WWW.IJRPR.COM

International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Review Paper on Deep Fake Detection Using Deep Learning

Anjali Mahantesh Mudakavi[1], Lalitha Virupakshi Mudakavi[2], Prateek Kataraki[3], Shreya Mahesh Desai[4], Prof. Vilas Jarali[5]

[1,2,3,4,5]Department of Computer Science and Engineering, Angadi Institute of Technology and Management, Belagavi-590009, India

A B S T R A C T

. In recent months, a slew of machine learning-based software has appeared to facilitate the seamless creation of convincingly altered facial features in videos, a phenomenon often referred to as "deepfake. Video. These manipulations are so sophisticated that they often leave little trace of tampering, giving rise to concerns about possible misuse in various scenarios, such as inciting political unrest, orchestrating blackmail schemes or orchestrating fake terrorist activity. The purpose of this article is to provide a comprehensive overview of recent research projects focused on the comprehensive detection of fraudulent content using advanced deep learning methods.

We aim to build research excellence through methodical research of different categories in the field of fake content detection. Many studies have already delved into the development of detection methods aimed at mitigating the potential harmful effects associated with widespread deep forgery. The application of neural networks and deep learning stands out as a central approach to reformulate these sentences and deal with the multifaceted challenges presented by the prevalence of deep falsification.

Keywords: — Deepfakes, Deep Learning,Adversarial media detection,Machine learning for video forgery analysis.

1. INTRODUCTION

This article discusses various deep learning solutions designed to automatically classify and detect deep fake videos. In particular, FaceForensics++ serves as the main source of video data used to train two neural networks, Xception and MobileNet, using pre-processed images. Each online course produces four different models corresponding to the four main deepfaking software platforms: Deepfakes, Face2Face, FaceSwap, and NeuralTextures.

Evaluation of Modeland #039 ratings shows remarkable accuracy in distinguishing between authentic and manipulated video. However, this accuracy is particularly sensitive and highly dependent on the deep forging platform used. To solve this problem, we propose a voting mechanism that utilizes the outputs of different models and provides a more stable and efficient solution(Deep fake Detection through deep learning).

As presented in the Malicious Artificial Intelligence Report [11], researchers show in artificial intelligence one should always consider the duplicity of this work and allow abuse affects research priorities and standards. Given the severity for malicious attack vectors caused by deep fakes, In this paper, we introduce a new solution for detection such a video. The main contributions of this work are summarized as follows.First, we propose a two-step analysis of CNN to extract features at the frame level followed by a time-aware RNN to capture temporal inconsistencies between frames caused by face change process Second, we used a collection of 600 videos evaluate the proposed method on half of the videos deep fakes collected from various video hosts. Third, we experimentally demonstrate the effectiveness of the described approach for suspect identification video is deep fake manipulation with 94% accuracy as the baseline of a random detector in a balanced setup(deep fake detection using recurrent neural network)

As the 2020 US election approaches, fake videos have become the focus of media attention. In an era dominated by fake news, there is a growing fear about the credibility of online content. To address these issues, Facebook and Instagram introduced policies in January 2020 that prohibit manipulation of artificial intelligence and quot;deepfake and quot; videos that can mislead viewers during elections [1]. However, the effectiveness of this approach depends on the ability to accurately distinguish between genuine and manipulated videos, which is the focus of this article (and quot;Deepfake Detect through Deep Learning quot;). Advances in technology have led to the misuse of fake technologies, especially the creation of explicit content featuring celebrities and politicians. This trend fueled the spread of propaganda and fake news and contributed to many social problems [4]. (Reference: "DeepVision: Deep Fake Detection Using the Blink Pattern of the Human Eye")

Manipulation of faces in photos or videos is a serious threat to global security. Faces play a central role in human-to-human communication and biometric authentication and identification services. Thus, changes in face frames can undermine trust in security applications and digital communication [1]. Therefore, analyzing and detecting manipulated faces in photos or videos plays a crucial role in detecting fraudulent content (and quot;a new deep learning-based method for video deep fake detection using XGBoosandquot;). The widespread use of social networking sites such as Facebook, Twitter

and YouTube, along with the proliferation of sophisticated camera phones, has made it easier than ever to create, share and edit videos and images. The rise of hyper-realistic fake images and videos produced using deep-fake techniques and distributed on these platforms has raised public privacy concerns. Deepfake, a technique based on deep learning, can seamlessly replace the face of the source person with the target person in the video, resulting in a fake story. Misuse of deepfake technology creates threats such as creation of fake news to disrupt financial markets, defamation of celebrities and confusion. Therefore, addressing the challenges of deep counterfeiting technology is necessary to reduce its potential disadvantages.

2. METHODOLOGY

The main goal of this study was to authenticate the videos by determining whether they are real or created by deep fake technology. Since deep learning models usually work with images, converting the video stream into a suitable format for these models required the development of a preprocessing module. In addition to adjusting the types of input data, this module had to consider the possible influence of non- video factors on the model training process.

In a video frame, it is worth noting that the face is not the only content; the frame mainly consists of the person and body parts and the background. These extra elements can negatively affect model training. Thus, the focus of the preprocessing module was on isolating the face region of the image for model input. The module shown in Figure includes three main steps: capture frames of video, detect faces in those frames, and save face regions as images.

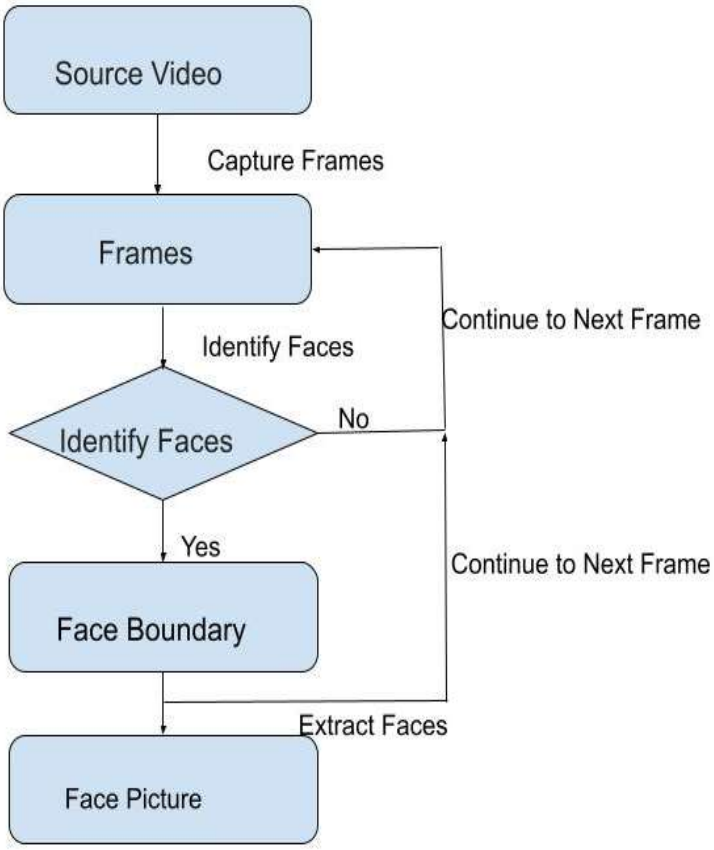

Fig. Preprpcessing flowchart

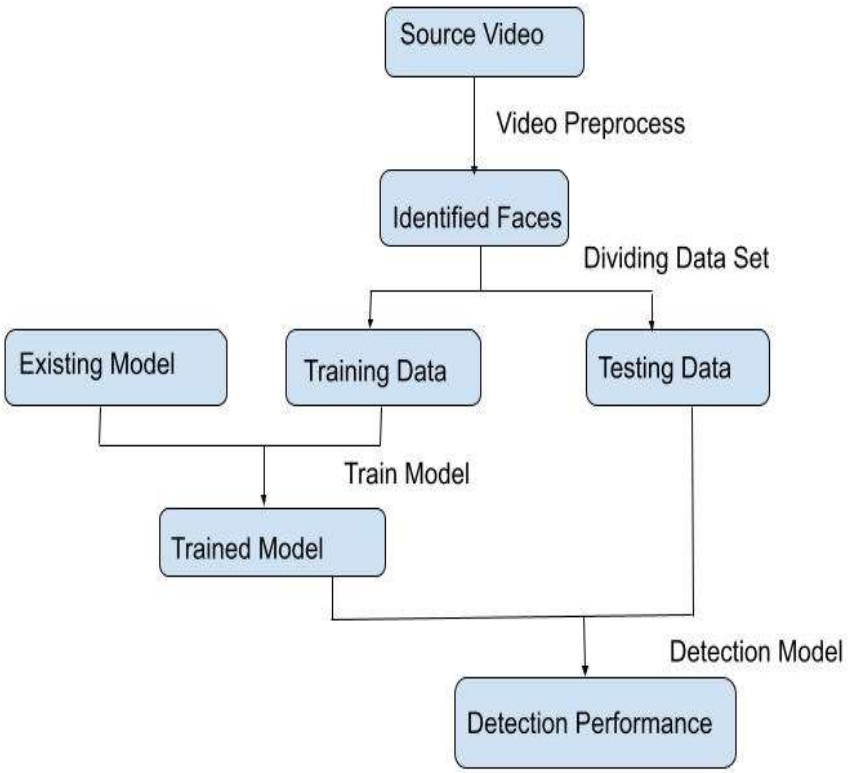

Fig. Overall process

The first step involved capturing the input video into frames, which was done using the video capture function of the OpenCV Python package. Since the project and # 039; relying on a single image as input, information between frames was considered unnecessary, as the high similarity of adjacent frames can lead to reduced training efficiency and potential over-fitting problems. Selecting one image out of every four frames was found to be a pragmatic approach given that the videos run at 30 frames per second. Moving to the second step, face recognition and image labeling were performed using the cascade classifier provided by OpenCV. After evaluating different classifiers, the haarcascade_frontalface_alt classifier was chosen due to its accuracy in delineating facial regions. To solve the problems of non-face selection, a criterion was defined to keep the largest detected area, which ensures accurate face recognition.

In the final step, the detected face regions were saved as new images, which were uniformly resized. The Xception model required an image size of 299x299, while the MobileNet model required smaller sizes of 224x224. The goal of this systematic video processing process was to optimize the input for subsequent deep learning model analysis, emphasizing the isolation of facial features that improves the accuracy of distinguishing between real and deep fake videos

3. COMPARISON TABLE

This a table that summarizes the results of 16 different face recognition techniques on a variety of video datasets. Each row in the table shows the year the technique was published, its name, the datasets it was tested on, the type of input it takes (videos or images), and the best result it achieved on a given metric (accuracy, AUC, or variance).

Ref	Year	Technique	Dataset	Input	Best result
[1]	2020	Xception	FaceForensics++	Videos	Accuracy 90%
[3]	2022	UV Texture Map	FaceForensics++ and DFDC	Video	Accuracy 99.79%
[4]	2022	Efficient Net-B0	FaceForensics++ and DFDC	Videos	AUC:0.951
[5]	2022	Efficient Net-B0	FaceForensics++ and DFDC	Videos	AUC:0.951
[6]	2022	CNN on multiple face organs	FaceForensics++ and DFDC,DFDC-P and Celeb-DF	Video	AUC:99.93%
[7]	2022	VGG	FaceForensics++ DFDC and DFD	Video	AUC:98.40%

[1]	2020	MobileNet	FaceForensics++	Videos	Accuracy 90%
[3]	2022	Face cut-out and Random cut-out	FaceForensics++ and DFDC	Video	Accuracy 98.24%
[8]	2022	Patch extraction and embedding	FaceForensics++, Celeb-DF and WildDeepfake	Video	Accuracy 99.41%
[9]	2022	Vision Transformer	ForgeryNet[126]	Image	Variance:0.004
[10]	2022	CNN+Frequency Filter	FaceForensics++, Celeb-DF and SR-DF	Video	AUC:91.20%
[2]	2022	Dropout rate to discard image patches	FaceForensics++, Celeb-DF and FaceForensics++	Video	AUC:99.8%
[11]	2023	2D and 3D CNNs	Celeb-DF, DFDC and FaceForensics++,	Video	AUC:0.9624
[12]	2023	Efficient Net-B7	DFDC and Celeb-DF(v2)	Video	AUC:0.982
[13]	2023	Efficient Net-B4	Celeb-DF, DFDC, FaceForensics++, and WildDeepfake	Video	***Accuracy:99.80%***

3.CONCLUSION

Different approaches exist, each with strengths and weaknesses. Supervised learning excels in specific datasets but can be vulnerable to adversarial adaptation. Unsupervised learning excels in anomaly detection but lacks interpretability and robustness. Hybrid approaches that combine both methods show promise for improved accuracy and generalizability.

Data scarcity of diverse deepfakes makes it difficult to train robust models. Continuously evolving deepfake creation techniques necessitate the development of adaptive detection models. Interpretability and explainability of model decisions are crucial for building trust and accountability.

Overall, the field of deepfake detection is brimming with potential, but overcoming current challenges and actively pursuing promising research directions are essential to build reliable and robust solutions that can safeguard online integrity and trust.

REFERENCES

[1] Deng Pan, Lixian Sun, Rui Wang,Xingjian Zhang, Richard O Sinnott,"Deepfake Detection Through Deep Learning" 2020 IEEE/ACM International Conference on Big Data Computing, Applications and Technologies (BDCAT) .

[2] D. Zhang, F. Lin, Y. Hua, P. Wang, D. Zeng, S. Ge, Deepfake video detection with spatiotemporal dropout transformer, in: Proceedings of the 30th ACM International Conference on Multimedia, 2022, pp. 5833–5841..

[3] S. A. Khan, D.-T. Dang-Nguyen, Hybrid transformer network for deepfake detection, in: Proceedings of the 19th international conference on content-based multimedia indexing, 2022, pp. 8–14.

[4] D. A. Coccomini, N. Messina, C. Gennaro, F. Falchi, Combining effi-cientnet and vision transformers for video deepfake detection, in: Image Analysis and Processing (ICIAP 2022),Springer International Publishing,Cham, 2022, pp. 219–229.

[5] J. Feinland, J. Barkovitch, D. Lee, A. Kaforey, U. A. Ciftci, Poker bluffdetection dataset based on facial analysis, in: International Conference on Image Analysis and Processing, Springer, 2022, pp. 400–410.

[6] Z. Xue, Q. Liu, H. Shi, R. Zou, X. Jiang, A transformer-based deepfake-detection method for facial organs, Electronics 11 (24).

[7] Y. Zhang, T. Wang, M. Shu, Y. Wang, A robust lightweight deepfake detection network using transformers, in: Pacific Rim International Conference on Artificial Intelligence, Springer, 2022, pp. 275–288.

[8] A. Khormali, J.-S. Yuan, DFDT: an end-to-end deepfake detection frame-work using vision transformer, Applied Sciences 12 (6) (2022) 2953.

[9] D. A. Coccomini, R. Caldelli, F. Falchi, C. Gennaro, G. Amato, Cross-Forgery Analysis of Vision Transformers and CNNs for Deepfake ImageDetection, in: Proceedings of the 1st International Workshop on Multi-media AI against Disinformation, 2022, pp. 52–58.

[10] J. Wang, Z. Wu, W. Ouyang, X. Han, J. Chen, Y.-G. Jiang, S.-N. Li,M2tr: Multi-modal multi-scale transformers for deepfake detection, in:Proceedings of the 2022 international conference on multimedia retrieval,2022, pp. 615–623.

[11] *M. A. Raza, K. M. Malik, I. U. Haq, Holisticdfd: Infusing spatiotemporal transformer embeddings for deepfake detection, Information Sciences (2023) 119352.*

[12] Y. Heo, W. Yeo, B. Kim, Deepfake detection algorithm based on improved vision transformer, Applied Intelligence 53 (2023) 7512–7527.

[13] H. Lin, W. Huang, W. Luo, W. Lu, Deepfake detection with multi-scale convolution and vision transformer, Digital Signal Processing 134 (2023)103895.

[14] F. Khalid, M. H. Akbar, S. Gul, Swynt: Swin y-net transformers for deepfake detection, in: 2023 International Conference on Robotics and Automation in Industry (ICRAI), 2023, pp. 1–6. doi:10.1109/ICRAI57502.2023.10089585..

[15] David Guera, Edward J. Delp,"Deepfake Detection Using Recurrent Neural Network" based on research sponsored by the defense advanced Projects agency.

[16] Aya Ismail, Marwa Elpeltagy, Mervat S. Zaki, Kamal Eldahshan, "A New deep Learning-based methodology for video deepfake detection using XGBoost.sensors2021,21,5413.https://doi.org/10.3390/s21165413.

[17] TackHyun Jung, SangWon Kim, KeeCheon Kim, "Deepvision: DEepfakes detection using human eye blinking pattern DOI.10.1109/ACCESS.2020.2988660,IEEE Access.