Guarding Authenticity: A Deepfake Detection System

Prof. N. R. Thorat[1], Prajwal Mohite[2], Anubhaw Mishra[3], Rohit Sawant[4]

[1]Assistant Professor, Department of Computer Engineering, SITS, Pune
[2]BE Graduate (IV year), Department of Computer Engineering, SITS, Pune
[3]BE Graduate (IV year), Department of Computer Engineering, SITS, Pune
[4]BE Graduate (IV year), Department of Computer Engineering, SITS, Pune

---***---

ABSTRACT -

This Recent advancements in computer vision have led to the development of powerful tools that can create realistic deepfakes. A generative adversarial network (GAN) can manipulate captured media streams, such as images, audio, and video, to make them appear to fit other environments. The spread of these fake media streams can cause chaos in social communities and damage the reputation of individuals or groups. It can also influence public sentiments and opinions toward the targeted person or community. Researchers have suggested using convolutional neural networks (CNNs) as an effective method for detecting deepfakes in the network. However, most existing techniques struggle to capture the dissimilarities between frames in the collected media streams. Motivated by this challenge, this paper presents a novel and improved deep-CNN (D-CNN) architecture for deepfake detection. The proposed approach aims to achieve reasonable accuracy and high generalizability. The model is trained on images from multiple sources, which enhances its overall generalizability capabilities.A binary-cross entropy and Adam optimizer are utilized to improve the learning rate of the D-CNN mode.

Key Words: cnn, deepfake, image, network, media stream, detection

1.INTRODUCTION

In this As digital technology rapidly evolves, a disturbing new threat has emerged - the ability to create synthetic or "deepfake" media using advanced artificial intelligence. Deepfakes leverage cutting-edge generative adversarial networks (GANs) to fabricate highly realistic fake images, videos and audio that can be disturbingly difficult to distinguish from authentic content. While holding creative potential, deepfakes also pose grave risks of being weaponized for misinformation campaigns, defamation, and undermining societal trust. Traditional forensic techniques are no match for the sophistication of modern deepfake generators, driving an urgent need for robust detection solutions powered by deep learning and computer vision. However, a critical challenge lies in developing models that can generalize effectively across diverse deepfake manipulations created by various GAN architectures and training data sources. Many existing approaches exhibit high accuracy on specific deepfake sources used during training, but falter dramatically when faced with unseen generation techniques - a generalization gap that leaves them vulnerable. To tackle this pressing issue, our work proposes a cutting-edge deep convolutional neural network (D-CNN) architecture meticulously designed to learn generalizable visual features for reliably distinguishing deepfakes across multiple sources. By training on a diverse dataset spanning numerous deepfake varieties and authentic samples, our model aims to maintain high detection accuracy in an ever-evolving landscape.

2. LITERATURE SURVEY

The paper [1] "DeepFake detection for human face images and videos: A survey" discusses the survey over deepfake detection. Research on DeepFake detection using deep neural networks (DNN) to identify and classify DeepFakes has attracted attention. Basically, DeepFake is an ad achieved by injecting or changing some information into the DNN model. In this review, we will introduce DeepFake's face and video detection methods according to results, performance, methods and detection types.We will review the existing types of DeepFake creation techniques and sort them into Moreover, we will summarize the available DeepFake dataset trends, focusing on their improvements. Additionally, the problem of how DeepFake detection generalizes the DeepFake detection model will also be analyzed. Finally, issues related to creating and detecting DeepFake will be discussed. We hope that the information contained in this survey will shed light on the use of deep learning for DeepFake detection in facial images and videos.

The research paper [2] 'Optical Stream-Based CNN for Detection of Unseen Deepfake Manipulations' delves into the emerging threat of deepfakes, which are realistic but fabricated videos created using AI technology. As AI continues to make it easier to produce these deceptive videos, it's crucial to develop methods to identify them, especially for various research purposes. The study introduces new statistical techniques for distinguishing between fake and authentic video sequences, showing improved performance compared to existing methods that usually analyze only a single video frame. Moreover, the proposed optical-based detection method demonstrates strong performance in enhancing crime detection operations and can be integrated with standard procedures to increase its overall effectiveness.In addition, the Deepfake Discovery Challenge (DFDC) Dataset was created to address the growing threat posed by deepfakes and other GANbased facial manipulations. This dataset, the largest of its kind to date, comprises over 100,000 video clips featuring 3,426 actors, and was generated using several non-academic, deep GAN-based models. To combat this issue, a competition named the DeepFake Discovery Challenge (DFDC) was launched on Kaggle, attracting widespread participation and resulting in the enhancement of the facial replacement dataset.The paper provides a detailed account of the dataset creation process and offers insights into top submissions in Kaggle competitions. Although deepfake detection remains a highly challenging and unresolved issue, the study demonstrates that the deepfake detection model considered by DFDC can provide in-depth analysis of videos, serving as a valuable tool for identifying potential issues in video content.

Article [3] - Deepfake Detection Challenge (DFDC) Dataset - Deepfakes is the latest version of technology that allows anyone to replace two numbers in a video. Besides deepfakes, various GAN-based facial manipulations have also been released with codes. To address this emerging threat, we created a massive video exchange database to demonstrate the detection model and launched a DeepFake Detection Challenge (DFDC) competition with Kaggle. Importantly, all recorded data agreed to participate and have their appearance updated during the creation of the face replacement dataset. The DFDC dataset is the largest face-swapping video dataset to date, containing over 100,000 clips from 3,426 paid actors, created using several non-academic, deep GAN-based models. In addition to explaining the process used to generate the data, we also provide detailed information about the top submissions in Kaggle competitions. Although deep search is a very difficult and still unsolved problem, we show that the deep search model studied by DFDC can "crazy" video in depth and that such models can become a useful deep analysis tool when identifying potential problems. learn video

The paper [4] propose a multi-modal detection for deepfake videos, called the Incompatibility Between Multiple Modes (IBMM) detection. The detection algorithm can detect whether the video is real or fake, and may be embedded in the monitoring equipment in the future. The model uses EfficientNet and simple 3D-CNN to recognize three types of deep videos. In the facial motion mode and lip motion mode, we use the EfficientNet for feature learning. This network uses a series of fixed scaling coefficients to scale the dimensions of the network uniformly and achieves good results in learning image features. In the audio mode, we adopt 3D-CNN network to train the hot coding diagram of audio data. Besides, for a single mode, we use the cross-entropy loss to calculate the irrationality of the mode. Negative ratio is used to calculate differences between models, such as the difference between lips and voice for different models. Experimental results show that, compared with other existing fake detection methods, the method presented in this paper has higher accuracy (95.87%) on DFDC datasets. And compared with the existing methods, the accuracy increases by 5.21%.

The paper [5] "a machine learning based free software tool has made it easy to create believable face swaps in videos that leaves few traces of manipulation, in what are known as "deepfake" videos. It is easy to imagine situations in which these fake videos could be used to create political pressure, to blackmail someone, or to fake. This paper proposes a temporal-aware pipeline to automatically detect deepfake videos. Our system uses a convolutional neural network (CNN) to extract frame-level features. Those features are then used to teach a recurrent neural community (RNN) that learns to classify if a video has been issue to manipulation or not. We evaluate our method towards a large set of deepfake motion pictures amassed from multiple video web sites. We show how our device can achieve competitive consequences on this undertaking while the usage of a easy architecture.

3. CONCLUSIONS

In conclusion, effectively employs a Convolutional Neural Network (CNN) model to discern between genuine and deepfake images, providing a robust means of identifying manipulated content. By training on a curated dataset comprising paired images, the model achieves high accuracy in distinguishing authentic images from altered ones. This underscores the critical role of machine learning in combating the proliferation of deepfake media, which poses significant risks to various domains, including journalism, entertainment, and national security. Continued research and innovation are imperative to stay ahead of evolving deepfake techniques and protect the integrity of digital content. Moreover, collaborations between academia, industry, and policymakers are essential to develop comprehensive strategies for mitigating the adverse impacts of deepfake technology on society.

ACKNOWLEDGEMENT

Authors we extend our sincere gratitude to our supervisor for their invaluable guidance, support, and encouragement throughout this project. Their expertise and insights have been pivotal in shaping the direction of our research. We also express appreciation to the researchers and organizations who provided access to datasets, software tools, and resources crucial for our experiments, enriching the quality of our work. Furthermore, we acknowledge the unwavering support of our colleagues, friends, and family members who stood by us with encouragement and understanding during challenging phases. Finally, we thank the academic community and society for fostering an environment conducive to collaborative research and innovation. This project would not have been possible without the collective efforts and contributions of all involved parties.

REFERENCES

[1] A. Malik, M. Kuribayashi, S. M. Abdullahi, and A. N. Khan, “DeepFake detection for human face images and videos: A survey, ”IEEE Access, vol. 10, pp. 18757–18775, 2022

[2] R.Caldelli, L.Galteri, I.Amerini, and A.D.Bimbo,“Opticalflowbased CNN for detection of unlearnt deepfake manipulations, ”Pattern Recognit. Lett., vol. 146, pp. 31–37, Jun. 2021.

[3] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. C. Ferrer, “The deepfake detection challenge (DFDC) dataset, ”2020, arXiv:2006.07397

[4] Y. Zhang, J. Zhan, W. Jiang, and Z. Fan, “Deepfake detection based on incompatibility between multiple modes, ”in Proc. Int. Conf. Intell. Technol. Embedded Syst. (ICITES), Oct. 2021, pp. 1–7.

[5]D. Guera and E. J. Delp, “Deepfake video detection using recurrent neural networks, ”in Proc. 15th IEEE Int. Conf. Adv. Video Signal Based Surveill. (AVSS), Nov. 2018, pp. 1–6. VOLUME 11, 2023

[6] F. Marra, D. Gragnaniello, D. Cozzolino, and L. Verdoliva, “Detection of GAN-generated fake images over social networks, ”in Proc. IEEE Conf.Multimedia Inf. Process. Retr. (MIPR), Apr. 2018, pp. 384–389..

[7] D. Yadav and S. Salmani, “Deepfake: A survey on facial forgery technique using generative adversarial network, ”in Proc. Int. Conf. Intell. Comput. Control Syst. (ICCS), May 2019, pp. 852–857..

8] F.F.Kharbat, T.Elamsy, A.Mahmoud,and R.Abdullah,“Imagefeature detectors for deepfake video detection, ”in Proc. IEEE/ACS 16th Int. Conf. Comput. Syst. Appl. (AICCSA), Nov. 2019, pp. 1–4. .