DFDA: An Analysis of Deep Learning Models to Detect Deepfake Videos

Munleef Bhat[1,†], Prateek Agrawal[1,2,*,†] and Charu Gupta[3,†]

[1]School of Computer Science and Engineering, Lovely Professional University, Punjab, India
[2]Shree Guru Gobind Singh Tricentenary University, Gurugram, India
[3]Department of Computer Science and Engineering, Bhagwan Parshuram Institute of Technology, Delhi, India

Abstract

Emerging technology dubbed Generative Artificial Intelligence (AI) has facilitated the fabrication of counterfeit videos and images that exhibit striking realism. This presents a significant apprehension as these simulated visuals, referred to as DeepFakes, possess the capacity to disseminate disinformation and ensnare individuals with ease. To confront this issue, scholars are employing sophisticated computational algorithms and methodologies to identify DeepFakes. This manuscript provides a comprehensive examination of the strategies employed for DeepFake detection. It delves into the integration of diverse forms of media (such as images, videos, and speech) with machine learning to discern counterfeit content. Additionally, it discusses the pivotal datasets utilized by researchers for evaluating their DeepFake detection techniques. We scrutinized conference and journal articles published on DeepFake video detection from 2015 to 2023. These articles explored diverse methodologies aimed at identifying counterfeit images, videos, and even fabricated vocalizations. They revealed that amalgamating distinct techniques, such as integrating images and videos or employing varied machine learning methodologies, can yield highly efficacious results in DeepFake detection. Moreover, the paper offers recommendations for prospective investigations to enhance DeepFake detection, thus fostering a safer cyberspace. Additionally, it introduces a novel dataset termed Celeb-DF, comprising numerous high-fidelity counterfeit videos featuring renowned personalities. This dataset is crafted to facilitate researchers in refining their techniques for detecting DeepFakes. In essence, this paper endeavours to augment ongoing endeavors to mitigate the proliferation of counterfeit content online by enhancing our capacity to identify DeepFakes.

Keywords

Deepfake, machine learning, generative AI, secure society, equality, deep learning

1. Introduction

Fueled by AI and deep learning advancements, DeepFakes [1] have emerged as a powerful instrument for manipulating digital media with unparalleled realism. Their capacity to seamlessly integrate faces into existing footage has triggered concerns across various sectors, owing to their ability to easily mislead unsuspecting viewers. These synthetic videos, often indistinguishable from genuine recordings, present a multifaceted threat to societal trust, personal privacy, and the credibility of information dissemination platforms. In the political arena, DeepFakes

ACI'23: Workshop on Advances in Computational Intelligence at ICAIDS 2023, December 29-30, 2023, Hyderabad, India
*Corresponding author.
†These authors contributed equally.
munleefbhat@gmail.com (M. Bhat); dr.agrawal.prateek@gmail.com (P. Agrawal); charu.wa1987@gmail.com (C. Gupta)

have the potential to disrupt democratic processes by fabricating speeches or events, fostering uncertainty, and undermining public trust in institutions. Socially, they could instigate turmoil or perpetuate detrimental stereotypes by falsely attributing statements or deeds to public figures. Financially, malicious entities might exploit DeepFakes for fraudulent purposes, such as impersonating executives or manipulating stock prices through fabricated announcements [2]. From a legal standpoint, the proliferation of DeepFakes poses intricate inquiries regarding authenticity, liability, and the boundaries of free expression. Courts may encounter difficulties in adjudicating cases involving digitally manipulated evidence, while individuals could find themselves unfairly targeted or defamed by synthetic content. Moreover, the malevolent exploitation of DeepFakes for generating illicit material, such as non-consensual pornography or grooming vulnerable individuals, underscores the pressing necessity for robust protections and legal frameworks. The potential psychological harm inflicted on victims and the erosion of trust in digital communications necessitate prompt and decisive action from policymakers, law enforcement agencies, and technology firms alike. As society wrestles with the ramifications of this disruptive technology, interdisciplinary collaboration and continuous research are imperative to devise effective countermeasures. Ethical considerations should steer the responsible development and deployment of AI tools, ensuring that technological advancements serve the collective welfare rather than morphing into instruments of mass deceit. Generally, humans struggle to discern between authentic videos and DeepFakes unaided by technology. DeepFakes are generated using a blend of techniques like merging, substituting, and overlaying images and video clips, crafting exceedingly realistic yet ultimately counterfeit content. Utilizing advanced AI techniques like DeepFakes and computational adversarial networks (GANs) have progressed to incorporate audio, further amplifying their authenticity. Recent Analyzing longitudinal and time data in video and audio formats, together with spatial and time data in pictures, is part of the process of trying to identify altered information. To propel the frontiers of DeepFake detection, researchers have made benchmarking datasets publicly accessible. By amalgamating these datasets with existing techniques, cutting-edge methods now leverage information fusion to robustly identify counterfeit media. While numerous surveys delve into DeepFake detection (DFD), elucidating advancements and hurdles, this paper zooms in on media modality fusion in DFD, supplementing existing critiques. It explores contemporary approaches in DFD, citing pertinent studies and benchmarking datasets alongside their findings. Additionally, it deliberates on challenges and potential future trajectories to further elevate the condition of DeepFake identification at the moment.

2. Datasets

In the contemporary digital landscape, the notion of ”Digital Transformation” has garnered global attention, offering a plethora of lucrative applications spanning from facial recognition systems to centralized data management and intelligent automation. This transformation harnesses advanced technologies to streamline everyday human tasks and bolster efficiency. Since 2018, there has been a noticeable uptick in the advancement of contemporary generative models, particularly in vision-related realms such as facial and frame synthesis, as well as tone synthesis. Acknowledging the potential harm posed by manipulated images and videos, numerous

multinational corporations (MNCs) and academic institutions have taken the lead in crafting their own synthesized datasets tailored specifically for identifying fraudulent media using deep-learning-based methodologies. Maintaining benchmark datasets regularly updated with diverse and evolving DeepFake content is imperative to ensure that detection models undergo rigorous testing against a broad range of manipulative strategies. Assessment measures are included in each benchmark dataset and research publication to help determine their dependability. for subsequent comparison with enhanced iterations of DeepFake detection algorithms [3]. Although these datasets undergo meticulous training and testing phases, benchmarking serves to demonstrate the improved efficacy of new or older approaches to detection on the updated dataset [4].

In 2018, the collective count of DeepFake videos tallied at 3,038, encompassing 1,669 counterfeit videos and 1,369 authentic videos. By 2020, this number surged to 188,154 videos, comprising 114,500 counterfeit videos and 73,654 authentic videos [5]. Remarkably, The UADFV collection was the lowest in scale, whilst the DeepFake is Detecting Challenge (DFDC)-Full Dataset was the largest repository for DeepFake databases. The size of reference databases for DeepFake detection continued to expand, with each dataset surpassing 100,000 videos by 2023, inclusive of the DF-Platter dataset. Researchers commonly employed both historical and contemporary benchmark datasets to assess DeepFake detection efficacy, ensuring equitable and comprehensive evaluations across various studies [6, 7].

Table 1
Dataset Details

Dataset	Pristine Video	Fake Video	Total Videos
FF-DF [8]	1001	1227	2228
UADFV [7]	45	48	93
DF-TIMIT [6]	333	682	1015
FF++ DF [9]	1325	1422	2747
Google DFD [18]	3633	2337	5970
DFDC-Preview Dataset [10]	1126	4432	5214
Celeb-DF [10]	512	5886	6238
DeeperForensics-1.0 [21]	50,000	10,000	60,000
DFDC-Full Dataset [22]	23,544	104,200	127,144

3. Deepfake detection in images and videos

In a general sense DeepFake methods for identification may be divided into two primary groups: frame forgery assessment for picture recognition and behavioral and geographic analysis for video classification. Li et.al [11] introduced a new method for discerning altered faces in images or videos. Their methodology cantered on a crucial element of human facial behaviour—the frequency of eye blinking—to verify physiological cues frequently absent in synthesized fraudulent videos, as depicted in Fig. 1

In this investigation, scholars explored the blinking frequencies of eyes in genuine videos versus DeepFake videos, formulating an innovative approach for DeepFake detection. The

Figure 1: Sample figure derived Celeb-DF database. The actual video frame is in the corresponding DeepFake shots in the right columns, and the column on the left five columns are made using various donor subjects.

results suggested that irregular blinking rates could indicate a fabricated or synthetic video. Fig, 1 illustrates the meticulous examination of eye blinking across frames in both authentic and DeepFake videos, utilizing computations based on the average interval between successive blinks and the average duration of blinks to ascertain authenticity. This technique comprises two stages: (a) facial recognition through image or frame analysis, identification of facial landmarks, alignment of faces, and extraction of the eye area, and (b) counting eye blinks by using features from the first stage of a long-term recurrent neuronal network (LRCN), as shown in Fig. 2. Afchar et al. [2] focused on using deep learning-inspired systems for detecting to examine mesoscopic features of pictures. They presented two recognition approaches with two different function of activation, namely Meso-4 and MesoInception-4. To support expansion, the authors used four layers of pools and convolutions in Meso-4, which was followed by a large network with an activation function for Rectified Linear Units (ReLU), as shown in Fig. 3. On the other hand, the MesoInception-4 architecture, which was used to evaluate the Face2Face and DeepFake datasets, replaced the original convolution layers with inception models, as shown in Figure 4. Excellent detection rates were demonstrated by the results, which achieved 98% for the DeepFake technology information and 95% for the Face2Face datasets.

Hinton et al. [12] introduced an innovative capsule architecture to overcome the A convolutional neuro networks' (CNNs') restrictions. Building upon this concept, Nguyen et al. [8] extended its application to detect a variety of image and video forgeries, including replay attacks, utilizing Convolutional Neural Networks (CNNs) and Capsule Networks. They integrated They use maxing out expectations methods and anticipate traffic into their system. In their methodology, video streams are initially divided into frames, followed by face detection, extraction, and resizing. The extracted faces are then processed in order to derive latent characteristics that are used as inputs by the Capsule Network via the VGG-19 system for forgery detection. Post-processing involves calculating average probabilities, resulting in a detection rate of 99.23% at the image level and 95.93% at every frame region for the DeepFake Set. A pipeline that is

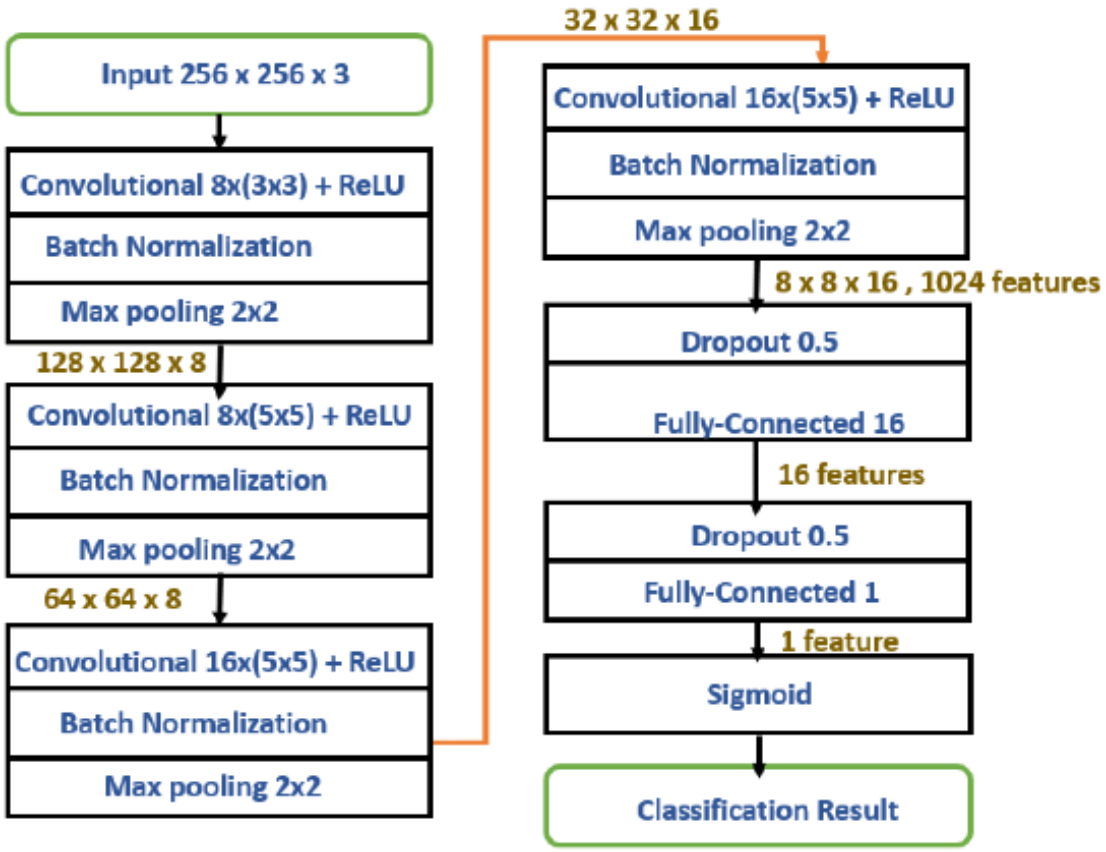

Figure 2: topology of the future recurrent convoluted neural (LRCN)

computerised was developed by Rossler et al. [9] to identify fake faces., employing tracking algorithms to trace human faces in videos or images. Subsequently, faces are analyzed by various classifiers to identify forgery, achieving high precision across multiple DeepFake datasets. Dolhansky et al. [10] used the DeepFake Detection Challenge (DFDC) sample to develop three methods for detecting using various characteristics. including TamperNet, a lightweight DNN model, designed for identifying acute level manipulations and digitally fabricated images. They achieved high accuracy on DeepFake and digitally fabricated images.

In the subsequent approach, two additional detection models were deployed utilizing XceptionNet on both facial and complete image datasets for forensic examination. These frame-oriented models implemented two levels: a per-frame detector limitation and one related corresponding to the video's captured frames per secondly. dictating the number of frames required to surpass the cutoff point per frames for categorizing a footage as counterfeit. During validation, maximizing log-WP across each fold unveiled optimal recall reminders of -3.352 for XceptionNet (full), -2.14 for XceptionNet (facial), and -3.044 for TamperNet. Korshunov et al. [13] employed the VID-TIMIT dataset to fabricate DeepFake videos using open-source GAN-based software, focusing on the influence of criteria for training and mixing on video fidelity. They generated low

Figure 3: Face swapping examples

and high-quality renditions for each subject and showcased that cutting-edge facial recognition algorithms, grounded on VGG and FaceNet, were susceptible to DeepFake videos, displaying false acceptance rates of up to 95.00%. In an audio-visual integrated system, feature extraction precedes the classification of altered videos from authentic ones making use of a two-classifier framework. A comparable approach was used by Chugh et al. [14], who used auditory properties called Mel-frequency cepstral factors (MFCC). and distances between mouth landmarks as visual features. Digital presentation attacks in DeepFake videos comprised PCA, LDA, IQMs, and SVMs. Feature blocks underwent dimensionality reduction via PCA before being fed to LSTM to distinguish altered from unaltered videos. Kaur et al. [15] assessed basic face-swap recognition techniques, noting the shortcomings of lip-sync-based methods in detecting disparities between lip movements and speech. They also shown how a Support Vector Machine (SVM) classifier in conjunction with image quality evaluations could identify DeepFake movies of superior quality having a comparably high error rate of 8.97%. A statistical technique based on hypothesis testing was developed by Agarwal and Varshney [4] to detect dishonest or faked content in photographs. As part of their approach, they had to calculate a mathematical threshold value that matched the error likelihood of identifying real or GAN-generated pictures.. Lyu [16] underscored the difficulties in identifying DeepFakes in audio recordings and high-definition videos that have been definition-synthesized. The writer highlighted concerns regarding the incapacity of current DeepFake generation techniques to precisely map hues of hair in relation to humansface. Considering the aforementioned research, this paper furnishes a succinct overview of the proposed DeepFake detection framework and stresses the necessity for future advancements in DeepFake detection methodologies. The author advocates for an adversarial perturbation-enabled model that reduces dependence on face detectors based on DNNs. The proposed detection methodology consists of two stages: an AI system for DeepFake detection

and a face detection phase with adversarial disruption. A comparison of the efficacy of several deep learning (DL) algorithms was conducted by Kumar et al. [6]. in DeepFake classification employing metric learning. They utilized a Multitask Cascaded Convolutional Neural network (MTCNN), which consists of three networks—a proposal network, a refining the internet, and output networks—for the purpose of extracting faces from pictures or recordings. Xception architecture facilitated transfer learning, while sequence classification employed LSTM and 3D convolution alongside a triplet network for metric learning. The triplet network determined how many frames a video clip has., comparing it to authentic video frames to assess realism. Given the spacing between the triplets, three different triplet-generation techniques were investigated: easy, lightly, and hard ones. embedding vectors. The proposed architecture utilized XceptionNet and MTCNN, with FaceNet for facial detection and feature extraction. With triplet loss, semi-hard triplets were able to discern amongst phony and real frames, achieving an AUC score of 99.2% on Celeb-DF and 99.71% precision on extremely dense neuron texture. data. For the purpose of recognizing fake movies, Mittal et al. [7] presented a deep-learning network structure that was influenced by the networks of Siamese rats and triplet loss. In their study, Mittal et al. evaluated the model's performance using the AUC measure on the DF-TIMIT and DFDC large-scale DFD samples. Their methodology attained a per-video AUC of 84.4% on the DFDC datasets and 96.6% on the DF-TIMIT information set, which is better than numerous state-of-the-art (SOTA) DFD approaches like Two-Stream, MesoNet, HeadPose, FWA, VA, Xception, Multi-task, Capsules, and DSP-FWA. Interestingly, it's the first technique to use video and audio hybrid modality at the same time. for DeepFake detection. In their study, they elucidated the correlation between audio and visual paradigms taken from the same footage, using speaking and face attributes identified as Sreal and Freal, correspondingly. Important characteristics were extracted from verbal and visual faces using Open-Face and PyAudio analysis. Huang et al. [17] introduced a counterfeit uperficial reconstruction technique, devoid polishing technique that needs after-processing of prior knowledge about the GAN, effectively bypassing existing state-of-the-art (SOA) techniques for detection. As of right now, GAN-based picture generating techniques frequently leave artifact patterns in synthesized images due to inherent limitations. To tackle this issue, the authors proposed techniques to identify and diminish such artifact patterns. Their approach entails training a dictionary model to capture genuine image motifs and using limited coding with linear projected for storing DeepFake pictures in a low-dimensional region. The DeepFake is image's artifact-free version's cursory repair subsequently minimizes artifact patterns. Evaluation involved testing against three SOA DFD methods—GANFingerprint, DCTA, and CNNDetector—along with 16 well-liked GAN-based methods for creating phony pictures to determine an image's legitimacy. A DFD technique based on a two-stage network structure was presented by Masi et al. [18] with the goal of separating digitally changed faces by enhancing artifacts and reducing high-level face information. This methodology is illustrated in Fig. 4.

4. Audio Modality Fusion in Deepfake detection

Comparable to DeepFakes in pictures and videos, audio material manipulation presents a significant hurdle for researchers in distinguishing genuine from counterfeit audio. An impactful

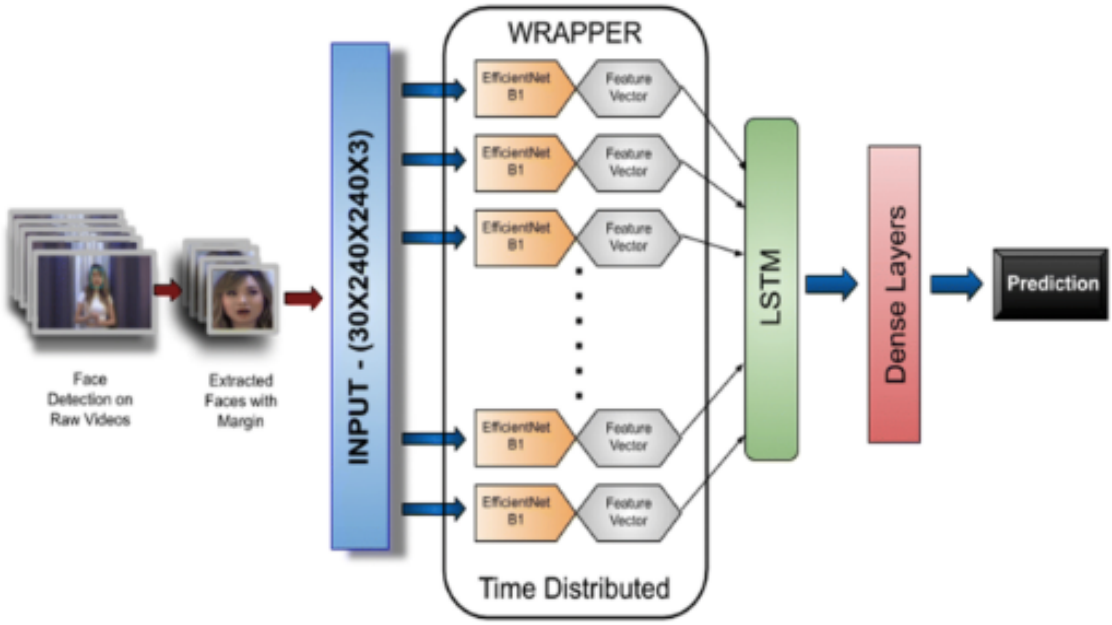

Figure 4: DFD technique utilizing a two-phase network architecture.

incident in mid-2019 involved criminals utilizing A machine learning program to imitate the speaking tone of a CEO and carry out a 243,000 USD fraudulent transfer. Systems that automate identification of speakers (ASV) are especially vulnerable to attacks including voice the conversion (VC), audio phishing, replay, and speech synthesizer (SS). which are exploited for illicit purposes. Advancements in SS and VC techniques have markedly complicated the differentiation between counterfeit and authentic speech, exacerbating the menace posed by synthetic audio and DeepFakes. This heightens the risk of misinformation influencing emotions and viewpoints, potentially culminating in organized and detrimental actions founded on false Initial perceptions. Engineers have combined ASV approaches with audio spoofed detection systems that use defensive measures rankings to discriminate between real and fake speech in order to counteract SS and VC assaults..

5. Deepfake detection methods

Li et al. [19] delved into the advancements in computational modeling and face location recognition integration like GANs and VAEs, which have markedly elevated the realism of DeepFakes in both visuals and videos. Following suit, Cozzolino et al. [9] demonstrated the utilization of DeepFake technology in forensic inquiries, employing to scrutinize categorization frameworks constructed with data mining visual anomalies and disparities. They underscored the efficacy of temporal amalgamation of convolutional representations and deep learning methodologies in identifying DeepFakes. Vignesh et al. [20] tackled the escalating menace posed by DeepFake videos, adeptly portraying contrived scenarios or personalities. The increasingly intricate generation processes of DeepFake videos present hurdles in place of conventional detecting methods.

The authors proposed a multi-attentional method that combines mechanisms of self-control, focus on space, and temporal attentiveness to overcome this. This tactic enables the model to focus on important areas and motifs while ignoring irrelevant information, allowing for the efficient extraction of local as well as international context-specific data from films.. By combining these attention processes, the suggested DeepFake detector model recognizes antiques, contradictions, or unusual patterns suggestive of DeepFake tampering. The model examines chronological and visual clues, including motion sequences, eye movements, and smiles in order to make decisions. The authors stressed how crucial it is to train the multi-attentional DeepFake detect model using large datasets that include a variety of DeepFake versions. By using this approach, the model becomes more resilient to new manipulation techniques and more broadly applicable.Transfer acquisition and domain-adaptation approaches may be utilized by the model to achieve superior performance across many DeepFake video formats. However, the authors pointed out that DeepFake detection techniques and manufacturing processes are still in combat with one another. They emphasized the necessity of ongoing study, creativity and and cooperation between academic institutions, business, and governmental bodies in order to remain ahead of hostile actors and guarantee the establishment of efficient DeepFake detectors entities. Mittal et al. [7] underscored the significance of incorporating audio cues in DeepFake detection, as discrepancies between audio and visual elements frequently arise in manipulated videos.. They suggested a combined audio-visual DeepFake detection method that simultaneously analyzes both forms in order to solve this. The model makes use of convolutional artificial neural networks (CNNs) and deep mining to gather characteristics in visual data. audio using spectrogram analysis, capturing facial expressions and speech patterns, respectively. These features are fused using using awareness or synthesis techniques to produce a shared description fed into a classification model. Training on diverse datasets enhances the model's ability to detect various DeepFake variations, resulting in improved precision and resilience to alteration methods.

In 2022, Varma and Rattani [21] addressed gender bias in DeepFake datasets by introducing the Gender-Balanced DeepFake (GBDF) dataset, tailored for Face-in-Video (FIR) DeepFake detection. This dataset aims to rectify the gender imbalance, crucial for unbiased model performance. GBDF encompasses diverse subjects, facial emotions, backdrops, and perspective adjustments showcasing real and DeepFake footage produced using a variety of editing methods. The paper delineated the collection and curation process of GBDF, ensuring It provides a thorough and well-rounded collection for studies regarding DeepFake diagnosis. It discussed annotation methods, data preprocessing, and challenges encountered in building a gender-balanced dataset. Experimental evaluations showcased GBDF's efficacy demonstrating the advantages of a balanced gender in improving detection efficiency and resilience in the training and assessment of DeepFake detectors.

6. Conclusion

The above sections discussing image and video features in DeepFake detection research have shed light on the progress made by researchers since late 2017. While significant strides have been taken to refine existing models, there remains considerable scope for further research to

improve the inspection of the pipeline's cost, effectiveness, and performance -effectiveness, and practical applicability in real-world contexts. A major challenge faced by DeepFake detection models is their limited ability to become less successful in situations where there are differences in lighting, face phrases, because and video quality. This is because the model cannot extrapolate across distinct datasets. Additionally, the presence of "unseen classes" in testing datasets compared to training datasets presents another hurdle. To tackle these challenges, researchers are exploring various methods, including incorporating concentration processes, using transferred information from learned models, and expanding training sets with a variety of specimens to improve conversion skills. These endeavors are aimed at driving forward the progress of DeepFake detection technology and its utility in real-world applications.

References

[1] R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, J. Ortega-Garcia, Deepfakes and beyond: A survey of face manipulation and fake detection, Information Fusion 64 (2020) 131–148.

[2] K. Narayan, H. Agarwal, K. Thakral, S. Mittal, M. Vatsa, R. Singh, Df-platter: multi-face heterogeneous deepfake dataset, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2023, pp. 9739–9748.

[3] S. Tariq, S. Lee, S. S. Woo, A convolutional lstm based residual network for deepfake video detection, arXiv preprint arXiv:2009.07480 (2020).

[4] S. Agarwal, L. R. Varshney, Limits of deepfake detection: A robust estimation viewpoint, arXiv preprint arXiv:1905.03493 (2019).

[5] S. Lyu, Deepfake detection: Current challenges and next steps, in: 2020 IEEE international conference on multimedia & expo workshops (ICMEW), IEEE, 2020, pp. 1–6.

[6] A. Kumar, A. Bhavsar, R. Verma, Detecting deepfakes with metric learning, in: 2020 8th international workshop on biometrics and forensics (IWBF), IEEE, 2020, pp. 1–6.

[7] T. Mittal, U. Bhattacharya, R. Chandra, A. Bera, D. Manocha, Emotions don't lie: An audio-visual deepfake detection method using affective cues, in: Proceedings of the 28th ACM international conference on multimedia, 2020, pp. 2823–2832.

[8] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, M. Nießner, FaceForensics++: Learning to detect manipulated facial images, in: International Conference on Computer Vision (ICCV), 2019.

[9] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, M. Nießner, Faceforensics++: Learning to detect manipulated facial images, in: Proceedings of the IEEE/CVF international conference on computer vision, 2019, pp. 1–11.

[10] B. Dolhansky, R. Howes, B. Pflaum, N. Baram, C. C. Ferrer, The deepfake detection challenge (dfdc) preview dataset, arXiv preprint arXiv:1910.08854 (2019).

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

[12] Y. Li, M.-C. Chang, S. Lyu, In ictu oculi: Exposing ai created fake videos by detecting eye

blinking, in: 2018 IEEE International workshop on information forensics and security (WIFS), IEEE, 2018, pp. 1–7.
[13] P. Korshunov, S. Marcel, Vulnerability assessment and detection of deepfake videos, in: 2019 International Conference on Biometrics (ICB), IEEE, 2019, pp. 1–6.
[14] K. Chugh, P. Gupta, A. Dhall, R. Subramanian, Not made for each other-audio-visual dissonance-based deepfake detection and localization, in: Proceedings of the 28th ACM international conference on multimedia, 2020, pp. 439–447.
[15] S. Kaur, P. Kumar, P. Kumaraguru, Deepfakes: temporal sequential analysis to detect face-swapped video clips using convolutional long short-term memory, Journal of electronic imaging 29 (2020) 033013.
[16] F. Vakhshiteh, R. Ramachandra, A. Nickabadi, Threat of adversarial attacks on face recognition: A comprehensive survey, arXiv preprint arXiv:2007.11709 (2020).
[17] Y. Huang, F. Juefei-Xu, R. Wang, Q. Guo, L. Ma, X. Xie, J. Li, W. Miao, Y. Liu, G. Pu, Fakepolisher: Making deepfakes more detection-evasive by shallow reconstruction, in: Proceedings of the 28th ACM international conference on multimedia, 2020, pp. 1217–1226.
[18] I. Masi, A. Killekar, R. M. Mascarenhas, S. P. Gurudatt, W. AbdAlmageed, Two-branch recurrent network for isolating deepfakes in videos, in: Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VII 16, Springer, 2020, pp. 667–684.
[19] L. Li, J. Bao, H. Yang, D. Chen, F. Wen, Advancing high fidelity identity swapping for forgery detection, in: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 5074–5083.
[20] R. U*, R. M, R. Vignesh K, T. K, Deepfake video forensics based on transfer learning, 2020. URL: http://dx.doi.org/10.35940/ijrte.F9747.038620. doi:10.35940/ijrte.f9747.038620.
[21] A. V. Nadimpalli, A. Rattani, On improving cross-dataset generalization of deepfake detectors, 2022. arXiv:2204.04285.