited visual artifacts to expose deepfakes and face manipulations. **Model they use:**

* **Technology:** Computer Vision
* **Models/Algorithms:** Artifact Detection

**Reference:** Matern, F., Riess, C., & Stamminger, M. (2019). Exploiting visual artifacts to expose deepfakes and face manipulations. In 2019 IEEE Winter Applications of Computer Vision Workshops (WACVW) (pp. 1-10). [https://doi.org/10.1109/WACVW.2019.00020](https://doi.org/10.1109/WACVW.2019.00020" \t "_new).

### Research Paper 9: The Upside of Deep Fakes

**What they do:** Explored potential positive uses of deep fakes. **Model they use:**

* **Technology:** Legal and social analysis

**Reference:** Hartzog, J., & Woodrow, J. (2019). The upside of deep fakes. Maryland Law Review, 78(4).

### Research Paper 10: You Thought Fake News Was Bad? Deep Fakes are Where the Truth Goes to Die

**What they do:** Discussed the dangers of deepfakes and their impact on truth. **Model they use:**

* **Technology:** Journalism and media analysis

**Reference:** Schwartz, O. (2018, November 12). You thought fake news was bad? Deep fakes are where the truth goes to die. The Guardian. Retrieved from [https://www.theguardian.com](https://www.theguardian.com" \t "_new).

### Research Paper 11: Review: Inception-v3 — 1st Runner Up (Image Classification) in ILSVRC 2015

**What they do:** Reviewed the Inception-v3 model and its performance in the ILSVRC 2015 competition. **Model they use:**

* **Technology:** Convolutional Neural Network (CNN)
* **Models/Algorithms:** Inception-v3

**Reference:** Tsang, S.-H. (2015). Review: Inception-v3 — 1st runner up (Image classification) in ILSVRC 2015. Retrieved from [https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c](https://medium.com/@sh.tsang/review-inception-v3-1st-runner-up-image-classification-in-ilsvrc-2015-17915421f77c" \t "_new).

### Research Paper 12: Deepfake Video Detection Using Recurrent Neural Networks

**What they do:** Developed a method for deepfake video detection using recurrent neural networks. **Model they use:**

* **Technology:** Recurrent Neural Networks (RNN)
* **Models/Algorithms:** RNN for Video Analysis

**Reference:** Güera, D., & Delp, E. J. (2018). Deepfake video detection using recurrent neural networks. In 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS) (pp. 1-6). [https://doi.org/10.1109/AVSS.2018.8639163](https://doi.org/10.1109/AVSS.2018.8639163" \t "_new).

### Research Paper 13: Face Manipulation with Generative Adversarial Network

**What they do:** Discussed face manipulation using GANs. **Model they use:**

* **Technology:** Generative Adversarial Network (GAN)
* **Models/Algorithms:** GAN for Face Manipulation

**Reference:** Xu, L. (2021, April). Face manipulation with generative adversarial network. In Journal of Physics: Conference Series (Vol. 1848, No. 1, p. 012081). IOP Publishing.

### Research Paper 14: Deep Detection for Face Manipulation

**What they do:** Developed methods for deep detection of face manipulation. **Model they use:**

* **Technology:** Neural Networks
* **Models/Algorithms:** Face Manipulation Detection

**Reference:** Feng, D., Lu, X., & Lin, X. (2020, November). Deep detection for face manipulation. In International Conference on Neural Information Processing (pp. 316-323). Springer, Cham.

### Research Paper 15: Advances in Neural Information Processing Systems

**What they do:** Overview of various advancements in neural information processing systems. **Model they use:**

* **Technology:** Neural Networks
* **Models/Algorithms:** Various models including GANs

**Reference:** Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... & Bengio, Y. (2014). Advances in neural information processing systems. Curran Associates, Inc, 27, 2672-2680.

### Research Paper 16: Conditional Generative Adversarial Nets

**What they do:** Introduced conditional GANs for various applications. **Model they use:**

* **Technology:** Generative Adversarial Network (GAN)
* **Models/Algorithms:** Conditional GAN

**Reference:** Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. arXiv preprint arXiv:1411.1784.

### Research Paper 17: Effective and Fast Deepfake Detection Method Based on Haar Wavelet Transform

**What they do:** Developed a fast deepfake detection method using Haar wavelet transform. **Model they use:**

* **Technology:** Wavelet Transform
* **Models/Algorithms:** Haar Wavelet Transform

**Reference:** Younus, M. A., & Hasan, T. M. (2020, April). Effective and fast deepfake detection method based on haar wavelet transform. In 2020 International Conference on Computer Science and Software Engineering (CSASE) (pp. 186-190). IEEE.

### Research Paper 18: How Do the Hearts of Deep Fakes Beat? Deep Fake Source Detection via Interpreting Residuals with Biological Signals

**What they do:** Detected deep fake sources by interpreting residuals with biological signals. **Model they use:**

* **Technology:** Biological Signal Analysis
* **Models/Algorithms:** Residual Analysis

**Reference:** Ciftci, U. A., Demir, I., & Yin, L. (2020, September). How do the hearts of deep fakes beat? Deep fake source detection via interpreting residuals with biological signals. In 2020 IEEE International Joint Conference on Biometrics (IJCB) (pp. 1-10). IEEE.

### Research Paper 19: The Creation and Detection of Deepfakes: A Survey

**What they do:** Surveyed the creation and detection methods for deepfakes. **Model they use:**

* **Technology:** Survey and Analysis
* **Models/Algorithms:** Various deepfake detection methods

**Reference:** Mirsky, Y., & Lee, W. (2021). The creation and detection of deepfakes: A survey. ACM Computing Surveys (CSUR), 54(1), 1-41.

### Research Paper 20: Media Forensics and Deepfakes: An Overview

**What they do:** Provided an overview of media forensics and deepfake detection. **Model they use:**

* **Technology:** Media Forensics
* **Models/Algorithms:** Various forensic analysis techniques

**Reference:** Verdoliva, L. (2020). Media forensics and deepfakes: an overview. IEEE Journal of Selected Topics in Signal Processing, 14(5), 910-932.

### Research Paper 21: A Landscape View of Deepfake Techniques and Detection Methods

**What they do:** Reviewed deepfake techniques and detection methods. **Model they use:**

* **Technology:** Review and Analysis
* **Models/Algorithms:** Various deepfake detection methods

**Reference:** Abdulredaa, A. S., & Obaida, A. J. (2022). A landscape view of deepfake techniques and detection methods. International Journal of Nonlinear Analysis and Applications, 13(1), 745-755.

### Research Paper 22: Deepfakes: A New Threat to Face Recognition? Assessment and Detection

**What they do:** Assessed the threat of deepfakes to face recognition and proposed detection methods. **Model they use:**

* **Technology:** Face Recognition and Detection
* **Models/Algorithms:** Deepfake Detection

**Reference:** Korshunov, P., & Marcel, S. (2018). Deepfakes: a new threat to face recognition? Assessment and detection. arXiv preprint arXiv:1812.08685.

窗体顶端

窗体底端