WWW.IJRPR.COM

International Journal of Research Publication and Reviews

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

Deepfake Video Detection using Neural Networks

1st Mr. Vaibhav Nangare,2nd Prof. Bhosale Sachin,3rd Prof. Dr. Khatri Anand,4th Prof. Mundhe Bhalchandra

[1] Computer Engineering Jaihind College of Engineering, Kuran, India
vaibhavnangare5@gmail.com
[2] Computer Engineering Jaihind College of Engineering, Kuran, India
sachinbhosale@gmail.com
[3] Computer Engineering Jaihind College of Engineering, Kuran, India
khatrianand@gmail.com
[4] AI & DS Engineering Jaihind College of Engineering, Kuran, India
mundheraj.mundhe@gmail.com

ABSTRACT: —

Deepfake Video Detection using Neural Networks involves using machine learning models, particularly neural networks, to identify manipulated videos created with deepfake technology. Deepfakes use advanced AI algorithms, typically based on deep learning, to generate realistic fake videos, images, or audio. Detecting these fake videos is crucial, especially in contexts like politics, social media, and entertainment, to avoid misinformation, fraud, or malicious content. Deepfake video detection using neural networks relies on sophisticated models like CNNs, RNNs, and autoencoders to analyze visual and temporal data for inconsistencies. As deepfake technology evolves, these models must continuously adapt and improve to remain effective. Neural networks, especially CNNs and RNNs, are powerful tools in detecting deepfake videos. They help spot visual and temporal inconsistencies, but continuous development is needed to keep up with advancing deepfake technology.

Keywords: Repeated neural networks (RNN), folding networks (CNN), and deep-fark video detection.

Introduction

Deepfake technology has advanced rapidly, raising significant concerns regarding misinformation, privacy, and digital security. Detecting deepfake videos has become a critical research area, with neural networks demonstrating promising results. This paper explores various deepfake detection methods utilizing deep learning models, including Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), and transformer-based architectures. We provide an in-depth analysis of datasets, training methodologies, and performance evaluation metrics. Experimental results indicate that neural networks can achieve high accuracy in distinguishing authentic videos from deepfakes, though challenges remain in generalization and robustness. Deepfake technology leverages artificial intelligence to manipulate audiovisual content, creating highly realistic but falsified media. The proliferation of deepfake videos poses threats to cybersecurity, public trust, and digital forensics. Detecting deepfakes is crucial for mitigating their impact, and neural networks have emerged as a leading approach in this domain. This paper explores various deep learning techniques used in deepfake video detection, emphasizing their strengths, limitations, and future directions. Several deep learning-based approaches have been proposed for deepfake detection. Early methods relied on handcrafted features and traditional machine learning classifiers. With advancements in deep learning, CNNs, Long Short-Term Memory (LSTM) networks, and transformer models have demonstrated superior performance in detecting deepfake artifacts. Recent studies highlight the importance of using large-scale datasets, transfer learning, and adversarial training to enhance model robustness.
Numerous research efforts have explored deepfake detection using deep learning. Early methods relied on traditional classifiers, while recent approaches leverage CNNs, LSTMs, and transformer models. Studies emphasize the importance of large datasets, adversarial training, and hybrid architectures to improve detection performance.

Literature survey

Deepfake videos, generated using deep learning techniques, pose significant threats to digital security, privacy, and information integrity. The rapid advancement of generative adversarial networks (GANs) and autoencoders has enabled the creation of highly realistic fake videos, making their detection a critical research area. This literature survey explores various neural network-based methodologies for deepfake detection, highlighting key approaches, challenges, and recent advancements.

CNNs have been widely used for image-based deepfake detection due to their ability to extract spatial features from video frames. Several studies leverage CNN architectures such as: **XceptionNet:** Rossler et al. (2019) proposed using XceptionNet for deepfake detection, achieving high accuracy on the FaceForensics++ dataset.

EfficientNet: This lightweight yet powerful CNN has demonstrated superior performance in detecting manipulated facial features in deepfake videos.

ResNet: Variants of ResNet have been used to capture fine-grained artifacts introduced by deepfake generation algorithms.

Given that deepfake videos are temporal sequences, RNNs and LSTMs are employed to analyze frame-to-frame inconsistencies. **Li et al. (2020)** utilized LSTMs in conjunction with CNN feature extractors to detect temporal artifacts in video sequences. **BiLSTM** networks have been employed to enhance the detection of subtle facial movement inconsistencies.

Their approach is based solely on not flashing as an identification note. To recognize the deep pope, several additional factors must be considered, such as facial creases and dental magic. It is recommended to use the method to take all of these factors into consideration. To identify modified photos and videos in many contexts, for example, B. detection of repeat attacks and computer-aided video detection can be used to use techniques used as "using capsule networks to recognize forged images and videos". Even if the model works well with data records, noise in the training process can lead to poor performance on actual data. Biological signals from the facial regions of the substantial and false portrait paintings are extracted using recognition of synthetic portrait videos using biological signals. To calculate temporal consistency and spatial coherence, apply CNN and probability SVM, train and train CNN and probability SVM.

Proposed system

Their approach is based solely on not flashing as an identification note. To recognize the deep pope, several additional factors must be considered, such as facial creases and dental magic. It is recommended to use the method to take all of these factors into consideration. To identify modified photos and videos in many contexts, for example, B. detection of repeat attacks and computer-aided video detection can be used to use techniques used as "using capsule networks to recognize forged images and videos". Even if the model works well with data records, noise in the training process can lead to poor performance on actual data. Biological signals from the facial regions of the substantial and false portrait paintings are extracted using recognition of synthetic portrait videos using biological signals. To calculate temporal consistency and spatial coherence, apply CNN and probability SVM, train and train CNN and probability SVM. The proposed system intends to create an artificial intelligence-powered deepfake video detection model through the use of neural networks. The system will utilize advanced deep learning techniques to examine video frames, identify facial features, and determine the authenticity of the video.

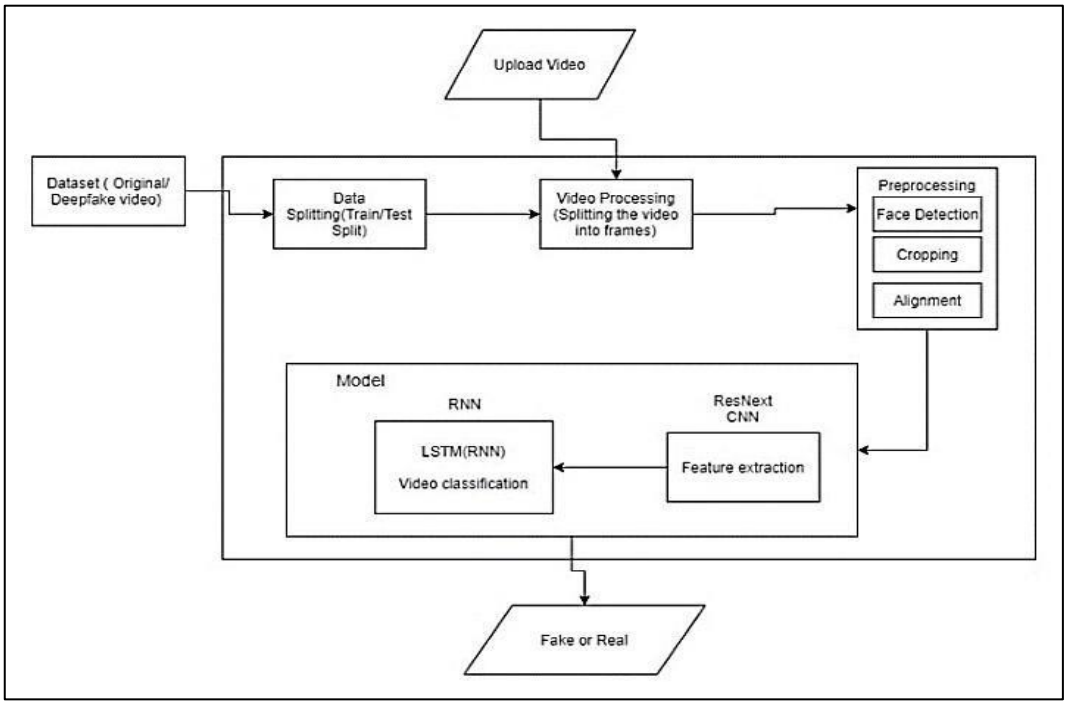

Fig. 1: System Architecture

Deepfake videos using neural networks are intended to automatically identify videos that have been manipulated using deep learning techniques. Deep learning models (such as reset, efficient Cynet, XceptionNet, or Vision Transformers) are trained to classify videos as real or fake. A Vision Transformer (VIT) or self-machine can help you identify subtle inconsistencies in your video. Ensemble learning techniques combine several deep learning models to improve identification.

The proposed method uses spatial and temporal features to realize fake content. Experimental results show a high level of accuracy and robustness for a variety of techniques for deeper generation. The rapid development of AI-generated videos using geese (such as Stylegan, Deepfake, Faceswap) distinguishes real content from fake content.
Continuous improvements, including controversial training and data records expansion, will further improve effectiveness in real-world applications.

A. ***Dataset:***

Out of the data records, 30% are utilized for testing purposes, while the remaining 70% are employed for training. Developing a deepfake recognition model necessitates a substantial dataset comprising genuine and fabricated videos. Some frequently encountered data sets include faceforensics++, deepfake detection challenge dataset, and celeb-df (celeb deepfake).

B. ***Preprocessing:***

As part of the initial data preparation, the video is segmented into individual frames. Following that, the face is identified, and the frame containing the detected face is cropped. Processing the 10-second movie at 30 frames per second would require a substantial amount of computing power, totaling 300 frames.

C. ***Model:***

This model comprises the lstm layer, following the architecture of resnext50_32x4d. Furthermore, the model is provided with frames in small quantities from the processed video for training and evaluating its performance. The process of deepfake video recognition using neuronal networks involves training a model that can differentiate between genuine videos and those that have been manipulated.

D. ***ResNext CNN for Feature Extraction:***

ResNext is a "deep diagram neural network" that improves reset by inspiring the concept of grouped convolution inspired by recording models. A born regeneration model (usually trained with Imagenet) and final classification layer removal are used. Next, add the required additional layers and remove the network by selecting the correct learning rate to ensure that the gradient descent of the model converges correctly. The 2048-dimensional characteristic vector of the billiard layer is fed sequentially to the LSTM.

E. ***LSTM for Sequence Processing:***

Assume that a sequence of resnext cnn feature vectors of input frames is fed into a two-node neural network. The likelihood of the sequence suggest whether it is a genuine or deepfake movie. Our primary challenge is to create a model that can effectively process a sequence of events in a logical and coherent manner. By comparing the frame captured at "t" second with the frame captured at "t-n" seconds, the temporal analysis of the video can be conducted by utilizing lstm to process the frames in a sequential manner. The number of frames preceding t, n can vary and is not limited to a specific value.

F. ***Evaluation Metrix:***

We will evaluate the system's performance by considering metrics such as accuracy, precision, recall, and f1-score. he region below the curve (auc) of the receiver operating characteristic (roc) curve. Frame-level and video-level detection rates.Robustness against adversarial deepfake attacks.

G. ***Provisioning and actual time detection:***

Set up the trained model as a web-based or mobile application. Optimize the model using quantization and cut for faster conclusions.

Predict:

Additionally, new videos are pre-processed to introduce a trained model format. The video is split into frames, followed by face cuts, and instead of storing the video in local memory, the abbreviated frames are handed over directly to the model trained for detection.

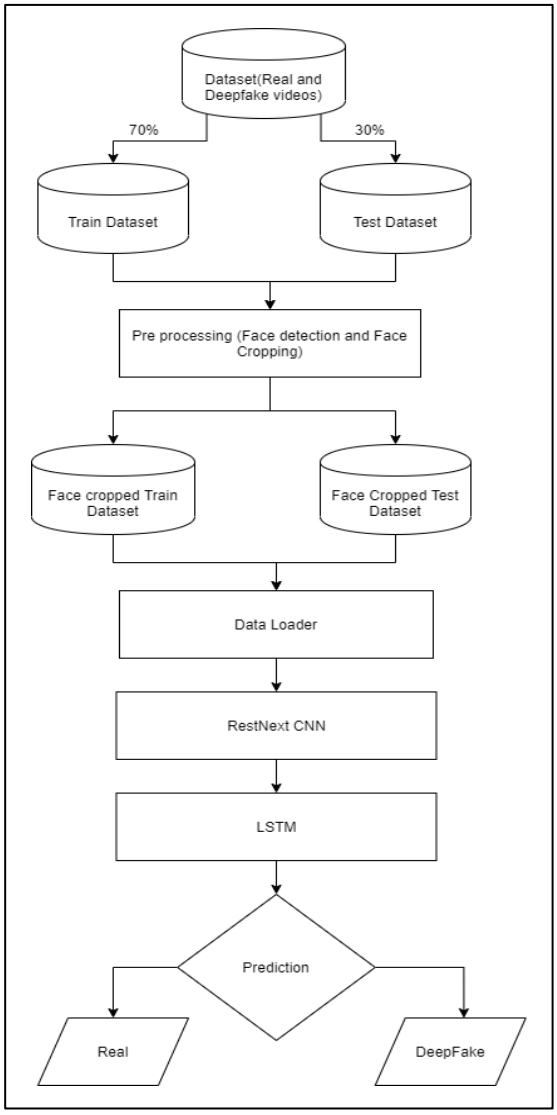

Fig. 2: Training Flow

RESULT

The purpose of this proposed system is to create reliable and efficient Deepke Pake frames with advanced deep learning techniques. One example is shown in the figure 3.

Fig. 3: Expected Results

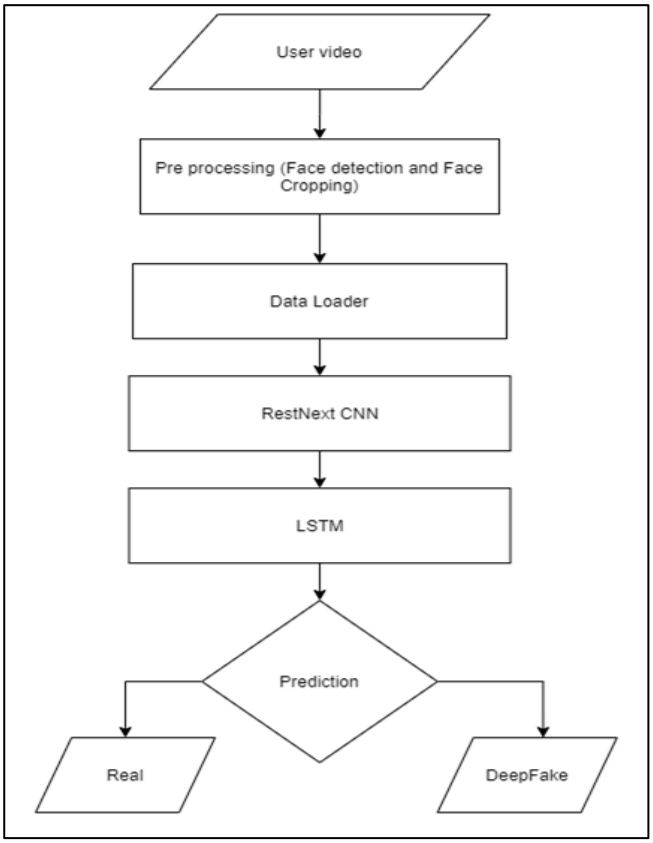

Fig. 4: Prediction flow

Conclusion

Our approach uses ResNext CNN for frame level and RNN and LSTM detection, and uses it for video classification. The criteria mentioned in this study allow the proposed approach to determine whether the video is real or profound. We believe that the actual data will be extremely accurate. Deepfake video detection using Neural Networks is a developed field where AI plays a key role in combating the growing abuse of generated media. This study highlights the effectiveness of deep learning models such as foldable networks (CNNS), recurrent neural networks (RNNS), and trans-based architectures in the identification of Deepepake videos. These models analyse temporal and spatial inconsistencies, pixel-level anomalies, and subtle facial distortions to distinguish fake content from real videos.

Limitations

The audio has not been accounted for in our method. Our approach won't be able to identify the audio deepfake because of this. However, we are putting up the idea of detecting audio deepfakes in the future.

REFERENCES

1. Yuezun Li, Siwei Lyu, ―ExposingDF Videos By Detecting Face Warping Artifacts,‖ in arXiv:1811.00656v3.
2. Yuezun Li, Ming-Ching Chang and Siwei Lyu
3. ―Exposing AI Created Fake Videos by Detecting Eye Blinking‖ in arxiv.
4. Huy H. Nguyen , Junichi Yamagishi, and Isao Echizen
5. ― Using capsule networks to detect forged images and videos ‖.
6. Hyeongwoo Kim, Pablo Garrido, Ayush Tewari and Weipeng Xu ―Deep Video Portraits‖ in arXiv:1901.02212v2.
7. Umur Aybars Ciftci, ˙Ilke Demir, Lijun Yin ―Detection of Synthetic Portrait Videos using Biological Signals‖ in arXiv:1901.02212v2.
8. Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In NIPS, 2014.
9. David G¨uera and Edward J Delp. Deepfake video detection using recurrent neural networks. In AVSS, 2018.
10. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, 2016.

11. An Overview of ResNet and its Variants : https://towardsdatascience.com/an-overview-of-resnet- and-its-variants-5281e2f56035
12. Long Short-Term Memory: From Zero to Hero with Pytorch: https://blog.floydhub.com/long-short-term- memory-from-zero-to-hero-with-pytorch/
13. Sequence Models And LSTM Networks https://pytorch.org/tutorials/beginner/nlp/sequence_mod els_tutorial.html
14. https://discuss.pytorch.org/t/confused-about-the-image- preprocessing-in-classification/3965
15. https://www.kaggle.com/c/deepfake-detection- challenge/data
16. https://github.com/ondyari/FaceForensics
17. Y. Qian et al. Recurrent color constancy. Proceedings of the IEEE International Conference on Computer Vision, pages 5459–5467, Oct. 2017. Venice, Italy.
18. P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to- image translation with conditional adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 5967–5976, July 2017. Honolulu, HI.
19. R. Raghavendra, Kiran B. Raja, Sushma Venkatesh, and Christoph Busch, ―Transferable deep-CNN features for detecting digital and print-scanned morphed face images,‖ in CVPRW. IEEE, 2017.
20. Tiago de Freitas Pereira, Andr´e Anjos, Jos´e Mario De Martino, and S´ebastien Marcel, ―Can face anti spoofing countermeasures work in a real world scenario?,‖in ICB. IEEE, 2013.
21. Nicolas Rahmouni, Vincent Nozick, Junichi Yamagishi, and Isao Echizen, ―Distinguishing computer graphics from natural images is using convolution neural networks,‖ in WIFS. IEEE, 2017.
22. F. Song, X. Tan, X. Liu, and S. Chen, ―Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients,‖ Pattern Recognition, vol. 47, no. 9, pp. 2825–2838, 2014.
    a. D. E. King, ―Dlib-ml: A machine learning toolkit,‖ JMLR, vol. 10, pp. 1755–1758, 2009.