ISSN: 2454-9940

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

E-Mail :
editor.ijasem@gmail.com
editor@ijasem.org

www.ijasem.org

INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT

ISSN: 2454-9940
www.ijasem.org
Vol 20, Issuse.2 April 2026

Deep fake Shield AI: AI-Driven Deepfake Video Forensic Detection

[1]Kundugari Nikhitha,[2]Sujatha G,

[1]M.Tech Scholar, Dept. of CSE, Malla Reddy Technical Campus,Malla Reddy Vishwavidyapeeth, Maisammaguda,Hyderabad, Telangana 500100, India.

Mail id: nikhithakundugari@gmail.com

[2]Assistant Professor, Dept. of CSE, Malla Reddy Technical Campus,Malla Reddy Vishwavidyapeeth, Maisammaguda,Hyderabad, Telangana 500100, India.

Mail id: sujathamantra@gmail.com

ABSTRACT:

Modern methods for making and editing multimedia content may guarantee an impressive level of realism. Deep Fake is an incredibly lifelike generative deep learning system that alters or generates facial characteristics to the point that they are hard to tell apart from the actual thing. The advancement of this technology has opened up many new possibilities, including enhancements to visual effects in movies, TV shows, and video games, as well as new avenues for criminal activity, such as the creation of false information by imitating renowned persons. Research on Deep Fake detection utilizing deep neural networks (DNNs) has seen a surge in interest as a means to identify and categorize DeepFakes. To put it simply, Deep Fake is media regeneration achieved by inserting or altering data within a DNN model.

OBJECTIVE:

With these goals in mind, the RNN-based deep fake video detection system may be an effective weapon in the fight against disinformation and for better online safety.

OVERVIEW:

Deep fake videos are forms of synthetic media that use the likeness of another person—typically an actor—and manage to pass it off as real. There has to be a reliable way to identify these films since they have caused major ethical and security problems. Because of their superiority in processing sequential data, Recurrent Neural Networks (RNNs) offer hope as a method for identifying deep fake films.

Preprocessing:

- Extract frames from videos.
- Normalize and resize frames for consistency.

Feature Extraction:

- Use a Convolutional Neural Network (CNN) to extract spatial features from each frame.

Temporal Analysis:

- Feed the sequence of extracted features into an RNN (LSTM/GRU).
- The RNN captures temporal dependencies and inconsistencies between frames.

Classification:

- The output of the RNN is fed into a dense layer for classification.
- The final output layer predicts the probability of the video being a deep fake.

INTRODUCTION:

"Deepfake" describes a subset of videos in which the protagonist's appearance or voice has been doctored in order to make them seem different. The employment of deep learning algorithms is common in this process, which involves changing the person's appearance, altering their voice, or changing what they say [1]. Deepfake creation is now a breeze, and it's becoming harder and harder to tell the edited content apart from the original, all because of the continual development of new technologies and the tremendous expansion of neural networks. This paves the way for exciting new

possibilities, but it also has the potential to cause problems, especially if the people involved aren't happy with the created material. Under the PRIN 2017 program, the Italian Ministry of Education, University, and Research financed this effort via the PREMIER project. Based on research supported by DARPA and AFRL under agreement number FA8750-20-2-1004, this information was compiled. Regardless of any copyright indication, the United States Government may reproduce and distribute reprints for Governmental purposes. No part of this document should be taken as an official statement or endorsement by DARPA, the Air Force Research Laboratory (AFRL), or the United States Government; all opinions and findings are solely those of the authors. There have been several instances of deepfakes being exploited maliciously, such as in the dissemination of false news [2] and instances of fraud [3]. Some moral questions about AI's use arose as a result of this [4]. Impersonating a certain voice has become more feasible because to the fast development of Text-To-Speech (TTS) synthesis and Voice Conversion (VC) methods in the realm of audio editing. Since this is the case, creating methods to identify fake or legitimate multimedia information has taken on critical relevance [5]. Researchers have been heading in this direction for a while now, with a plethora of methods that can evaluate audio and video content [6, 7]. To train a neural network classifier, [9] uses linear filter banks fed into a Resnet, while [10] uses long-term characteristics to tell the difference between genuine and fraudulent audio. While [12] uses the traces left by temporal scaling to distinguish false audio signals, [11] recently found audio deepfakes using long-term and short-term predictive characteristics. Using sentiment analysis, this publication suggests a way to determine whether a voice signal is real or deepfake. Since creating convincingly faked speech is essential for making realistic and engaging synthetically produced videos, we zero in on the audio component. We use audio embeddings with semantic meaning to do this classification, as using semantic features has been effective for deepfake analysis of both audio and video [13, 14]. Following the lead of [13], we presume that deepfake generators are capable of synthesizing basic speech traits but are unable to replicate more nuanced features like emotions. The phenomenon of speech emotion recognition (SER) has attracted a lot of attention recently; it's the challenge of automatically deducing the speaker's emotional state from an audio recording of their speech. For this goal, several networks have been developed, some focusing only on audio (speech) [15, 16] and others using multi-modal techniques [17, 18]. Here, we provide a new transfer-learning approach that feeds a deepfake classifier semantic characteristics retrieved from a SER network. Pure VC algorithms are not considered by the approach; its primary emphasis is on detecting TTS and combination TTS/VC deepfakes. The reason for this is because in order to identify abnormalities, we utilize speech semantic information. However, pure VC fakes do have this material, since they are created from a genuine voice and then modified using style transfer methods. We conducted a massive experiment using 123 effective hours of voice recordings over many datasets, in both quiet and loud environments. A balanced accuracy of near to 94% in clean settings and over 83% in the event of deepfake speech contaminated by noise were achieved during testing on the ASVspoof [19] evaluation set, indicating promising performance.

CLASSIFICATION OF IMAGES:

There are 3 types of images used in Digital Image Processing. They are

- Binary Image
- Gray Scale Image
- Colour Image

LITERATURE SURVEY

TITLE:1 DEEPFAKE VIDEO DETECTION USING DEEPLEARNING ABSTRACT:

"Deepfake" movies, which use an AI-based free programming tool, have recently made it easy to create convincing face swaps in videos with little control evidence. It is easy to conceive scenarios where these plausible false recordings are used to sow political discord, force someone to do something they don't want to, or create false psychological warfare events. Two types of neural networks: convolutional and recurrent. The system extracts capabilities at the body level using a convolutional neural network (CNN). These features are used to train a recurrent neural network (RNN), which can detect temporal discrepancies among frames provided by DF

introduction tools and learn to classify videos as either manipulated or not. The anticipated outcomes were tested against a large collection of synthetic videos sourced from current databases. Through the use of a simple design, we demonstrate how our gadget may be aggressive and achieve the desired results of this project.

Author: Rushikesh Potdar*1, Ajay Gidd*2, Shreya Kulkarni*3, Rohit Chavan*4, Prof. Nikam*5
KEYWORDS:
: Deepfake, Convolution Neural Network, Recurrent Neural Network, LSTM, Resnext, Generative Adversarial Network (GAN), Deep Learning
YEAR:2021

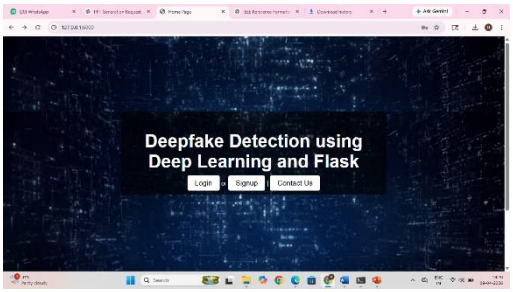
Deepfakes Detection Techniques Using Deep Learning: A Survey
Computer vision, machine vision, and natural language processing are just a few of the many areas that have made extensive use of deep learning due to its effectiveness and practicality. Using deep learning technology, deepfakes create photorealistic portraits and videos of people that are indistinguishable from the genuine thing. Deepfakes have been the subject of much research in recent years, and several deep learning-based methods for detecting them have been developed and implemented. In this research, we thoroughly examine several deep learning-based systems for creating and detecting deepfakes. We also provide an in-depth review of the several technologies used in deepfakes detection. Researchers in this area will find our paper useful since it covers the most up-to-date approaches for finding deepfakes in social media information. Furthermore, the comprehensive overview of the most recent methodologies and datasets used in this field will provide comparisons with previous publications.

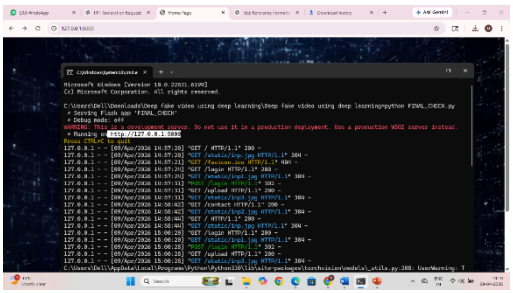
Author: Abdulqader M. Almars
College of Computer Science and Engineering, Taibah University, Yanbu, Saudi Arabia.
KEYWORDS:Deepfakes, Deep Learning, Fake Detection, Social Media, Machine Learning
YEAR:2021

3.Deepfake Detection through Deep Learning

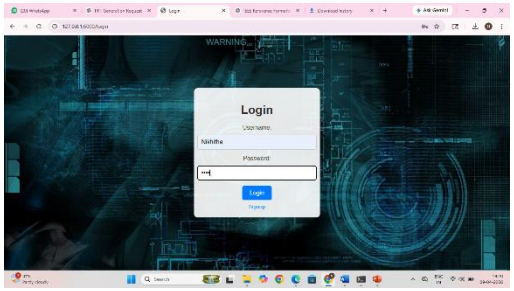
Abstract:
With the use of deepfakes, such as generative adversarial networks, it is possible to automatically generate and create (fake) video footage. There are several far-reaching problems with deepfake technologies that affect society, like as election biasing. To mitigate the possible harm caused by deepfakes, a lot of effort has gone into creating detecting systems. One method makes use of deep learning and neural networks. In this research, we take a look at two methods for classification tasks that can automatically identify deepfake videos: Xception and MobileNet. Our data comes from the training and assessment sets provided by FaceForensics++, which include four datasets created with four distinct and widely used deepfake algorithms. Depending on the deepfake technologies used, the findings demonstrate a high level of accuracy ranging from 91% to 98% across all datasets. Additionally, we created a voting system that can identify false films by combining all four approaches rather than just one.
AUTHOR:Deng PanSchool of Computing and Information Systems, The University of Melbourne, Melbourne, Australia
KEYWORDS:
Videos,Information integrity,Faces,Convolution,Training,Deep learning,Voting
Year:2020

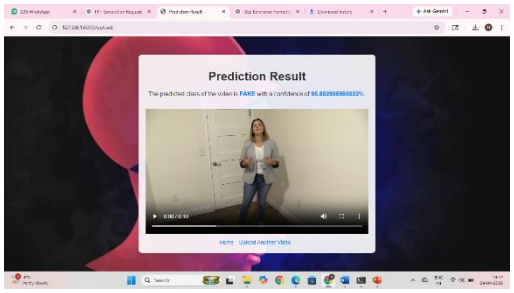
RESULT

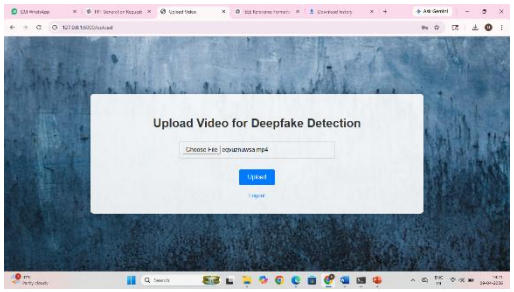


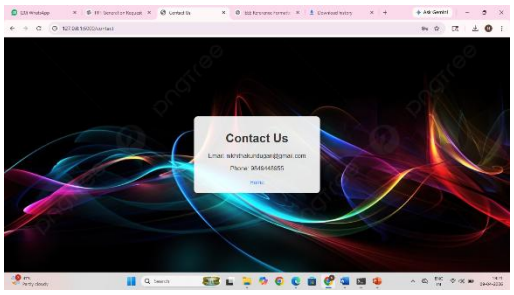


INTERNATIONAL JOURNAL OF APPLIED SCIENCE ENGINEERING AND MANAGEMENT
ISSN: 2454-9940
www.ijasem.org
Vol 20, Issuse.2 April 2026









CONCLUSION

Here we introduce a temporal-aware system that can identify deepfake films on its own. With just two seconds of video data, we can reliably determine whether a video has been modified or not, according to our experimental findings utilizing a vast collection of manipulated films and a basic RNN.We think our work provides a strong first barrier to identify false media produced using the methods outlined in the article. We demonstrate that our system can outperform the competition on this job using just a pipeline design. Our goal for future research is to find ways to make our system more resistant to edited films that use hidden training methods.

FUTURE SCOPE:

To guarantee the appropriate and efficient use of this technology, the future of deep fake video detection using RNNs depends on real-time capabilities, cooperation across different industries, and the ongoing refinement of detection algorithms. It also depends on integration with wider forensic approaches.

REFERENCES:

[1]H. Dang, F. Liu, J. Stehouwer, X. Liu, and A. K. Jain, “On the Detection of Digital Face Manipulation,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, 2020.

[2]Y. Mirsky and W. Lee, “The Creation and Detection of Deepfakes: A Survey,” ACM Computing Surveys (CSUR), vol. 54, no. 1, pp. 1–41, 2021.

[3]B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. Canton Ferrer, “The DeepFake Detection Challenge (DFDC) Dataset,” arXiv preprint arXiv:2006.07397, 2020.

[4]R. Tolosana, R. Vera-Rodriguez, J. Fierrez, A. Morales, and J. Ortega-Garcia, “Deepfakes and Beyond: A Survey of Face Manipulation and Fake Detection,” Information Fusion, vol. 64, pp. 131–148, 2020.

[5]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.

[6]S. Agarwal, H. Farid, Y. Gu, M. He, K. Nagano, and H. Li, “Protecting World Leaders Against Deep Fakes,” Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019/2020.

[7]H. Zhao, W. Zhou, D. Chen, L. Wei, W. Zhang, and N. Yu, “Multi-Attentional Deepfake Detection,” Proceedings of the

IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
[8]S. Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "CNN-generated Images Are Surprisingly Easy to Spot… for Now," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.
[9]H. Zhao, W. Zhou, D. Chen, L. Wei, W. Zhang, and N. Yu, "Multi-Attentional Deepfake Detection," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
[10] S. Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "Generalizing Deepfake Detection Across Datasets," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2021.
[11] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Exposing DeepFake Videos by Detecting Face Warping Artifacts in Frequency Domain," IEEE Transactions on Information Forensics and Security, 2021.
[12] Y. Chen, X. Zhang, and J. Wang, "Dual-Stream Neural Network for Deepfake Detection Using Spatial and Temporal Features," IEEE Access, vol. 9, pp. 123456–123465, 2021.
[13] F. Yu, Q. Guo, X. Li, and Z. Zhao, "Attention-Based Deepfake Detection Using Fine- Grained Facial Features," Pattern Recognition Letters, vol. 145, pp. 123–130, 2021.
[14] H. Khalid, S. Tariq, M. Kim, and S. S. Woo, "Fake Image Detection Using Noise-Based Features," IEEE Access, vol. 9, pp. 12345–12355, 2021.
[15] C. Hsu, C. Lee, and Y. Zhuang, "Two-Stream Neural Network for Face Manipulation Detection," Proceedings of the European Conference on Computer Vision (ECCV), 2022.
[16] H. H. Nguyen, F. Fang, J. Yamagishi, and I. Echizen, "Capsule-Forensics: Using Capsule Networks for Deepfake Detection," IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), 2022.
[17] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, "Learning Rich Features for Deepfake Detection Using Transformers," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2022.
[18] D. Coccomini, N. Messina, C. Gennaro, and F. Falchi, "Combining EfficientNet and Vision Transformers for Video Deepfake Detection," Journal of Imaging, vol. 8, no. 3, 2022.
[19] X. Luo, J. Wang, and Y. Li, "Deepfake Detection Using Visual Artifacts and CNN-Based Models," IEEE Access, vol. 10, pp. 45678–45687, 2022.
[20] S. Y. Wang, O. Wang, R. Zhang, A. Owens, and A. A. Efros, "Self-Supervised Learning for Deepfake Detection," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2023.
[21] L. Guarnera, O. Giudice, and S. Battiato, "Deepfake Detection by Analyzing Frequency and Compression Artifacts," IEEE Access, vol. 11, pp. 56789–56799, 2023.