Video Forensic Using Deep Learning

Santosh Ghawate, Kailash Choudhary, Manisha Mehrotra
Student, Student, Guide
Computer Department
Dhole College Of Engineering, Pune, India

Abstract – The contemporary world is witnessing a rapidsurge in the adoption of Deep Learning (DL) applications. Deep Learning has proven instrumental in addressing critical issues suchas big data analysis, computer vision, and human brain interfacing. However, this advancement in technology also brings forth potential threats to privacy, democracy, and national security on both national and international scales. A notable example is the proliferation of deepfake videos, generated using artificial intelligence, which poses significant risks to political, social, and personal spheres.Deepfake videos are becoming increasingly sophisticated, often deceiving even well-trained observers. These videos, known for their potential to tarnish reputations, raise serious concerns among the general public. Consequently, it is imperative to develop effective methods for detecting deepfakes. This survey paper not only explores various deepfake creation algorithmsbut also delves into the methodologies proposed by researchers for detecting deepfakes.

The paper comprehensively addresses the challenges and current trends in the field, shedding light on the potential impact of deepfake technology on privacy, democracy, and national security. Furthermore, it outlinesfuture directions for the evolution of deepfaketechnology. The survey provides a holistic overview of deepfake creation approaches and emphasizes the need for the implementation of novel and reliable methodsto counter the complexities associated with deepfakes. By investigating the background of deepfakes and examining state-of-the-art deepfake detection methods, this paper aims to contribute to the development of robust solutions for addressing the challenges posedby deepfake technology.

Index Terms - Deepfake, Convolution Neural Network, Recurrent Neural Network, LSTM, Resnext, Generative Adversarial Network(GAN), Deep Learning

I. INTRODUCTION

In the contemporary digital landscape, safeguarding sensitive information and securing online identities is of utmost importance. The conventional reliance on passwords for digital account security has shown vulnerabilities, necessitating exploration into more sophisticated and dependable methods of user authentication [1]. Within this context, deep learning, a subset of artificial intelligence, has emerged as a promising solution, particularly in the realm of facial authentication for login processes [2].

This review aims to thoroughly examine the application of deep learning methodologies, with a specific focus on convolutional neural networks (CNNs) [3]. By tracing the evolution and implementation of these techniques, we aim to assess their efficacy in enhancing the security of online platforms while ensuring a user-friendly experience. A critical evaluation of existing research will shed light on both the advantages and challenges associated with integrating deep learning- based face authentication systems [2].

Additionally, this review delves into the implications of these technologies on overall user experience, highlighting their potential to streamline and fortify digital authentication processes [4]. By synthesizing insights from various scholarly works, we aim to contribute to the ongoing discourse regarding the importance of incorporating advanced technologies to address contemporary security challenges, paving the way for a more secure and seamless digital future.

II. LITERATURE SURVEY

1. Dataset

To built any machine learning and deep learning model we require a real-world data. First we collected data from different platform like Kaggle's Deepfake Detection challenge[7], Celeb-DF[8], Face Forensic ++[9].
The Kaggle DeepFake Detection challenge dataset comprises 3000 videos, evenly split between real and manipulated data. Celeb-DF, featuring videos of renowned celebrities, contains a total of 1000 videos, half of which are real, and the other half manipulated. The Face Forensic ++ dataset includes 2000 videos, with 1000 being real and the rest
manipulated. Subsequently, these three
datasets were amalgamated and subjected to data preprocessing.

2. Face-based Video Manipulation Methods.

Various approaches targeting face manipulations in video sequences have been proposed since the 1990s. Thies et al. pioneered real-time expression transfer for faces and introduced Face2Face, a real-time facial reenactment system capable of altering facial movements in diverse video streams.

Alternative methods to Face2Face have also been suggested.

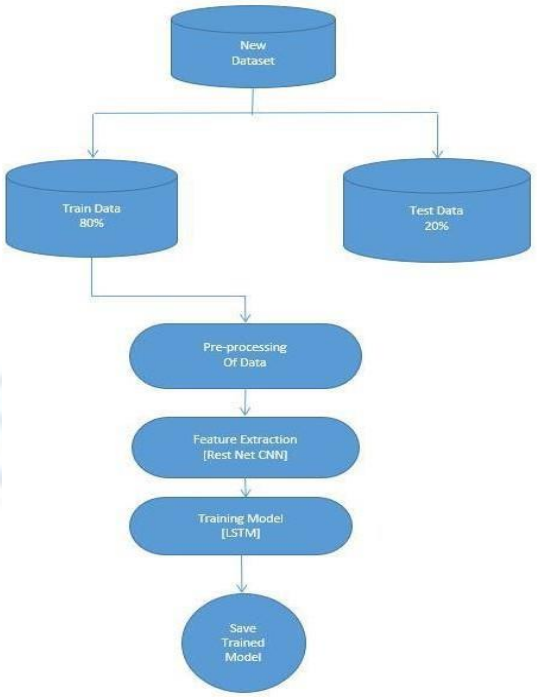
Deep learning techniques, particularly generative adversarial networks (GANs), have been employed for face image synthesis, including aging alterations and modifications to attributes like skin color. Deep feature interpolation has demonstrated notable success in altering face attributes such as age, facial hair, or mouth expressions. Despite some limitations, recent advancements, such as progressive GANs introduced by Karras et al., have significantly improved the synthesis quality of faces.

3. Creating Deepfake

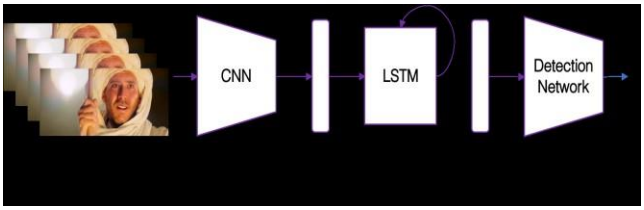
Deep learning techniques, known for enhancing image compression performance, have been employed, notably autoencoders for dimensionality reduction and generative models learning. Autoencoders extract compressed representations of images, forming the basis for face-swapping capabilities. Two sets of encoder-decoders with shared weights for the encoder networks play a pivotal role in this process.

ResNextCNN for Feature Extraction: The ResNext CNN classifier is proposed for extracting features and reliably recognizing frame-level characteristics. Fine-tuning the network involves adding extra layers as needed and setting an appropriate learning rate to ensure proper convergence of the model's gradient descent.

LSTM for Sequence Processing: In the context of a 2-node neural network with probabilities as input and ResNext CNN feature vectors as output, an LSTM with 2048 units and a 0.4 dropout likelihood is proposed for sequence processing. This LSTM facilitates a meaningful temporal analysis of the video frames by sequentially comparing them, addressing the challenge of recursively processing a sequence.



Recurrent Network for Deepfake Detection: This section outlines our comprehensive end- to-end trainable recurrent deepfake video detection system, which incorporates a convolutional LSTM structure for processing frame sequences. The fundamental components of this structure include a CNN for frame feature extraction and an LSTM for temporal sequence analysis.



III. RESULT

It is not uncommon to encounter deepfake videos where manipulation is confined to a small portion of the video, such as when the target face appears briefly. To address this, we extract continuous subsequences of fixed frame length from every video in the training, validation, and test sets, serving as input to our system.

The entire pipeline undergoes end-to-end training until we reach a 10-epoch loss plateau in the validation set. Our results demonstrate that, with less than 2 seconds of video (equivalent to 40 frames for videos sampled at 24 frames per second), our system achieves an accuracy exceeding 97% in accurately predicting whether the analyzed fragment originates from a deepfake video or not.

IV. CONCLUSIONS

This paper introduces a temporal-aware system designed to automatically detect deepfake videos. Our experimental findings, based on a substantial collection of manipulated videos, indicate that a simple convolutional LSTM structure enables accurate prediction of video manipulation with as little as 2 seconds of video data.

We posit that our work establishes a robust initial line of defense for identifying fake media created using the described tools. The demonstrated competitive results, achieved through a straightforward pipeline architecture, underscore the effectiveness of our system.

In future research, our focus will be on enhancing the robustness of our system against manipulated videos using techniques not encountered during training.

V. REFERENCES

[1] M. Abadi et al. Tensorflow: A system for large-scale machine learning. Proceedings of the USENIX Conference on Operating Systems Design and Implementation, 16:265–283, Nov. 2016. Savannah, GA.

[2] G. Antipov, M. Baccouche, and J.-L. Dugelay. Face aging with conditional generative adversarialnetworks. rXiv:1702.01983, Feb. 2017.

[3] H. Averbuch-Elor et al. Bringing portraits to life. ACM Transactions on Graphics, 36(6):196:1–196:13, Nov. 2017.

[4] P. Bestagini et al. Local tampering detection in video sequences. Proceedings of the IEEE International Workshop on Multimedia Signal Processing, pages 488– 493, Sept. 2013. Pula, Italy.

[5] Koopman, M., Rodriguez, A. M., and Geradts, Z. (2018). Detection of deepfake video manipulation. In The 20th Irish Machine Vision and Image Processing Conference (IMVIP) (pp. 133-136).

[6] C. Bregler, M. Covell, and M. Slaney. Video rewrite: Driving visual speech with audio. Proceedings of the ACM Annual

[7] Bayar, B., and Stamm, M. C. (2016, June). 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 (pp. 5-10).ACM.

[8] Yang, W., Hui, C., Chen, Z., Xue, J. H., and Liao, Q. (2019). FV-GAN: Finger vein representation using generative adversarial networks. IEEE Transactions on Information Forensics and Security, 14(9), 2512-2524

[9] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv:1412.6980, Dec. 2014.

[10] N. Rahmouni, V. Nozick, J. Yamagishi, and I. Echizen. Distinguishing computer graphics from natural images using convolution neural networks. Proceedings of the IEEEWorkshop on Information Forensics and Security, pages 1–6, Dec.2017. Rennes, France

[11] N. Rahmouni, V. Nozick, J. Yamagishi, and I. Echizen. Distinguishing computer graphics from natural images using convolution neural networks. Proceedings of the IEEEWorkshop on Information Forensics and Security, pages 1–6, Dec. 2017. Rennes, France

[12] J. Thies et al. Face2Face: Real-time face capture and reenactment of rgb videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 2387– 2395, June 2016. Las Vegas, NV.

[13] M. Brundage et al. The malicious use of artificial intelligence: Forecasting, prevention, and mitigation. arXiv:1802.07228, Feb. 2018

[14] D. G¨uera, S. K. Yarlagadda, P. Bestagini, F. Zhu, S. Tubaro, and E. J. Delp. Reliability map estimation for cnn-based camera model attribution. Proceedings of the IEEE Winter Conference on Applications of Computer Vision, Mar. 2018. Lake Tahoe, NV.

[15] Thanh Thi Nguyen, Cuong M. Nguyen, Dung Tien Nguyen, Duc Thanh Nguyen, Saeid Nahavandi, " Deep Learning for Deepfakes Creation and Detection: A Survey ", Jul 2020.