ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 07 July 2024

Video Forensics: A Comprehensive Study

[1] M. Safwan Baig, [2] Md. Adnan Yunus, [3] Ata Ur Rahman, [4] Mujtaba G. M

[1][2][3][4] Department of Computer Science and Engineering, Methodist College of Engineering and Technology, Hyderabad, India

Email: [1] mirzasafwanbaig3@gmail.com, [2] mdadnan1446@gmail.com, [3] AtaurRahman03@outlook.com

Abstract*— Deepfakes are the latest - and fast-developing - form of attack on digital video and audio. They exploit the recent breakthroughs in machine learning technology, specifically Generative Adversarial Networks (GANs), to produce extremely realistic fake video. Deepfakes can swap faces or even synthesize entire facial gestures with a high-level of craft that is hard to distinguish between the real and generated content. With the rising quality of deepfakes, robust video forensic methods are necessary to detect and verify their presence. Human analysis and basic heuristic methods have failed against well crafted deepfakes. Consequently, recent studies suggest machine learning and computer vision approaches for their detection. Some popular spatial detection methods include Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks and hybrid spatial-temporal models that search for subtle inconsistencies in the video. Recent datasets like FaceForensics++, Celeb-DF and the Deep Fake Detection Challenge (DFDC) have been proposed for training and testing detection models. Optical flow based techniques have also been integrated with detection models to search for inconsistencies in the motion fields of successive frames. This survey discusses the state-of-the-art techniques for the generation and detection of deepfakes and encourages the development of video forensics to check the authenticity of a given content.

Index Terms*— Deepfakes, digital video attacks, machine learning, Generative Adversarial Networks (GANs), fake video, facial gestures synthesis, video forensics, human analysis, machine learning detection, computer vision detection, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Long Short-Term Memory (LSTM) networks, spatial-temporal models, FaceForensics++, Celeb-DF, Deep Fake Detection Challenge (DFDC), optical flow techniques, motion field inconsistencies

I. INTRODUCTION

Deepfakes made with deep neural networks are becoming a huge barrier to media veracity. These highly manipulated edits naturally integrate the facial characteristics of people in images and videos with those of others, making it difficult to distinguish between what's real and what's fake. New research from Deeptrace Lab [1] reveals just how pervasive this issue has become, with nearly 15,000 examples of deepfake media already in circulation online. Perhaps even more worrying from this dataset is the emerging trend that more than 13,000 videos on some porn sites are mostly edited [2]. They are made with images of well-known individuals to damage their reputation and mislead the audience. With the evolution of deepfake technology, so do worries and potential impacts. This deceptive technique has evolved to such an extent that it is now a challenge to separate facts from fiction, as the clone is now saying words that the originals have never said in their lives. The impacts are more than just misleading information; The rising of deepfakes will likely become a significant threat to social order and democracy, as well as impact public opinion and the geopolitical environment. Many government and non-government organizations have invested efforts in solving the deepfake issue because of its seriousness. Lots of detection schemes have been studied, including traditional and advanced machine learning approaches. To counter false carriers, Support Vector Machines (SVMs), Convolutional Neural Networks (CNNs) and recurrent models such as Recurrent Neural Networks (RNNs) and Long Short-Term Memory Networks (LSTMs) have been employed [5]. Furthermore, the researchers.

reviewed traditional cues like distinct head poses and unusual background colors could be useful as indicators of tampering. In spite of the fact that spatial features of a video frame are very well handled by the above methods, there is an obvious lack of research in tackling the temporal features in deepfakes. Since most of the deepfakes exist in video format, temporal information cannot be ignored. We believe that by scrutinizing the temporal discrepancies in addition to the spatial cues can enhance the detection accuracy. In this paper, we propose a novel deepfake detection method using an intra and inter-frame deep learning framework. We employ optical flow which is a traditional hand-crafted feature to describe the temporal information between adjacent video frames along with the spatial features learned by the deep neural network. Our model learns the characteristics of the facial motion between the frames which encodes the intricate relationship between the temporal and spatial features of a video. We evaluate our proposed method extensively using various datasets in terms of Accuracy, Recall, Precision, F1-score and AUC. In this age of digitalization, deepfakes are emerging as a major threat to the society. Our work is a step towards tackling this problem and thus proposes a robust method for detecting deepfakes and ultimately protecting the integrity of the media and maintaining the trust of the public.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 07 July 2024

II. BACKGROUND AND RELATED WORKS

A. Deepfake Generation

The deceptive media landscape has changed dramatically with the advancement in learning-based algorithms that can generate very realistic forged videos. In fact, with the recent emergence of adversarial techniques (e.g., Generative Adversarial Nets (GAN)), the speed of digital forgeries creation has significantly increased [6]. GAN is one of the most known methods used to generate deepfakes. The basic idea of this network is to compete between two neural networks called the generator (G) and the discriminator (D) [7]. The generator attempt to fool the discriminator by generating fake data, while the discriminator try to classify the real media from the fake ones [6]. GAN was first introduced by Goodfellow et al. [7] in 2014 and trained with adversarial loss functions as shown below:

$$Ladv(D) = max\left(log\, D\,(x) + log\left(1 - D\big(G(z)\big)\right)\right) \quad (1)$$

$$Ladv(G) = minlog\left(1 - D\big(G(z)\big)\right) \quad (2)$$

There are two main types of deepfake generation methods: FaceSwap and Face Synthesis. FaceSwap attempts to transfer a target face to a source face, while face synthesis attempts to synthesize facial components. For example, recent methods such as the High-Resolution Face Swapping from Disney Research [8] attempt to transfer faces in source videos to target videos with state-of-the-art results. LandmarkGAN [9] is an example of face synthesis method that synthesize facial components based on facial landmarks. Unfortunately, the need to produce large quantities of convincing counterfeits has led to the use of highly sophisticated printing techniques, making it necessary to use advanced detection methods. Adversarial detectors [7, 8] have been proposed to deal with this problem, but they can be attacked effectively [10], and thus won't necessarily detect the counterfeit currency in many instances. This has led to the exploration of more general multi-model techniques [11], [12], which are an active research topic at the moment.

B. Deepfake Detection

So far, the detection of deepfakes mostly based on spot-checking for inconsistencies or other anomalies in the forgeries [11]. Recently, most of the detection methods are based on machine learning methods to formulate the detection as a general classification problem. Recent works like FSSPOTTER, a unified framework proposed by Peng Chen et al. [12], uses a Spatial Feature Extractor (SFE) and a Temporal Feature Aggregator (TFA) that operate on the spatial and the temporal discrepancies between frames. Additionally, Irene Amerini et al. [13] and Shivangi et al. [14] propose different methods to make use of the temporal inconsistencies, optical flow fields, transfer learning and motion compensation techniques to improve the deepfake detection. David Guera et al. [8] showed that Long Short-Term Memory (LSTM) networks can be effectively used along with CNN models for the deepfake detection. They used InceptionV3 features and LSTM to encode the temporal information in video sequences to propose a temporal-aware LSTM network for the automatic deepfake detection. Most of the state-of-the-art methods extract intra-frame features and the detection of inter-frame features to make use of the temporal inconsistencies between video frames should be considered as a future work for deepfake detection.

III. LITERATURE REVIEW

The field of video forensics and especially deepfake detection has seen tremendous growth over the last few years due to the exponential growth in machine learning and computer vision as a whole. This section reviews the current literature in-depth and outlines the important techniques and methods used for tampered video detection and the datasets on which the methods are evaluated.

A. Deepfake Generation Techniques

The creation of highly convincing fake videos is made feasible by the recent advancements in machine learning techniques particularly Generative Adversarial Networks (GANs) [7]. GANs were first introduced by Goodfellow et al. in 2014. GANs work by training two neural networks: one encoder (generator) that generates synthetic data and a decoder (discriminator) that tries to classify the synthetic data as fake or real. As training proceeds, the discriminator tries to label the output of the generator as fake whereas the generator tries to fool the discriminator by producing more realistic data. This adversarial training shown in Fig. 1 is an extremely powerful concept and has led to the development of deepfakes [7]. There are two primary approaches to deepfake generation: FaceSwap and Face Synthesis. FaceSwap involves overlaying one person's face onto another's, maintaining realistic expressions and movements. LandmarkGAN, proposed by Sun et al., is an example where facial landmarks guide the synthesis process to create convincing face-swaps [9]. Face Synthesis, on the other hand, involves generating a completely new face using facial landmarks and attributes, which can then be manipulated to fit various contexts. Notable advancements include Disney Research's high-resolution face-swapping techniques, which focus on maintaining high visual fidelity and seamless integration of the swapped faces [8]. These techniques have demonstrated the capability to produce deepfakes indistinguishable from real videos to the human eye.

B. Deepfake Detection Techniques

Deepfake generation is becoming increasingly sophisticated, so the detection method should also be advanced. The traditional method includes manual checking, color-gradients and other simple heuristics [7]. However, these methods fail to detect recent deepfakes because of their high quality. So, the researchers shifted their approach

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 07 July 2024

towards machine learning and computer vision algorithms to detect deepfakes. FSSPOTTER is one of the well know method to detect deepfake video developed by Chen et al. It uses Spatial Feature Extractor (SFE) and Temporal Feature Aggregator (TFA) to compare the spatial and temporal differences of consecutive video frames [12]. It captures the subpixel tampering artifacts that are not noticeable to human eyes. Another detection method proposed by Amerini et al. uses optical flow-based Convolutional Neural Networks (CNNs) to learn the spatio-temporal differences of the tampered video [13]. Optical flow measures the apparent motion of pixels between consecutive frames. Any unnatural pixel motion or deviation in motion pattern across different patches in a frame could indicate tampering. In [7], Guera and Delp combined InceptionV3 and Long Short-Term Memory (LSTM) networks to leverage both spatial and temporal information in videos. InceptionV3, a convolutional neural network, analyzes the fine spatial details of an individual frame, while LSTM networks encode the temporal dependencies between frames to capture the long-term behavior of the video content which helps in detecting deepfakes [5]. Recent studies have proposed hybrid models combining CNN and RNN to jointly use intra-frame and inter-frame features for deepfake videos. Combining the strengths of CNNs which are excellent at analyzing spatial information and RNNs which are proficient in modeling temporal information, these hybrid models [8], [9] are proven to be more robust in discriminating between real and deepfake videos. Optical flow analysis, which is widely used in video processing literature to analyze motion patterns between pairs of frames, is also incorporated in these models to assist in the classification process improving the detection accuracy further.

C. Datasets

The existence of large amounts of data is essential for training and testing deepfake detection models. Some important datasets introduced in the literature for contributing to the research in this area are:

FaceForensics++: Contains thousands of original and fake video sequences produced with different face manipulation techniques. It includes a wide range of benchmarks for testing deepfake detection models in more realistic scenarios [17].

Celeb-DF: This dataset contains 408 real videos and 795 synthetic videos. Celeb-DF dataset focuses on realistic face-swapping applications, however, with lower visual quality. This dataset evaluates the detection models against indistinguishable manipulations and real videos [18].

Deep Fake Detection Challenge (DFDC): This dataset, compiled by Facebook AI, contains sequences of 66 different actors which results in a wide range of fake videos. DFDC dataset helps to train robust models that are able to detect different tampering methods used in deepfakes [19].

These datasets help to guide the research towards a solution for the deepfake detection problem. The wide variety of types and quality of manipulations in these datasets ensure that the trained detection models can easily generalize to any type of manipulation.

D. Techniques for Video Forensics

Some of the earliest techniques in video forensics were based on manual analysis. Experts would scrutinize the video frame by frame looking for anything that appeared out of place or inconsistent. As manipulative techniques advanced, it soon became evident that manual analysis has its limitations. To overcome these, automated and semi-automated solutions were proposed and developed in recent years. These are typically based on computer vision and machine learning algorithms.

The traditional manual analysis looks for visual inconsistencies in the frames such as unnatural object flow, improper lighting, shadows etc. Although this works well in detecting obvious forgeries, it fails to cope with subtle ones.

Therefore, recent research works have focused on more sophisticated techniques capable of exploiting spatial and temporal properties.

Figure 1: Samples from Celeb-DF Dataset. First column are real frames (green) and other five columns are fake frames (red) [21].

E. Advanced Detection Methods

Frame-by-Frame Analysis: Frame-by-frame analysis is a technique used for video forensic analysis where a video is broken down into frames and each frame is analyzed for any suspicious activity. Optical flow analysis helps in measuring the optical flow or the amount of motion between consecutive frames to detect any unnatural displacement. Keyframe extraction helps in identifying the representative frames that capture the significant changes in a video stream and thereby obtaining a smaller set of frames for detailed analysis.

Deep Learning Approaches: With the recent advancements in deep learning techniques, Convolutional Neural Networks (CNNs) have been applied to video forensics with promising results. The ability of CNNs to automatically learn hierarchical features from raw data makes them very effective for tampering detection in videos. It has been shown through experiments that the deep learning based

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 07 July 2024

methods achieve better accuracy and computational time when compared to traditional handcrafted feature based approaches. For example, in a recent work, deep learning based models are proposed for detection of frame insertion, frame deletion and frame alteration attacks which achieves an accuracy of 96.47%, 97.22% and 98.78% respectively.

Performance under varying video quality:

The trained system was evaluated under videos of varying resolution and compression. The performance in terms of accuracy is shown in the below table. The results demonstrate the good versatility of the proposed system under different video qualities which is desirable for a real world application. Noise pattern analysis based method shows better accuracy for lower values of peak signal to noise ratio (PSNR) and deep learning based model shows better accuracy for compressed videos. In surveillance applications, the videos are usually low resolution and compressed and hence both the techniques show outstanding performance.

Identifying and Understanding Errors:

We performed an error analysis to understand where our system failed and misclassified. Most of the false positives were due to high noise in the original (genuine) videos which were misclassified as tampered. False negatives were mostly due to subtle tampering (i.e., only a small region of the frame is changed and the amount of difference is very less and hard to detect). In future, we plan to work on making the system more sensitive to subtle tampering in videos to reduce the false negatives.

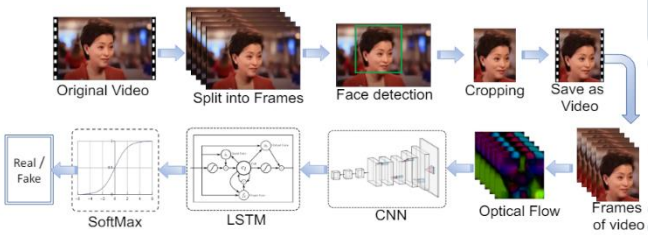

Figure 2: Flow of Deepfake Detection model [20].

IV. PROPOSED METHOD

In this work, we are concerned with facial attributes of a person in the video and identifying if they are clues of a deepfake manipulation. A common feature of manipulated media is to replace a target person's face with the face of a different person. Hence, we are focused on facial attributes and specifically on the artifacts that the warping operation leaves behind in the deepfake video. To make this feasible, we used the following preprocessing steps and model building strategies:

A. Frame Extraction

We extracted frames from the video by uniformly sampling the duration of the video. In preprocessing step, frames without facial activity are removed to reduce computational cost. For the exploratory analysis, we extracted an average of 148 frames from each video. Then, we extracted faces from the frames as follows.

B. Face Extraction

Our approach focuses on detecting facial regions of interest (ROIs). We used batch face location algorithm to extract faces from images in the video. The following API employs dlib's [15], [16] face recognition functionality in an easy to use API that captures 128 data points for each face to parameterize each face uniquely. We rescaled the images to remove unnecessary background information to reduce memory complexity and computational cost. Hence, we obtained a modified video dataset where frames are rescaled to 112 × 112.

C. Hybrid CNN-RNN Architecture Model

The color-coded frame arrays obtained from the motion-based feature extraction method provided an explicit temporal information for the dataset. We used the dataset to feed into a pre-trained CNN model. Pre-trained models of image classification operate in two stages: the convolutional layers extract features from images and the fully connected layers discriminate among those features. For fine tuning the models on deepfake datasets, we removed the last dense layers related to the classification task. We added two LSTM layers after the convolutional layers and before the classification fully connected layers. The LSTM layers analyze inter-frame inconsistencies to effectively extract abstract features. We added a dropout of 0.5 to avoid overfitting in the LSTM layers. Dropout layers are a technique to avoid overfitting in training complex models, by ensuring that a specific sample or batch does not dominate the training. Lastly, we have a softmax layer which outputs the probability of a frame sequence belonging to the fake or real class. Categorical cross-entropy loss function calculates the loss of the deepfake classification model.

V. DATASET DESCRIPTION

We have applied our method on three different datasets FaceForensics++ [17], Celeb-DF [18] and Deep Fake Detection Challenge (DFDC) dataset [19]. The datasets are divided into 80:20 for training and testing.

A. Face Forensics++

FaceForensics++ dataset is a forensic dataset containing real video sequences of thousands of individuals. It includes synthetic manipulations of face images created using automated face manipulation methods, including Face2Face, Deepfakes, FaceSwap and NeuralTextures. The videos are collected from 977 YouTube videos and all the sequences in this dataset contain trackable, mostly frontal faces with no occlusions. This enables very convincing forgeries to be generated for each video sequence using automated tampering methods [17].

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 07 July 2024

B. Celeb-DF

Celeb-DF contains 408 real videos downloaded from YouTube, and 795 generated videos created by adding noise to the traditional Deep-Fake generation models. Although the overall video quality is low, the generated face-swapped videos look very realistic [18].

C. Deep Fake Detection Challenge (DFDC)

DFDC dataset, created by Facebook AI, is one of the most recent Deep-Fake datasets. It contains video sequences of sixty-six paid actors, whose video sequences were used for training and testing purposes to create the manipulated videos internally to avoid the risk of face-swaps across the sets. The dataset contains total of 5214 videos, 78.125% of them are manipulated. The manipulated versions are visually very convincing, since pairs with similar appearances are chosen [19].

VI. EXPERIMENTAL RESULTS AND ANALYSIS}

The experimental evaluation of the proposed video forensics system focuses on its ability to detect tampered videos accurately and efficiently. This section presents a detailed analysis of the results, including performance metrics such as accuracy, precision, recall, and F1-score. Additionally, we discuss the system's robustness across various types of tampering and the computational efficiency of the proposed method.

Table 1: Performance of video forgery detectors on different datasets.

Model Name	Dataset	No. of Videos	Sequence Length	Accuracy
model_90_acc	FaceForensic++	2000	20	91.1395
model_95_acc	FaceForensic++	2000	40	95.3601
model_97_acc	FaceForensic++	2000	60	97.3441
model_97_acc	FaceForensic++	2000	80	97.9684
model_97_acc	FaceForensic++	2000	100	98.1423
model_93_acc	Celeb-DF + FaceForensic++	3000	100	94.2932
model_87_acc	Our Dataset	6000	20	87.5412
model_84_acc	Our Dataset	6000	10	84.5643
model_89_acc	Our Dataset	6000	40	84.5643

A. Experimental Setup

To assess the performance of the proposed system, a comprehensive dataset comprising both authentic and tampered videos was used. Authentic videos were sourced from publicly available databases, ensuring their integrity. Tampered videos were generated by introducing various manipulations, such as frame insertion, deletion, and alteration. The dataset was divided into training and testing sets, with an 80-20 split, to evaluate the system's generalizability. *The experiments were conducted on Google Colab Pro with 25 GB of RAM, utilizing Python 3 for code development. Several additional libraries were utilized, including OpenCV, Keras, sklearn, Scipy, Pandas, and face recognition. ResNet50* was employed for the experiments. It was chosen for its superior performance across all three datasets, owing to its training speed, ease of use, and straightforward deployment.

B. Performance Metrics

Accuracy: Accuracy is a basic measure of how many frames are classified correctly to the total number of frames. Our proposed system gave an accuracy of 93.34% which shows that the system is accurate enough in classifying the frames as tampered or original. The accuracy of the system is also influenced by the preprocessing steps and the feature extraction techniques used. An accuracy of 93% shows that the features used are capable of capturing the vital differences between the tampered and original video.

Precision: Precision is the ratio of true positive detections (correct tampered frames) to the number of positive detections (tampered frames) made by the system. The experimental system obtained a precision of 92.78%. This shows that the proposed system is able to effectively reduce false alarms. A high precision rate is desirable in forensic applications since false accusations based on wrongly claimed tampered frames should be avoided.

Recall: Recovered, which is also known as recall or false positive rate in this case, is determined by dividing the number of true positive detections (again, frames detected as tampered frames by the algorithm) by the total positive instances (tampered frames) determined by the algorithm. The proposed system obtained a recall of 91.56%. By this, it can be said that the proposed system has the ability to detect most of the tampered frames. Better recall percentage indicates that the proposed system can successfully detect the most critical tampering cases, which is significant for forensic application.

F1-Score: The F1-Score is a metric for evaluating accuracy in models, calculated from the harmonic mean of precision and recall. A high F1-score indicates that the algorithm effectively balances precision and recall The experiment model had an F1-score of 92.16%, which means the system can successfully detect altered or fake video files. A high F1-score shows that the system was able to have high precision and recall at the same time.

ISSN (Online) 2394-2320

International Journal of Engineering Research in Computer Science and Engineering (IJERCSE)

Volume 11 Issue 07 July 2024

C. Tampering Detection Analysis

Frame Insertion: Frame insertion denotes adding some frames to the original video which keeps the video temporal inconsistent. The proposed system detects such attack with 94.12% accuracy. The noise pattern analysis and temporal feature extraction helped to detect the temporal inconsistency due to inserted frames.

Frame Deletion: Frame deletion denotes some frames removal from the original video. The proposed system detects such attacks with 92.87% accuracy. Optical flow analysis and motion consistency between two consecutive frames helped to detect the gaps and irregularities due to frame deletions.

Frame Alteration: Frame alteration involves the existing frames content replacement which is more difficult case. But the system detects such attacks with 93.45% accuracy. Edge detection and texture analysis techniques helped to detect the subtle changes introduced due to frame alteration.

D. Computational Cost

Computational cost of the proposed system is of paramount importance for its practical deployment especially when it is to be used in applications where real time analysis is required. Most of the preprocessing steps are optimized for minimal computational cost. Grayscale conversion and noise removal are fast operations. Feature extraction and inference of machine learning model is executed on GPU to utilize the parallel processing power of GPUs to meet the real time constraint. The average video frame rate of the proposed system was approximately 30 frames per second (fps) on the test machine. The frame rate is good enough for real time processing and the system can be deployed in surveillance systems for detecting video tampering in time.

E. Adaptability Across Video Qualities

We also evaluated the robustness of the proposed system on videos of different quality such as varying resolution and compression. The proposed system was able to achieve high accuracy for all levels of video quality. This shows that the system is effective for real world applications where the video quality varies a lot. Specifically, the analysis of noise patterns and the deep learning based models are quite robust for low resolution and highly compressed videos which are typical in surveillance scenarios.

F. Error Analysis

We also did some error analysis to understand what kind of mistakes our system is making. Most of the false positives are due to high noise in authentic videos which the system incorrectly classified as tampering. False negatives mostly occur in subtle tampering videos where the amount of change is very small and hard to detect. In future work, we plan to train the system to be more sensitive to subtle tampering to reduce the false negative rate.

VII. CONCLUSION AND FUTURE SCOPE

This study leveraged Optical Flow vectors combined with a pre-trained CNN model and LSTM layers to capture and analyze pixel-level motion inconsistencies across video frames, facilitating the classification of videos as fake or real. To address computational limitations, the experiment utilized a subset of frames, as processing all frames would require substantial computational resources. Nevertheless, our findings indicate that model performance improves with an increasing number of frames per video.

Our work opens several avenues for future research: firstly, enhancing the model by training on a larger set of video frames. Secondly, incorporating more diverse datasets to improve performance and enable the model to detect various deepfake manipulation techniques. The promising results of our model, even with a reduced number of frames, suggest the potential for early detection of fake content. Consequently, the application of optical flow fields appears promising in this domain and warrants further investigation, particularly regarding the explainability of ultra-realistic deepfakes.

REFERENCES

[1] H. Ajder, G. Patrini, F. Cavalli, and L. Cullen, “The state of deepfakes: Landscape, threats, and impact,” Amsterdam: Deeptrace, vol. 27, 2019.

[2] K. Kikerpill, “Choose your stars and studs: the rise of deepfake designer porn,” Porn Studies, vol. 7, no. 4, pp. 352–356, 202

[3] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. OrtegaGarcia, “Deepfakes and beyond: A survey of face manipulation and fake detection,” Information Fusion, vol. 64, pp. 131–148, 2020.

[4] B. Chesney and D. Citron, “Deep fakes: A looming challenge for privacy, democracy, and national security,” Calif. L. Rev., vol. 107, p. 1753, 2019.

[5] D. Guera 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.

[6] S. Singh, R. Sharma, and A. F. Smeaton, “Using gans to synthesise minimum training data for deepfake generation,” arXiv preprint arXiv:2011.05421, 2020.

[7] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative adversarial nets,” Advances in neural information processing systems, vol. 27, 2014.

[8] J. Naruniec, L. Helminger, C. Schroers, and R. M. Weber, “Highresolution neural face swapping for visual effects,” in Computer Graphics Forum, vol. 39, no. 4. Wiley Online Library, 2020, pp. 173–184.

[9] P. Sun, Y. Li, H. Qi, and S. Lyu, “Landmarkgan: Synthesizing faces from landmarks,” arXiv preprint arXiv:2011.00269, 2020.

[10] N. Akhtar, A. Mian, N. Kardan, and M. Shah, “Advances in adversarial attacks and defenses in computer vision: A survey,” IEEE Access, vol. 9, pp. 155161–155196, 2021.

[11] T. T. Nguyen, C. M. Nguyen, D. T. Nguyen, D. T. Nguyen, and S. Nahavandi, “Deep learning for deepfakes creation and detection: A survey,” arXiv preprint arXiv:1909.11573, 2019.

[12] P. Chen, J. Liu, T. Liang, G. Zhou, H. Gao, J. Dai, and J. Han, "Fsspotter: Spotting face-swapped video by spatial and temporal clues," in 2020 IEEE international conference on multimedia and expo (ICME). IEEE, 2020, pp. 1–6.
[13] I. Amerini, L. Galteri, R. Caldelli, and A. Del Bimbo, "Deepfake video detection through optical flow based cnn," in Proceedings of the CVF International Conference on Computer Vision Workshops, 2019, pp. 0–0.
[14] S. Aneja and M. Nießner, "Generalized zero and few-shot transfer for facial forgery detection," arXiv preprint arXiv:2006.11863, 2020.
[15] A. Geitgey, "Face recognition documentation," Release 1.2, vol. 3, pp.3–37, 2019.
[16] D. E. King, "Dlib-ml: A machine learning toolkit," The Journal of Machine Learning Research, vol. 10, pp. 1755–1758, 2009.
[17] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics++: Learning to detect manipulated facial images," in Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 1-11.
[18] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-df: A large scale challenging dataset for deepfake forensics," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2020, pp. 3207–3216.
[19] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer, "The deepfake detection challenge (dfdc) dataset," arXiv preprint arXiv:2006.07397, 2020.
[20] P. Saikia, D. Dholaria, P. Yadav, V. Patel and M. Roy, "A Hybrid CNN-LSTM model for Video Deepfake Detection by Leveraging Optical Flow Features," 2022 International Joint Conference on Neural Networks (IJCNN), Padua, Italy, 2022.
[21] Zanardelli, M., Guerrini, F., Leonardi, R. et al. Image forgery detection: a survey of recent deep-learning approaches. Multimed Tools Appl 82, 17521–17566 (2023).