IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

EXPOSING DEEP FAKES IN SOCIAL MEDIA PLATFORMS USING GENERATIVE ADVERSARIAL NETWORKS

Dr. Palson Kennedy.R[1], Nandhini.S[2], Priyadharshini.D[3],
[1]Professor, [2,3] Students,
Department of Computer Science and Engineering,
PERI Institute of Technology, Chennai, India.

Abstract: This project delves into the critical realm of combating deep fakes in online networking platforms by employing advanced deep learning techniques. Leveraging Generative Adversarial Networks (GANs) to simulate deep fake generation and utilizing the feature extraction capabilities of Inception ResNetV2, there search aims to develop a robust deep fake detection model. The proposed model undergoes comprehensive training, evaluation, and fine-tuning, with a focus on countering adversarial techniques employed in sophisticated deep fake generation. The study contributes are liable and effective means of discerning between genuine and synthetic content, offering heightened security for online networking platforms. Furthermore, insights gained into adversarial strategies and practical deployment recommendations provide a comprehensive approach to addressing the rising threat of deep fakes in the digital landscape. The proposed model undergoes comprehensive training, evaluation and fine-tuning, with focus on countering adversarial techniques employed in sophisticated deep fake generation.

I. INTRODUCTION

Theexpandingcomplexityofcellphonecamerasandtheaccessibilityofgood web association all around the world has expanded the always developing reach of online media also, media sharing entries have made the creation and transmission of computerized recordings more simple than any time in recent memory. The developing computational force has made deep learning so incredible that would have been thought unthinkable just a modest bunch of years prior. Like any extraordinary innovation, this has made new difficulties. Purported "DeepFake" created by deep generative adversarial models that can control video and brief snippets. Spreading of the DF over the online media stages have got ten extremely normal prompting spamming and peculating in correct data over the stage. These sorts of the DF will be awful, and lead to threatening, deluding of average citizens. To conquer this sort of situation, DF detection is very essential. So, we describe a brand new deep learning based approach that could successfully distinguish AI-generated fake videos (DFVideos) from actual videos. It’s incredibly essential to broaden technology that could spot fakes, so that the DF may be recognized and averted from spreading over the internet.

II. EXISTINGSYSTEM

The detection of digital face manipulation in video has attracted extensive attention due to the increased risk to public trust. To counteract the malicious usage of such techniques, deep learning-based deep fake detection methods have been developed and have shown impressive results. However, the performance of these detectors is often evaluated using benchmarks that hardly reflect real-world situations. For example, the impact of various video processing operations on detection accuracy has not been systematically assessed. To address this gap, this project first analyses numerous real-world influencingfactorsandtypicalvideoprocessingoperations.Then,amoresystematicassessment methodology is proposed, which allows for a quantitative evaluation of a detector's robustness under the influence of different processing operations. For this entire process they used the SVM model.

III. PROPOSED SYSTEM

Increasing computing power has made deep learning algorithms so powerful that creating a fake video generated by artificial intelligence, popularly called as deep fakes, is very simple. Scenarios where this realistic face has been replaced by deep fakes are used to create political unrest, fake terrorist acts, revenge porn, blackmailing nations can easily be imagined. In this work, a new method is based on deep learning that can effectively distinguish fake videos generated by artificial intelligence from real videos. This method is able toautomaticallydetectreplacementandreenactmentofdeepforgery.Using artificial intelligence (AI) to fight against artificial intelligence(AI). This system uses ResNext Convolution Neural Network to extract frame-level features and these features and further uses GAN and ResNet V2 to classify whether the video is subject to some kind of manipulation i.e..Whether the video is deep fake or real video. System can also achieve competitive results using a very simple and robust approach.

IV. SYSTEMARCHITECTURE

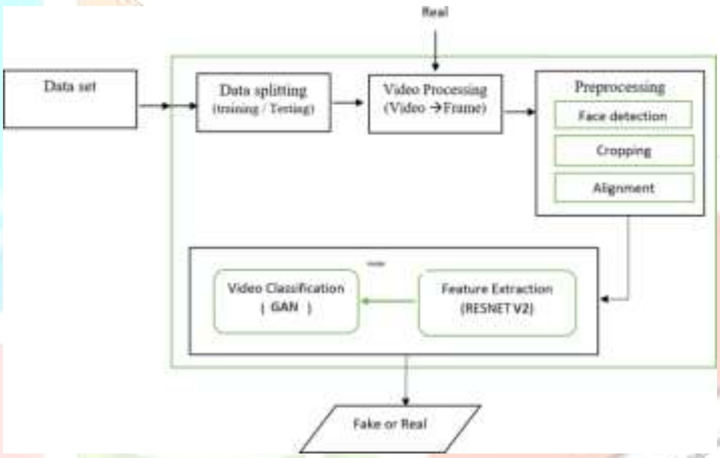
Real
Data set
Data splitting (training / Testing)
Video Processing (Video →Frame)
Preprocessing
Face detection
Cropping
Alignment
Video Classification (GAN)
Feature Extraction (RESNET V2)
Fake or Real

Fig No.1 System Architecture

4.1ARCHITECTUREEXPLANATION

A convolutional neural network (CNN or ConvNet is a network architecture for deep learning that learns directly from the data below in Fig. CNNs are useful for finding image patterns to recognize objects, classes and categories. Convolutional Neural Networks comprise node layers, convolution, pooling, hidden, and output layers. Our image passes through all these layers, and every step is important. As for the convolution layer, this layer is the main block of CNN. This is the layer where most of the computations occur. This layer has three inputs. The first is input data, these connection is a filter, and the third is a feature map. The convolution process consists of the input image used with filters, and filters are mainly 3x3 matrices that are iterated over the image. In the end, an activation function calculates the final result. The output is stored in an output matrix. The pooling layer is used for down sampling, reducing the number of parameters in our input image. The fully connected layer is the layer that performs classification on the image and decides whether an image is real or fake.

V.CONCLUSION

We provided a neural network-primarily based totally method to classify the video as deep fake or actual, at the side of the self-assurance of the proposed model. The proposed approach is stimulated with the aid of using the manner the deep fakes are created with the aid of using the GANs with the assist of Auto encoders. Our approach does the frame stage detection the use of ResNetV2 and video class the use of RNN at the side of GAN. The proposed approach is successful in detecting the video as a deep fake or actual primarily based totally on the listed parameters in the paper. We consider that it'll offer a very excessive accuracy on actual time data.

VI.FUTUREENHANCEMENT

The detection showed its best performance with the CNN, and more accurate results are aimed to be established in the future for a more authentic, precise website. It is also intended to work with different datasets and generalize the dataset more with several augmentation techniques so that the system can detect any data inserted by the user.

REFERENCES

[1] T. Mahara, V. L. H. Josephine, R. Srinivasan, P. Prakash, A. D. Algarni and O. P. Verma, "Deep vs. Shallow: A Comparative Study of Machine Learning and Deep Learning Approaches for Fake Health News Detection," in IEEE Access, vol. 11, pp. 79330-79340, 2023, doi: 10.1109/ACCESS.2023.3298441.

[2] Z. M. Almutairi and H. Elgibreen, "Detecting Fake Audio of Arabic Speakers Using Self-Supervised Deep Learning," in IEEE Access, vol. 11, pp. 72134-72147, 2023, doi: 10.1109/ACCESS.2023.3286864.

[3] A. A. Obaid, H. Khotanlou, M. Mansoorizadeh and D. Zabihzadeh, "Robust Semi-Supervised Fake News Recognition by Effective Augmentations and Ensemble of Diverse Deep Learners," in IEEE Access, vol. 11, pp. 54526-54543, 2023, doi: 10.1109/ACCESS.2023.3278323.

[4] M. Tajrian, A. Rahman, M. A. Kabir and M. R. Islam, "A Review of Methodologies for Fake News Analysis," in IEEE Access, vol. 11, pp. 73879-73893, 2023, doi: 10.1109/ACCESS.2023.3294989.

[5] K. Li, X. Lu, M. Akagi and M. Unoki, "Contributions of Jitter and Shimmer in the Voice for Fake Audio Detection," in IEEE Access, vol. 11, pp. 84689-84698, 2023, doi: 10.1109/ACCESS.2023.3301616.

[6] B. L. V. S. Aditya and S. N. Mohanty, "Heterogenous Social Media Analysis for Efficient Deep Learning Fake-Profile Identification," in IEEE Access, vol. 11, pp. 99339-99351, 2023, doi: 10.1109/ACCESS.2023.3313169.

[7] N. M. Alnaim, Z. M. Almutairi, M. S. Alsuwat, H. H. Alalawi, A. Alshobaili and F. S. Alenezi, "DFFMD: A Deepfake Face Mask Dataset for Infectious Disease Era WithDeepfake Detection Algorithms," in IEEE Access, vol. 11, pp. 16711-16722, 2023, doi: 10.1109/ACCESS.2023.3246661.

[8] R. Catelli et al., "A New Italian Cultural Heritage Data Set: Detecting Fake Reviews With BERT and ELECTRA Leveraging the Sentiment," in IEEE Access, vol. 11, pp. 52214-52225, 2023, doi: 10.1109/ACCESS.2023.3277490.

[9] M. O. Alassafi et al., "A Novel Deep Learning Architecture With Image Diffusion for Robust Face Presentation Attack Detection," in IEEE Access, vol. 11, pp. 59204-59216, 2023, doi: 10.1109/ACCESS.2023.3285826.

[10] M. Rajaee and K. Mazlumi, "Multi-Agent Distributed Deep Learning Algorithm to Detect Cyber-Attacks in Distance Relays," in IEEEAccess, vol. 11, pp. 10842-10849, 2023, doi: 10.1109/ACCESS.2023.3239684.

[11] D. Benalcazar, J. E. Tapia, S. Gonzalez and C. Busch, "Synthetic ID Card Image Generation for Improving Presentation Attack Detection," in IEEE Transactions on Information Forensics and Security, vol. 18, pp. 1814-1824, 2023, doi: 10.1109/TIFS.2023.3255585.

[12] A. H. Khalil, A. Z. Ghalwash, H. A. -G. Elsayed, G. I. Salama andH. A. Ghalwash, "Enhancing Digital Image
ForgeryDetection Using Transfer Learning," in IEEE Access, vol. 11, pp. 91583-91594, 2023, doi: 10.1109/ACCESS.2023.3307357.

[13] I. -Y. Kwak et al., "Voice Spoofing Detection Through Residual Network, Max Feature Map, and Depthwise Separable Convolution," in IEEE Access, vol. 11, pp. 49140-49152, 2023, doi: 10.1109/ACCESS.2023.3275790.

[14] H. Alamro, K. Mahmood, S. S. Aljameel, A. Yafoz, R. Alsini and
a. Mohamed, "Modified Red Fox Optimizer With Deep Learning Enabled False Data Injection Attack Detection," in IEEE Access, vol. 11, pp. 79256-79264, 2023, doi: 10.1109/ACCESS.2023.3298056.

[15] R. M. Albalawi, A. T. Jamal, A. O. Khadidos and A. M. Alhothali, "Multimodal Arabic Rumors Detection," in IEEE Access, vol. 11, pp. 9716-9730, 2023, doi: 10.1109/ACCESS.2023.3240373.

[16] R. A. Zayed, L. F. Ibrahim, H. A. Hefny, H. A. Salman and A. AlMohimeed, "Experimental and Theoretical Study for the Popular Shilling Attacks Detection Methods in Collaborative Recommender System," in IEEE Access, vol. 11, pp. 79358-79369, 2023, doi: 10.1109/ACCESS.2023.3289404.

[17] A. A. Shafee, M. M. E. A. Mahmoud, G. Srivastava, M. M. Fouda,M.Alsabaan and M. I. Ibrahem, "Detection of Distributed Denial of Charge (DDoC) Attacks Using Deep Neural Networks With VectorEmbedding," in IEEE Access, vol. 11, pp. 75381-75397, 2023, doi: 10.1109/ACCESS.2023.3296562.

[18] Y. Patel et al., "An Improved Dense CNN Architecture for Deepfake Image Detection," in IEEE Access, vol. 11, pp. 22081-22095, 2023, doi: 10.1109/ACCESS.2023.3251417.

[19] Z. Liu, J. Hu, Y. Liu, K. Roy, X. Yuan and J. Xu, "Anomaly-Based Intrusion on IoT Networks Using AIGAN-a Generative Adversarial Network," in IEEE Access, vol. 11, pp. 91116-91132, 2023, doi: 10.1109/ACCESS.2023.3307463.

[20] J. M. Pérez et al., "Assessing the Impact of Contextual Information in Hate Speech Detection," in IEEE Access, vol. 11, pp. 30575-30590, 2023, doi: 10.1109/ACCESS.2023.3258973.

[21] M. J. Abdulaal et al., "Privacy-Preserving Detection of Power Theft in Smart Grid Change and Transmit (CAT) Advanced Metering Infrastructure," in IEEE Access, vol. 11, pp. 68569-68587, 2023, doi: 10.1109/ACCESS.2023.3291217.

[22] S. Asiri, Y. Xiao, S. Alzahrani, S. Li and T. Li, "A Survey of Intelligent Detection Designs of HTML URL Phishing Attacks," in IEEE Access, vol. 11, pp. 6421-6443, 2023, doi: 10.1109/ACCESS.2023.3237798.

[23] W. M. B. Ateeq and H. S. Al-Khalifa, "Intelligent Framework for Detecting Predatory Publishing Venues," in IEEE Access, vol. 11, pp. 20582-20618, 2023, doi: 10.1109/ACCESS.2023.3250256.

[24] S. Y. Diaba, M. Shafie-Khah and M. Elmusrati, "Cyber Security in Power Systems Using Meta-Heuristic and Deep Learning Algorithms," in IEEE Access, vol. 11, pp. 18660-18672, 2023, doi: 10.1109/ACCESS.2023.3247193.

[25] R. Sepúlveda-Torres, A. Bonet-Jover and E. Saquete, "Detecting Misleading Headlines Through the Automatic Recognition ofContradiction in Spanish," in IEEE Access, vol. 11, pp. 72007-72026, 2023, doi: 10.1109/ACCESS.2023.3295781.