International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 04 April 2024
Page No: 830 - 835
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0117

Deep Fake Video Detection

Harsh Vardhan[1], Naman Varshney[2], Manoj Kiran R[3], Pradeep R[4], Dr. Latha N.R[5]
[1, 2, 3, 4] UG - Computer Science and Engineering, BMS College of Engineering, Bangalore, India.
[5]Associate Professor, Computer Science and Engineering, BMS College of Engineering, Bangalore, India.
Emails: *Harsh.cs18@bmsce.ac.in[1], naman.cs20@bmsce.ac.in[2], Manojkiran.cs21@bmsce.ac.in[3], Pradeepr.cs21@bmsce.ac.in[4], latha.cse@bmsce.ac.in[5]*

Abstract

Deep fake technology, driven by advancements in artificial intelligence, has garnered significant attention in recent years. This paper synthesizes findings from research papers on deep fake technology, focusing on its misuse and the need for further development. The abstracts of selected papers are analyzed to identify trends, methodologies, and challenges in the field. Common themes include the generation, detection, and mitigation of deep fakes, as well as their societal and ethical implications. Through interdisciplinary collaboration, researchers strive to address the risks associated with deep fake misuse while leveraging its potential for positive applications.

Keywords: *Convolutional neural networks, Deep Fake, fake face image forensics, ResNet, Long short-term memory (LSTM).*

1. Introduction

Deep fake technology, enabled by sophisticated AI algorithms, has revolutionized the creation and manipulation of audiovisual content. However, alongside its innovative potential, deep fake technology has been increasingly misused for deceptive purposes. [1] This paper explores the findings of research papers on deep fakes, shedding light on the various ways in which this technology is exploited. From fake news dissemination to malicious impersonation, the misuse of deep fakes poses significant challenges to society, underscoring the urgent need for robust detection and mitigation strategies. Despite these risks, the development of deep fake technologies remains crucial for countering their negative impacts and unlocking their positive potential [2].

2. Literature Survey

In this paper, we introduced a novel deep learning model for fake face media forensics, utilizing a multi-channel constrained convolution to extract content-excluded images and employing a pre-trained ResNet-18 model for feature extraction. Experiments were conducted on datasets manipulated by Face2Face and Deep Fake to validate the proposed model's performance, which demonstrated the highest accuracy across various video compression levels compared to baseline models. The study also involved visualizing hierarchical feature maps, CAM (Class Activation Maps), and comparing important parts for authenticity classification. Proposed Accuracy is 77.94%. In this study, we introduced a method named Deep Vision to analyze significant changes in eye blinking, utilizing machine learning, various algorithms, and a heuristic approach to detect Deep fakes generated by GANs.The proposed algorithm, [3] Deep Vision, observed Variations in blinking patterns related to gender, age, and cognitive behavior, employing machine learning techniques. The algorithm, based on previous studies, demonstrated a high accuracy of 87.5% in detecting Deep fakes and normal videos, although acknowledging a limitation related to correlations with mental illness and dopamine activity. The study suggests that while blinking patterns are correlated with mental health and nerve conduction pathways. This has a high accuracy rate of 87.5%. This paper introduces a block chain-based solution for ensuring the authenticity of digital videos through a decentralized approach, establishing a

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 04 April 2024
Page No: 830 - 835
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0117

secure and trusted traceability to the original video creator or source. The solution employs a decentralized storage system using IPFS, Ethereum name service, and a decentralized reputation system. This shows varying accuracy results between 84% to 99%. Explores the state-of-the-art in Deep fakes, focusing on detection methods. While frequency domain techniques lack accuracy, CNN-based [4] methods are effective but prone to overfitting and context-dependency. Tests with VGG-16 demonstrate the complexity of the task. Comparisons with Fakes potter and AutoGAN show the proposed approach's superior performance, achieving over 90% accuracy in most cases. The proposed MCNet is introduced as a manipulation classification network specifically designed to identify different manipulation algorithms applied to JPEG compressed images, [5] utilizing spatial, frequency, and compression domain features. MCNet employs appropriate preprocessing and network architectures for each domain learner, maximizing performance in manipulation classification. This Systematic Literature Review (SLR) compiles information from 112 studies conducted between 2018 and 2020, presenting a comprehensive overview of state-of-the-art methods for detecting Deep fake. Basic techniques and the efficacy of different detection models are discussed. The review highlights the widespread use of deep learning-based methods, particularly [6] Convolutional Neural Networks (CNN), in deep fake detection. This paper presents Enhanced-GAN, a method incorporating self-attention and normalization techniques to generate high-resolution DEEPFAKE images with enhanced class recognition compared to PGGAN. Enhanced-GAN outperforms PGGAN in terms of AM and [7] Mode scores at 128 and 256 resolutions, overcoming mode collapse during training. The synthesized DEEPFAKE data is then used for data augmentation in a U-net segmentation model, demonstrating superior performance. [8]The paper introduces a new spoofing scenario (PS) involving partially-spoofed speech segments and explores the application of spoofing detection at both the utterance and segment levels. A new database, Partial Spoof, is described, labeled at multiple temporal resolutions, and used for assessing spoofing classifiers. SSL models are introduced as an enhanced front-end, and novel neural architectures and training strategies are proposed for simultaneous, multi-resolution training. The accuracy is 77%. In this paper, we introduced a unified Gabor function capable of generating linear, elliptical, and circular Gabor filters, specifically designed for images with diverse shapes compared to traditional Gabor functions. [9] The proposed adaptive Gabor filters, unlike conventional adaptive weighted filters. The versatility of the underlying function allows it to be employed both independently in deep architectures and in conjunction with adaptive weighted filters within the same architecture. The accuracy is as high as 93.62%. Various researchers have developed deep-learning approaches to tackle the growing issue of deep fake images and videos, [10] particularly with the widespread availability of media on social networking sites. Despite the success of current deep learning approaches in detecting deep fakes, the paper concludes by highlighting the increasing quality of such content, emphasizing the need for enhanced methods. It addresses the challenge of determining the optimal number of layers and architecture for deep fake detection and notes the integration of deep fake detection tools by social media companies to mitigate the widespread impact of fake content. [11] The paper addresses deep fake detection by formulating it as a fine-grained classification problem and introduces a multi-attentional deep fake detection. Framework. The proposed framework explores discriminative local regions through multiple attention maps, enhancing texture features from shallow layers to capture subtle artifacts. [12] It aggregates low-level textural and high-level semantic features guided by attention maps. Accuracy is low at 67.44%. To address class imbalance in the dataset, the proposed model adjusts the loss function, penalizing false positives more than false negatives. The model detects and localizes forged areas in frames, highlighting them with a green outline. Evaluation is performed at pixel, frame, and video levels. Average pixel-level values for True Negative Rate (TNR), recall, and

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 04 April 2024
Page No: 830 - 835
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0117

precision are presented. CNN gives a better performance with an accuracy of 93% as compared to SVM which yields an accuracy of 75%. The proposed continual deep fake detection benchmark (CDDB) evaluates three families of continual incremental learning (CIL) methods, each with exploited variants (using BC, MC, MT) on the suggested CDDB benchmark across three scenes (EASY, HARD, LONG), utilizing state-of-the-art [13] CIL methods such as NSCIL, LRCIL, iCaRL, LUCIR, and DyTox. The evaluation includes adapting top-performing CIL methods using binary classification (BC), multi-class classification (MC), and multi-task learning (MT), employing a deep fake CNN detector as the backbone. [14] Face swapping is achieved through a novel feed-forward neural network, inspired by recent progress in artistic style transfer. The network transforms the identity of a person in a given input image to that of another person while preserving pose, facial expression, gaze direction, hairstyle, and lighting conditions. The transformation network [15] utilizes a multiscale architecture with inputs at different resolutions. It is trained on a collection of images from the target (replacement) identity and employs facial key point alignment, background/hair/skin segmentation, and a multi-image style loss. Morphed face images are created using GIMP and GAP tools with manual alignment. Detection relies on [16] features from pre-trained D-CNNs. The database has 352 bona fide and 431 morphed images. Realistic scenarios are simulated by printing and scanning digitally morphed images. Pre- and post-processing ensure image quality parity. We present a deep learning approach for image manipulation detection, utilizing a novel constrained convolutional layer to learn prediction error filters. This CNN adapts by suppressing content, learning low-level forensic features directly from data. Our proposed CNN architecture, equipped with constrained convolutional layers, accurately detects various image manipulations. Experiments demonstrate its effectiveness in detecting targeted manipulations and [17] outperforming the SRM-based approach, especially with extensive training data. In this study, an expert system is introduced to discern image authenticity, employing advanced techniques for identification of manipulated images. The research methodology involves the use of a substantial manipulated face dataset, MANFA, testing a customized deep learning model, XGB-MANFA, and achieving state-of-the-art performance with an AUC value of 93.4%. The proposed model showcases superiorities over existing systems, emphasizing high [18] performance, flexibility, robustness, and its ability to operate without specialized tools or expert knowledge. Table 1 shows the Methodology Overview Table. The study addresses the rising concern of facial manipulation in videos, proposing a network architecture utilizing five (CNNs) and (RNN) to effectively detect manipulations with low computational cost. The methodology involves an analysis of existing literature, utilizing a CNN face detector to extract face regions, employing ReLU with CNN for discriminant spatial feature extraction, achieving an average detection rate of 98% for Deep Fake movies and 95% for Face2Face videos.

Table 1 Methodology Overview Table

Paper	Methodology	Accuracy
1	Multi-channel constrained convolution, Pre-trained ResNet-18 for feature extraction, Experimentation on Face2Face and DeepFake datasets, Visualization of hierarchical feature maps and CAM.	High accuracy of 77.94% across various compression levels.
2	Machine learning and heuristic approach for Deepfake detection, Analysis of changes in eye blinking patterns related to gender, age, and cognitive behavior.	A high accuracy of 87.5% in detecting Deepfakes

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 04 April 2024
Page No: 830 - 835
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0117

3	Blockchain-based solution for video authenticity, Decentralized storage system using IPFS, Ethereum name service, and decentralized reputation system.	Accuracy ranges from 84% to 99%
4	Exploration of state-of-the-art Deepfake detection methods, Comparison with VGG-16, FakeSpotter, and AutoGAN.	Over 90% accuracy in most cases
5	Manipulation classification network for JPEG compressed images, Utilizes spatial, frequency, and compression domain features.	Random Result: 82.3%
6	Systematic Literature Review (SLR) of Deepfake detection methods, Emphasis on the use of deep learning-based methods, particularly CNNs.	Comprehensive overview of methodologies
7	Enhanced-GAN with self-attention and normalization techniques, Outperforms PGGAN in AM and Mode scores and utilizes synthesized DEEPFAKE data for data augmentation.	Superior performance demonstrated
8	Introduces a new spoofing scenario (PS), Describes the PartialSpoof database labeled at multiple temporal resolutions, and Uses SSL models and novel neural architectures for simultaneous, multi-resolution training.	Accuracy of 77%
9	Introduction of unified Gabor function for diverse Gabor filters, Versatility for independent or combined use with adaptive weighted filters.	Accuracy as high as 93.62%
10	Overview of deep-learning approaches for tackling deep fake images and videos, Discussion on the success of current methods, and increasing content quality.	Ongoing efforts to address challenges
11	Formulation of deep fake detection as a fine-grained classification problem, Introduction of a multi-attentional framework for enhanced detection	Accuracy of 67.44%
12	Adjustment of loss function to address class imbalance, Detection and localization of forged areas in frames, Evaluation at pixel, frame, and video levels	CNN achieves better performance with an accuracy of 93%
13	Introduction of continual deep fake detection benchmark (CDDB), Evaluation of three families of continual incremental learning methods, Adaptation of top-performing methods using binary classification, multi-class classification, and multi-task learning	Varying performance across scenes
14	Utilizes a feed-forward neural network for face swapping, Preserves facial features through a multiscale architecture	Demonstrates superior performance
15	Creation of morphed face images using GIMP and GAP tools with manual alignment, Utilizes features from pre-trained D-CNNs for detection	Morphed image detection accuracy of 75.6%
16	Utilizes constrained convolutional layer for prediction error filters in image manipulation detection, Adapts by suppressing content, Achieves effectiveness in detecting various manipulations	Detection accuracy of 88.9% for targeted manipulations
17	Introduction of an expert system for discerning image authenticity, Utilizes a substantial manipulated face dataset (MANFA), Testing with a customized deep learning model (XGB-MANFA)	Achieves state-of-the-art performance with an AUC value of 93.4%
18	Utilizes a network architecture with five CNNs and RNN for facial manipulation detection, Analysis of existing literature, Utilizing CNN face detector and ReLU with CNN for feature extraction	The average detection rate of 98% for DeepFake movies and 95% for Face2Face videos

Conclusions

The synthesis of research papers on deep fake technology highlights the multifaceted challenges posed by its misuse and the imperative to develop effective countermeasures. From the creation of fake news to the manipulation of public opinion, deep fakes have profound societal implications that demand urgent attention. Interdisciplinary collaboration and ongoing research efforts are essential for advancing detection and mitigation

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 04 April 2024
Page No: 830 - 835
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0117

techniques, safeguarding against the harmful effects of deep fake technology. Moreover, responsible development and deployment of deep fake technologies are crucial for harnessing their positive applications while mitigating their potential risks. Ultimately, by addressing the challenges posed by deep fake misuse, researchers can pave trustworthy digital media.

References

[1]. E. Kim and S. Cho, "Exposing Fake Faces through Deep Neural Networks Combining Content and Trace Feature Extractors," in IEEE Access, vol. 9, pp. 123493-123503, 2021, doi: 10.1109/ACCESS.2021.3110859.

[2]. T. Jung, S. Kim and K. Kim, "Deep Vision: Deepfakes Detection Using Human Eye Blinking Pattern," in IEEE Access,vol.8,pp.8314483154,2020,doi:10.1109/ACCESS.20 20.2988660.

[3]. A. H. Khalifa, N. A. Zaher, A. S. Abdallah and M. W. Fakhr, "Convolutional Neural

Network Based on Diverse Gabor Filters for Deepfake Recognition," in IEEE Access, vol.10,pp.2267822686,2022,doi:10.1109/ACCESS.2022.31 52029.

[4]. Rana, Md Nobi, Mohammad, Murali, Beddhu, Sung, Andrew, 2022/01/01 Deepfake Detection: A Systematic LiteratureReviewVL10DO10.1109/ACCESS .2022.3154404

[5]. H. R. Hasan and K. Salah, "Combating Deepfake Videos Using Blockchain and Smart Contracts," in IEEE Access, vol.7,pp.4159641606,2019,doi:10.1109/ACCESS.2019.290 5689.

[6]. N. Waqas, S. I. Safie, K. A. Kadir, S. Khan and M. H. Kaka Khel, "DEEPFAKE Image Synthesis for Data Augmentation," in IEEE Access, vol. 10, pp. 80847-80857, 2022, doi: 10.1109/ACCESS.2022.3193668.

[7]. L. Guarnera, O. Giudice and S. Battiato, "Fighting Deepfake by Exposing the Convolutional Traces on Images," in IEEE Access, vol. 8, pp. 165085-165098, 2020, doi: 10.1109/ACCESS.2020.3023037.

[8]. Lin Zhang, Xin Wang, Erica Cooper, Nicholas Evans, and Junichi Yamagishi. 2022. The Partial Spoof Database and Countermeasures for the Detection of Short Fake Speech Segments Embedded in an Utterance. IEEE/ACM Trans. Audio, Speech and Lang. Proc. 31 (2023), 813–825. https://doi.org/10.1109/TASLP.2022.3233236

[9]. -J. Yu, S. -H. Nam, W. Ahn, M. -J. Kwon and H. -K. Lee, "Manipulation Classification for JPEG Images Using Multi-Domain Features," in IEEE Access, vol. 8, pp. 210837-210854,2020,doi: 10.1109/ACCESS.2020.3037735.

[10]. A. Mary and A. Edison, "Deep fake Detection using deep learning techniques: A Literature Review," 2023 International Conference on Control, Communication and Computing (ICCC), Thiruvananthapuram, India, 2023, pp. 1-6, doi: 10.1109/ICCC57789.2023.10164881.

[11]. H. Zhao, et al., "Multi-attentional Deepfake Detection," in 2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), Nashville, TN, USA,2021pp.21852194.doi:10.1109/CVPR46437.2021.002 22

[12]. H. Mamtora, K. Doshi, S. Gokhale, S. Dholay and C. Gajbhiye, "Video Manipulation Detection and Localization Using Deep Learning," 2020 2nd International Conference on Advances in Computing, Communication Control and Networking (ICACCCN), Greater Noida, India,2020,pp.241248,doi:10.1109/ICACCCN51052.2020.9 362923.

[13]. C. Li, et al., "A Continual Deep fake Detection Benchmark: Dataset, Methods, and Essentials," in 2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV), Waikoloa, HI, USA, 2023 pp. 1339-1349.doi: 10.1109/WACV56688.2023.00139

[14]. Korshunova, W. Shi, J. Dambre and L. Theis, "Fast Face-Swap Using Convolutional Neural

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 04 April 2024
Page No: 830 - 835
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0117

Networks," in 2017 IEEE International Conference on Computer Vision (ICCV), Venice,Italy,2017pp.36973705.doi:10.1109/ICCV. 2017.397

[15]. R. Raghavendra, K. B. Raja, S. Venkatesh and C. Busch, "Transferable Deep-CNN Features for Detecting Digital and Print-Scanned Morphed Face Images," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, USA, 2017, pp. 1822-1830, doi: 10.1109/CVPRW.2017.228.

[16]. B. Bayar and M. C. Stamm, "Constrained Convolutional Neural Networks: A New Approach towards General Purpose Image Manipulation Detection," in IEEE Transactions on Information Forensics and Security, vol. 13, no. 11, pp. 2691-2706, Nov. 2018, doi: 10.1109/TIFS.2018.2825953.

[17]. L. Minh Dang, Syed Ibrahim Hassan, Suhyeon Im, Hyeonjoon Moon, Face image manipulation detection based on a convolutional neural network, Expert Systems with Applications, Volume 129, 2019, Pages 156-168, ISSN 0957-4174, https://doi.org/10.1016/j.eswa.2019.04.005.

[18]. Awotunde JB, Jimoh RG, Imoize AL, Abdulrazaq AT, Li C-T, Lee C-C. An Enhanced Deep Learning-Based Deep Fake Video Detection and Classification System. Electronics.2023; 12(1):87.https://doi.org/10.3390/electronics120