IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

Robust Deepfake Detection System with Deep Learning Techniques

Yanamala Bhuvaneswari[1], Sanjana Samba[2], Nemani Hiranmayai[3] and Dr Halavath Balaji[4]
Department of Computer Science and Engineering[1,2,3,4]
Sreenidhi Institute of Science and Technology, Hyderabad, India

Abstract: *This comprehensive study investigates the pervasive issue of deep fakes within the context of deep learning applications, focusing on their detection and production. Utilizing a diverse array of deep learning algorithms, including InceptionResnetV2, VGG19, CNN, Xception, InceptionV3, EfficientNetB1, DenseNet121, Hybrid Model, LSTM, ResNext-LSTM, and MRI-GAN, the research systematically evaluates their effectiveness in detecting deep fakes. Results reveal varying levels of accuracy, with Xception emerging as the most precise algorithm, achieving an accuracy of 99.32%. Notably, InceptionResnetV2 and DenseNet121 also demonstrate robust performance, with accuracies surpassing 99%. However, certain models like VGG19 and LSTM exhibit lower accuracy rates, underscoring the need for further refinement. These findings underscore the urgent necessity for robust detection mechanisms amidst the proliferation of malicious deep fakes, safeguarding against potential societal ramifications such as misinformation and privacy breaches.*

Keywords: Deep Learning, Fake Detection, InceptionResnetV2, VGG19, CNN, and Xception

I. INTRODUCTION

Deepfake technology, driven by advanced machine learning techniques, has emerged as a powerful tool for creating highly convincing fake videos and images by seamlessly superimposing one person's likeness onto another. This has raised significant concerns about the potential misuse of deepfakes for malicious purposes, such as spreading misinformation or manipulating public perception.

In response to this growing threat, researchers and technologists have turned to deep learning approaches to develop effective deepfake detection methods. Deep learning, particularly convolutional neural networks (CNNs) and recurrent neural networks (RNNs), has shown promise in identifying subtle inconsistencies and artifacts present in deepfake content. By leveraging the power of neural networks to learn complex patterns and features, these detection models aim to distinguish between genuine and manipulated media.

Several studies have contributed to the development of deepfake detection techniques. Notably, research by H. Li et al. introduced a deep learning-based method utilizing facial action units to identify manipulated facial expressions in videos (Li et al., 2020). Additionally, work by A. Rossler et al. proposed using deep learning to analyze subtle head movements and blinking patterns to uncover anomalies indicative of deepfake content (Rossler et al., 2019).

This introduction explores the burgeoning field of deepfake detection through deep learning, highlighting the urgency and significance of addressing the challenges posed by this rapidly evolving technology. The subsequent sections delve into specific methodologies, advancements, and challenges associated with deepfake detection, shedding light on the ongoing efforts to safeguard the integrity of digital media in an era dominated by sophisticated AI-generated manipulations.

This study explores deep learning algorithms for detecting deep fakes, addressing the escalating threat of misinformation and privacy breaches. By evaluating diverse models including InceptionResnetV2, VGG19, CNN, Xception, and others, it aims to develop robust detection mechanisms to mitigate the societal impact of maliciously manipulated content.The proliferation of deep fakes, convincingly altered videos/images, poses a grave threat in various domains, including misinformation dissemination and privacy breaches. This study addresses the pressing need to comprehend and counteract the malicious applications of deep fakes, emphasizing the urgency in developing effective detection methodologies.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

II. LITERATURE SURVEY

[9] The surge in the use of deepfake technology has fueled widespread apprehension due to its potential for malicious activities. To counteract this, deepfake detection has become a crucial area of research. While existing datasets like Deepfake Detection and FaceForensics++ have significantly advanced detection methods, they often rely on videos with volunteer actors in controlled environments, limiting their representation of real-world scenarios. To address this gap, this paper presents the WildDeepfake dataset, encompassing 7,314 face sequences from 707 deepfake videos sourced entirely from the internet. In contrast to previous datasets, Wild Deepfake aims to mirror the diversity and complexity of real-world deepfakes found online. The dataset's unique composition poses a more formidable challenge for deepfake detection algorithms, as it diverges from controlled settings and popular deepfake software. The authors conduct a comprehensive evaluation of baseline detection networks on both traditional datasets and WildDeepfake, revealing the latter's increased difficulty and reduced detection performance. To enhance detection capabilities, the paper introduces two Attention-based Deepfake Detection Networks (ADDNets), leveraging attention masks on real and fake faces. The proposed ADDNets demonstrate empirical effectiveness not only on established datasets but crucially on the more challenging WildDeepfake dataset, reinforcing their potential for combating real-world deepfake threats. This research contributes to the ongoing development of robust deepfake detection mechanisms capable of addressing the evolving landscape of online deepfake content.

[17] This study introduces an innovative approach to deepfake detection by focusing on the consistency of source features within forged images. The underlying hypothesis posits that distinct source features can persist and be discerned even after undergoing advanced deepfake generation processes. The proposed method, termed pair-wise self-consistency learning (PCL), employs Convolutional Neural Networks (ConvNets) for representation learning, aiming to extract and identify these source features indicative of deepfake manipulation. Complementing PCL is a novel image synthesis technique known as the inconsistency image generator (I2G), which generates richly annotated training data to facilitate the training of PCL. Through rigorous experimentation across seven prominent datasets, the models developed in this study exhibit notable improvements in deepfake detection. In the in-dataset evaluation, the average Area Under the Curve (AUC) increases from 96.45% to 98.05%, surpassing the current state-of-the-art performance. Furthermore, in the cross-dataset evaluation, the AUC enhances from 86.03% to 92.18%. These results underscore the efficacy of the proposed approach, demonstrating its capacity to enhance the accuracy and reliability of deepfake detection models across diverse datasets. The research contributes valuable insights and methodologies to the ongoing efforts in fortifying defenses against the proliferation of deceptive deepfake content.

[24] In response to the formidable challenge posed by the proliferation of fake videos, particularly those generated by advanced generative adversarial networks, this paper introduces a novel approach for detecting deepfake videos. Leveraging the state-of-the-art Attribution-Based Confidence (ABC) metric, the proposed method operates without the need for access to training data or the calibration model on validation data. Unlike traditional methods, the ABC metric enables inference solely based on the availability of the trained model. The methodology involves training a deep learning model exclusively on original videos, and subsequently employing the ABC metric to determine the authenticity of a given video. This metric generates confidence values, and for original videos, it establishes a threshold with confidence values exceeding 0.94. The utilization of the ABC metric provides a streamlined and efficient means of discerning between genuine and manipulated videos, presenting a promising avenue for deepfake detection without the necessity of accessing extensive training or validation datasets. This paper's contribution lies in its innovative application of the ABC metric, showcasing its effectiveness in distinguishing between authentic and deepfake videos. By focusing on attribution-based confidence, this approach represents a valuable addition to the arsenal of tools aimed at mitigating the challenges posed by the rapid evolution of deceptive video manipulation techniques.

[28] This research addresses a critical vulnerability in deepfake detection systems by demonstrating the efficacy of adversarial perturbations in deceiving common detectors. Employing the Fast Gradient Sign Method and the Carlini and Wagner L2 norm attack in both blackbox and whitebox scenarios, the study creates adversarial perturbations that significantly enhance deepfake images, resulting in a stark decrease in detection accuracy. While detectors achieved over 95% accuracy on unaltered deepfakes, their performance plummeted to less than 27% accuracy when faced with perturbed deepfakes. The paper also investigates two potential enhancements to deepfake detectors. Firstly, Lipschitz regularization is explored, constraining the gradient of the detector with respect to the input to boost robustness against

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

input perturbations. This regularization improves the detection of perturbed deepfakes, showcasing a noteworthy 10% accuracy increase in the blackbox setting. Secondly, the Deep Image Prior (DIP) defense is introduced, which utilizes generative convolutional neural networks to remove perturbations in an unsupervised manner. The DIP defense proves effective, achieving 95% accuracy on perturbed deepfakes that originally fooled the detector, while retaining a 98% accuracy rate in other cases on a 100-image subset. This research sheds light on the importance of fortifying deepfake detectors against adversarial attacks, offering valuable insights into potential strategies for enhancing the resilience of these systems in the face of evolving adversarial threats.

[26] In response to the escalating threat posed by highly realistic deepfake content generated by advanced technologies like Generative Adversarial Networks (GANs), this paper introduces DeepfakeStack, a robust deep ensemble-based learning technique for detecting manipulated videos. The proliferation of deepfake technology has given rise to numerous illicit applications, such as deceptive propaganda, cybercrimes, and political campaigns, emphasizing the critical need for effective countermeasures. DeepfakeStack leverages recent advancements in deep learning models to create a comprehensive solution for detecting manipulated multimedia. By combining a series of state-of-the-art classification models into an ensemble, DeepfakeStack forms an enhanced composite classifier. Experimental results demonstrate the superior performance of DeepfakeStack, outperforming other classifiers with an impressive accuracy of 99.65% and an AUROC (Area Under the Receiver Operating Characteristic) score of 1.0 in deepfake detection. These findings highlight the effectiveness of the proposed method and position DeepfakeStack as a promising tool for building real-time deepfake detectors, offering a robust defense against the misuse of hyper-realistic multimedia in various illicit activities. The research provides a significant contribution to the ongoing efforts in developing advanced technologies to counteract the escalating challenges associated with the deceptive manipulation of audio and video content.

III. METHODOLOGY

i) Proposed Work

The proposed system aims to combat the proliferation of deep fakes by integrating various deep learning algorithms for detection. Leveraging models such as InceptionResnetV2, VGG19, CNN, Xception, InceptionV3, EfficientNetB1, DenseNet121, Hybrid Model, LSTM, ResNext-LSTM, and MRI-GAN, the system will comprehensively analyze multimedia content to identify potential manipulations. Each algorithm offers unique strengths, from Xception's high accuracy to LSTM's sequence modeling capabilities. By combining these algorithms, the system aims to enhance detection accuracy across different types of deep fakes, including videos and images. The system's architecture will incorporate a modular design to facilitate scalability and flexibility, allowing for easy integration of new algorithms or improvements. Additionally, it will prioritize real-time processing capabilities to swiftly identify and mitigate the dissemination of maliciously fabricated content, thereby safeguarding against the societal consequences of deep fake proliferation.

ii) System Architecture

The deepfake detection system employs a multi-step process: importing deepfake videos, segmenting them into frames, conducting exploratory data analysis (EDA) with visualization, resizing images, and utilizing an image data generator. Deep learning algorithms including InceptionResNetV2, VGG19, CNN, and Xception are applied for detection. Each frame undergoes independent processing by these algorithms to extract distinctive features. The system evaluates detection performance using metrics like accuracy, precision, recall, F1-score, specificity, sensitivity, MAE, and MSE. This systematic architecture ensures efficient and comprehensive deepfake detection, combining preprocessing, exploratory analysis, and advanced deep learning techniques for enhanced efficacy.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

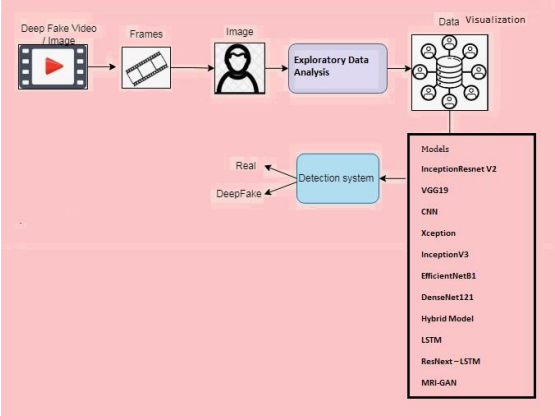

Fig. 1. System Architecture

iii) Algorithms:

Inception ResnetV2: Inception-ResNetV2 is a deep convolutional neural network architecture that combines the concepts of both Inception and ResNet. It aims to achieve better performance and efficiency by utilizing residual connections and inception modules.

VGG19: VGG19 is a convolutional neural network architecture with 19 layers, developed by the Visual Geometry Group at the University of Oxford. It is renowned for its simplicity and effectiveness in image classification tasks.

CNN (Convolutional Neural Network): CNN is a deep learning algorithm commonly used for image recognition and classification tasks. It comprises multiple layers of convolutional and pooling operations, followed by fully connected layers for high-level feature extraction and classification**.**

Xception: Xception is a depthwise separable convolutions network and an extention of Inception architecture. It aims to capture spatial and channel-wise correlations in feature maps efficiently, leading to improved performance and reduced computational complexity.

InceptionV3: InceptionV3 is another variant of the Inception architecture, designed to achieve better performance and efficiency. It introduces factorized convolutional layers and improved dimensionality reduction techniques.

EfficientNetB1: EfficientNetB1 is part of the EfficientNet family of neural network architectures, known for achieving state-of-the-art performance with significantly fewer parameters compared to traditional models. It employs a compound scaling method to balance model size and computational efficiency.

DenseNet121: DenseNet121 is a densely connected convolutional neural network architecture. It facilitates feature reuse by establishing direct connections between all layers within a dense block, leading to enhanced feature propagation and gradient flow.

Hybrid Model: The Hybrid Model is likely a custom architecture combining elements of different neural network architectures tailored to the specific requirements of the task at hand.

LSTM (Long Short-Term Memory): LSTM is a type of recurrent neural network (RNN) architecture, designed to capture long-term dependencies in sequential data. It includes specialized memory cells capable of retaining information over extended time intervals, making it well-suited for tasks involving sequential data such as text and time series analysis.

ResNext - LSTM: This is likely a combination of a Residual Neural Network (ResNet) architecture with an LSTM component, aiming to leverage the strengths of both architectures for tasks involving both image processing and sequential data analysis.

MRI-GAN: MRI-GAN is a type of generative adversarial network (GAN) specifically designed for tasks related to magnetic resonance imaging (MRI) data. GANs consist of two neural networks, a generator and a discriminator, trained simultaneously to generate realistic synthetic data. In the case of MRI-GAN, it may be used for tasks such as image denoising, reconstruction, or synthesis in the context of MRI scans. synthesis in the context of MRI scans.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

IV. EXPERIMENTAL RESULTS

Accuracy:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

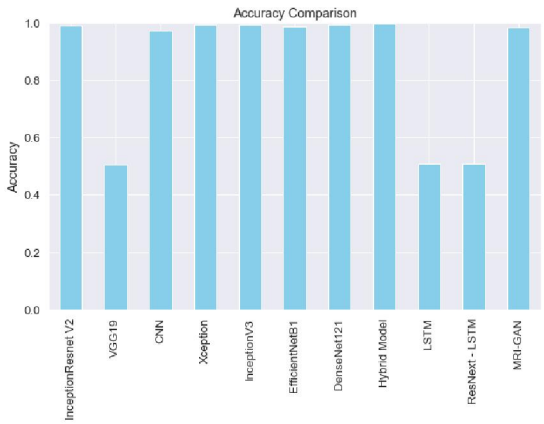

Fig 2 Accuracy comparison graph

Precision:

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

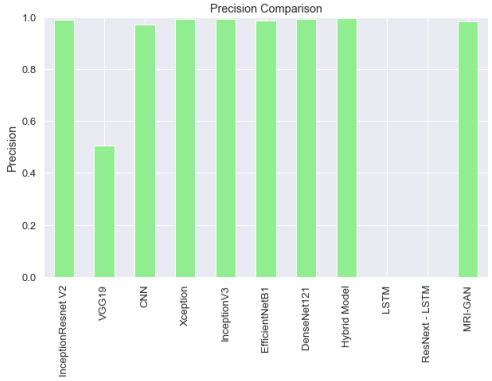

Fig 3 Precision comparison graph

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

Recall:

$$Recall = \frac{TP}{TP + FN}$$

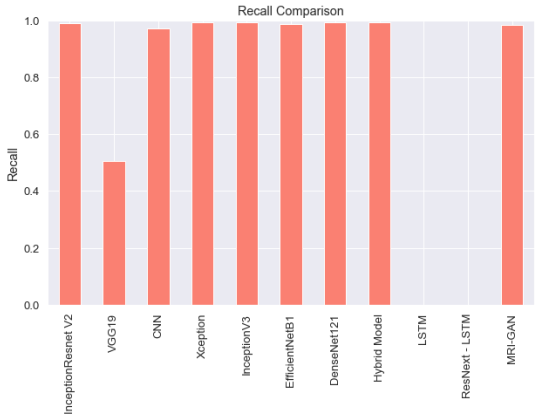

Fig 4 Recall comparison graph

F1-Score:

$$\textbf{F1 Score} = \frac{2}{\left(\frac{1}{\text{Precision}} + \frac{1}{\text{Recall}}\right)}$$

$$\textbf{F1 Score} = \frac{2 \times \text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

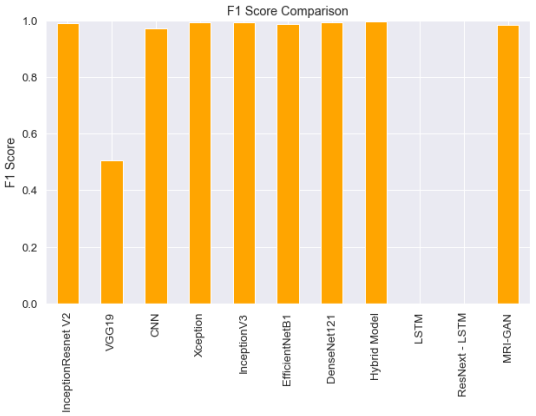

Fig 5 F1Score comparison graph

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

Sensitivity:

$$\text{Sensitivity} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Negatives}}$$

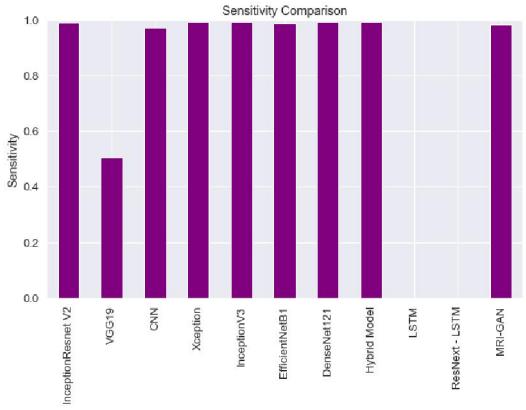

Fig 6 Sensitivity comparison graph

Specificity:

$$\text{Specificity} = \frac{\text{True Negatives}}{\text{True Negatives} + \text{False Positives}}$$

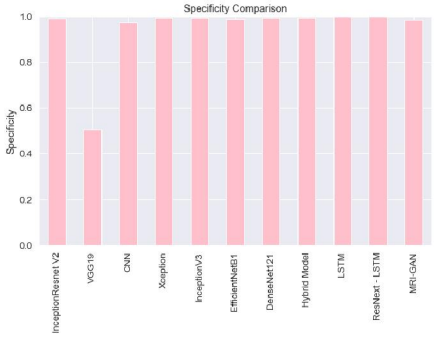

Fig 7 Specificity comparison graph

	Accuracy	Recall	Precision	F1 Score	Sensitivity	\
InceptionResnet V2	0.989809	0.989837	0.989837	0.989837	0.989837	
VGG19	0.505987	0.505904	0.505904	0.505904	0.505904	
CNN	0.972229	0.972107	0.972107	0.972107	0.972107	
Xception	0.993248	0.993267	0.993267	0.993267	0.993267	
InceptionV3	0.992866	0.992886	0.992886	0.992886	0.992886	
EfficientNetB1	0.987389	0.987424	0.987424	0.987424	0.987424	
DenseNet121	0.993503	0.993455	0.993455	0.993455	0.993455	
Hybrid Model	0.994777	0.994090	0.996754	0.995406	0.994090	
LSTM	0.507516	0.000000	0.000000	0.000000	0.000000	
ResNext - LSTM	0.507516	0.000000	0.000000	0.000000	0.000000	
MRI-GAN	0.985350	0.985391	0.985391	0.985391	0.985391	

	Specificity	MAE	\
InceptionResnet V2	0.989837	<function mae at 0x000002DEB5D4B168>	
VGG19	0.505904	0.497876	
CNN	0.972107	0.046275	
Xception	0.993267	0.01206	
InceptionV3	0.992886	0.012127	
EfficientNetB1	0.987424	0.020828	
DenseNet121	0.993455	0.010442	
Hybrid Model	0.994090	0.017597	
LSTM	1.000000	0.492226	
ResNext - LSTM	1.000000	0.62148	
MRI-GAN	0.985391	0.029025	

	MSE
InceptionResnet V2	<function mse at 0x000002DEB5D4BCA8>
VGG19	0.250147
CNN	0.021048
Xception	0.005788
InceptionV3	0.005822
EfficientNetB1	0.010759
DenseNet121	0.005295
Hybrid Model	0.007279
LSTM	0.492226
ResNext - LSTM	0.639702

Fig 9 Performance evaluation table

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

V. CONCLUSION

In conclusion, the comprehensive evaluation of deep learning algorithms for deep fake detection underscores the critical importance of robust mechanisms to combat the pervasive threat of manipulated multimedia content. Results reveal varying levels of effectiveness among the algorithms studied, with Xception emerging as the most accurate, achieving a remarkable accuracy rate of 99.32%. Additionally, models such as InceptionResnetV2 and DenseNet121 demonstrate commendable performance, surpassing the 99% accuracy threshold. However, challenges persist, as evidenced by the lower accuracy rates of certain models like VGG19 and LSTM, highlighting areas for further research and refinement. Despite these challenges, the study elucidates the potential of deep learning in addressing the pressing concerns posed by deep fakes, including the dissemination of fake news and the exploitation of public figures. Moving forward, it is imperative to continue advancing detection technologies, fostering collaboration between researchers, industry stakeholders, and policymakers to develop comprehensive strategies for mitigating the societal impact of maliciously manipulated content and preserving the integrity of digital media.

VI. FUTURE SCOPE

The future scope of this research lies in continual advancements to counter emerging deep fake threats. Further exploration of innovative algorithms and the integration of evolving technologies, such as reinforcement learning, could enhance detection capabilities. Collaborative efforts across academia, industry, and policymakers are crucial to developing standardized protocols for deep fake detection. Additionally, ongoing research could delve into real-time detection systems and the integration of explainability features to increase transparency and user trust in deep fake detection technologies.

REFERENCES

[1] M. Mirza and S. Osindero, "Conditional generative adversarial nets," arXiv preprint arXiv:1411.1784, 2014.
[2] Y. Bengio, P. Simard, and P. Frasconi, "Long short-term memory," IEEE Trans. Neural Netw, vol. 5, pp. 157–166, 1994.
[3] I. Goodfellow, Y. Bengio, and A. Courville, Deep learning. MIT press, 2016.
[4] S. Hochreiter, "Ja1 4 rgenschmidhuber (1997)."long short-term memory"," Neural Computation, vol. 9, no. 8.
[5] M. Schuster and K. Paliwal, "Networks bidirectional reccurent neural," IEEE Trans Signal Proces, vol. 45, pp. 2673–2681, 1997.
[6] J. Hopfield et al., "Rigorous bounds on the storage capacity of the dilute hopfield model," Proceedings of the National Academy of Sciences, vol. 79, pp. 2554–2558, 1982.
[7] Y. Wu, M. Schuster, Z. Chen, Q. V. Le, M. Norouzi, W. Macherey, M. Krikun, Y. Cao, Q. Gao, K. Macherey, et al., "Google's neural machine translation system: Bridging the gap between human and machine translation," arXiv preprint arXiv:1609.08144, 2016.
[8] L. Nataraj, T. M. Mohammed, B. Manjunath, S. Chandrasekaran, A. Flenner, J. H. Bappy, and A. K. Roy-Chowdhury, "Detecting gan generated fake images using co-occurrence matrices," Electronic Imaging, vol. 2019, no. 5, pp. 532–1, 2019.
[9] B. Zi, M. Chang, J. Chen, X. Ma, and Y.-G. Jiang, "Wilddeepfake: A challenging real-world dataset for deepfake detection," in Proceedings of the 28th ACM international conference on multimedia, 2020, pp. 2382– 2390.
[10] H. A. Khalil and S. A. Maged, "Deepfakes creation and detection using deep learning," in 2021 International Mobile, Intelligent, and Ubiquitous Computing Conference (MIUCC). IEEE, 2021, pp. 1–4.
[11] J. Luttrell, Z. Zhou, Y. Zhang, C. Zhang, P. Gong, B. Yang, and R. Li, "A deep transfer learning approach to fine-tuning facial recognition models," in 2018 13th IEEE Conference on Industrial Electronics and Applications (ICIEA). IEEE, 2018, pp. 2671–2676.
[12] S. Tariq, S. Lee, H. Kim, Y. Shin, and S. S. Woo, "Detecting both machine and human created fake face images in the wild," in Proceedings of the 2nd international workshop on multimedia privacy and security, 2018, pp. 81–87.
[13] N.-T. Do, I.-S. Na, and S.-H. Kim, "Forensics face detection from gans using convolutional neural network," ISITC, vol. 2018, pp. 376–379, 2018.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 6, April 2024

[14] X. Xuan, B. Peng, W. Wang, and J. Dong, "On the generalization of gan image forensics," in Chinese conference on biometric recognition. Springer, 2019, pp. 134–141.

[15] P. Yang, R. Ni, and Y. Zhao, "Recapture image forensics based on laplacian convolutional neural networks," in International Workshop on Digital Watermarking. Springer, 2016, pp. 119–128.

[16] B. Bayar and M. C. Stamm, "A deep learning approach to universal image manipulation detection using a new convolutional layer," in Proceedings of the 4th ACM workshop on information hiding and multimedia security, 2016, pp. 5–10.

[17] T. Zhao, X. Xu, M. Xu, H. Ding, Y. Xiong, and W. Xia, "Learning selfconsistency for deepfake detection," in Proceedings of the IEEE/CVF international conference on computer vision, 2021, pp. 15 023–15 033.

[18] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: a compact facial video forgery detection network," in 2018 IEEE international workshop on information forensics and security (WIFS). IEEE, 2018, pp. 1–7.

[19] E. Sabir, J. Cheng, A. Jaiswal, W. AbdAlmageed, I. Masi, and P. Natarajan, "Recurrent convolutional strategies for face manipulation detection in videos," Interfaces (GUI), vol. 3, no. 1, pp. 80–87, 2019.

[20] Y. Li, M.-C. Chang, and 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.

[21] J. Donahue, L. Anne Hendricks, S. Guadarrama, M. Rohrbach, S. Venugopalan, K. Saenko, and T. Darrell, "Long-term recurrent convolutional networks for visual recognition and description," in Proceedings of the IEEE conference on computer vision and pattern recognition, 2015, pp. 2625–2634.

[22] Y. Li and S. Lyu, "Exposing deepfake videos by detecting face warping artifacts," arXiv preprint arXiv:1811.00656, 2018.

[23] X. Yang, Y. Li, and S. Lyu, "Exposing deep fakes using inconsistent head poses," in ICASSP 2019-2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2019, pp. 8261–8265.

[24] S. Fernandes, S. Raj, R. Ewetz, J. S. Pannu, S. K. Jha, E. Ortiz, I. Vintila, and M. Salter, "Detecting deepfake videos using attributionbased confidence metric," in Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2020, pp. 308– 309.

[25] U. A. Ciftci, I. Demir, and L. Yin, "Fakecatcher: Detection of synthetic portrait videos using biological signals," IEEE transactions on pattern analysis and machine intelligence, 2020.

[26]M. S. Rana and A. H. Sung, "Deepfakestack: A deep ensemblebased learning technique for deepfake detection," in 2020 7th IEEE international conference on cyber security and cloud computing (CSCloud)/2020 6th IEEE international conference on edge computing and scalable cloud (EdgeCom). IEEE, 2020, pp. 70–75.

[27] S.-Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "Cnngenerated images are surprisingly easy to spot... for now," in Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, 2020, pp. 8695–8704.

[28] A. Gandhi and S. Jain, "Adversarial perturbations fool deepfake detectors," in 2020 International joint conference on neural networks (IJCNN). IEEE, 2020, pp. 1–8.