iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 12 **Issue:** IV **Month of publication:** April 2024

DOI: https://doi.org/10.22214/ijraset.2024.60765

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 12 Issue IV Apr 2024- Available at www.ijraset.com

Exposing Fake Faces Through Deep Neural Networks Combining Content and Trace Feature Extractors

Keerthi. P[1], MJ Mohammed Yunus[2], Ujjwal Joshi[3], Papan Roy[4], Kennis. M[5]
[1]*Assistant Professor, Dept of CSE Impact College of Engineering and Applied sciences, Bangalore, Affiliated to VTU*
[2, 3, 4, 5]*Students, Dept of CSE Impact College of Engineering and Applied sciences, Bangalore, Affiliated to VTU*

Abstract: ***In recent times, the proliferation of free deep learning-based software has facilitated the emergence of convincing facial swaps in videos, commonly referred to as 'DeepFake' (DF) videos. 'Deep learning' has improved the realism and accessibility of creating fake digital video content, which was previously attainable through traditional visual effects. These AI-generated media, often referred to as DF, present a dual challenge: their creation is relatively straightforward using AI tools, yet their detection poses a significant hurdle. We address this challenge by employing Convolutional Neural Networks (CNNs) and 'Recurrent Neural Networks' (RNNs) to identify 'DFs'. Specifically, our system utilizes a CNN to obtain frame level characteristics and apply them to train an RNN capable of identifying temporal inconsistencies introduced by DF creation tools. We evaluate our approach on a substantial dataset of fake videos and demonstrate competitive performance with a straightforward architecture.***

Keywords: DeepFake, Convolutional Neural Networks (CNNs), Recurrent Neural Networks (RNNs), Video manipulation, AI detection.

I. INTRODUCTION

The advancement of smartphone camera technology and the widespread availability of high-speed internet globally have greatly expanded the scope of Internet usage platforms and media sharing websites, facilitating the effortless creation and dissemination of digital videos. The increasing computational capabilities have enabled significant advancements in deep learning, surpassing what was once deemed unattainable just a few years ago.

However, with transformative technologies come new challenges, such as the rise of "DeepFake" content generated by sophisticated deep generative adversarial models capable of manipulating video and audio clips. DeepFake content is being widely shared on networking sites, leading to issues like spamming and the dissemination of misinformation. Such Deep Fakes can have dire consequences, misleading and threatening the general public.

Understanding the process by which 'Generative Adversarial Networks' (GANs) generate DeepFake content is essential for detection. GANs take a movie and a picture of a certain human (the 'target') as input to create another clip target's face replaced by that of a different individual (the 'source'). Deep 'adversarial neural networks' form the core of DeepFake generation, educated by human photos and goal videos to connect the source's facial traits to the target's. Through subsequent post-processing, the resultant films achieve an impressive degree of reality. The 'GAN' algorithm dissects the flim into frames, replacing the input picture in each frame, and then reconstructs the video, often employing auto encoders for this purpose.

We offer an approach to identifying DeepFake videos leverages the inherent properties of such videos. Due to computational limitations and production constraints, DeepFake algorithms can produce pictures of faces with a set dimension, necessitating an affine warping process to position the synthetic face with the target's configuration. This warping introduces discernible Iissues found in the resultant 'DeepFake' films, stemming from the resolution disparities between the warped facial area and its surroundings. Our detection method identifies these errors via comparison of computed facial portions to the surroundings, employing a combination of 'ResNext Convolutional Neural Networks' (CNNs) for gathering features and 'Recurrent Neural Networks' (RNNs) with 'Long Short-Term Memory' (LSTM) units to capture temporal inconsistencies introduced by GANs during the DeepFake reconstruction process. We trained the 'ResNext CNN model' by explicitly simulating resolution discrepancies in affine face wrappings.

II. LITREATURE SURVEY

The rapid proliferation of deep fake videos, along with their illicit utilization, poses a significant peril to democratic processes, legal systems, and societal trust. Consequently, this is an growing demand for the analysis, detection, and mitigation of counterfeit videos. Several methods have developed for recognizing deep fakes, such as the approach employed in 'Exposing DF Film via recognizing Facial Cracking Properties [1],' which leverages a specialized 'Convolutional Neural Network' to discern anomalies in facial regions and their surroundings. Another innovative method, as outlined in 'Exposing AI made counterfeit footage by recognizing eyes twitching [2],' focuses on detecting the absence of natural eye blinking, a physical indicator often absent in synthetic videos. However, the efficacy of such methods may be enhanced by considering additional parameters like dental characteristics and facial wrinkles. In 'Using Capsule networks identify counterfeit pictures and recordings. [3],' a 'capsule network' is utilized to identify manipulated visual content across various scenarios. While effective, this approach's reliance on random noise during training may limit its real-world applicability, unlike our proposed method trained on noise-free datasets. Similarly, in 'Detecting Artificial Webcam Films Utilizing Natural Indicators [5],' biological signals extracted from cheek areas are employed alongside advanced signal processing techniques to discern between real and phony videos. Finally, 'Fake Catcher' offers a promising solution for detecting counterfeit content, albeit facing challenges in preserving biological signals because of the absence of a discriminator. Addressing this limitation necessitates the formulation of a differentiable loss function aligning with the proposed signal processing steps.

III.PROPOSED SYSTEM

Numerous tools facilitate the creation of digital footprints (DF), yet their scarcity persists in the realm of DF detection. Our novel approach to DF detection promises to significantly mitigate DF dissemination across the internet. Introducing a user-friendly web-based platform, users can effortlessly upload videos for real-time classification as authentic or synthetic. This endeavour holds potential for expansion, evolving from a web foundation into an Internet Explorer extension for seamless DF identification. Notably, major platforms such as WhatsApp and Facebook could seamlessly integrate this solution for pre-emptive DF screening prior to content sharing. A pivotal objective entails rigorous assessment, spanning security, usability, precision, and dependability. Our methodology is tailored to discern various forms of DF, encompassing substitution, reduction, and interpersonal variations. Illustrated in Figure 1, our system architecture embodies simplicity and efficacy.

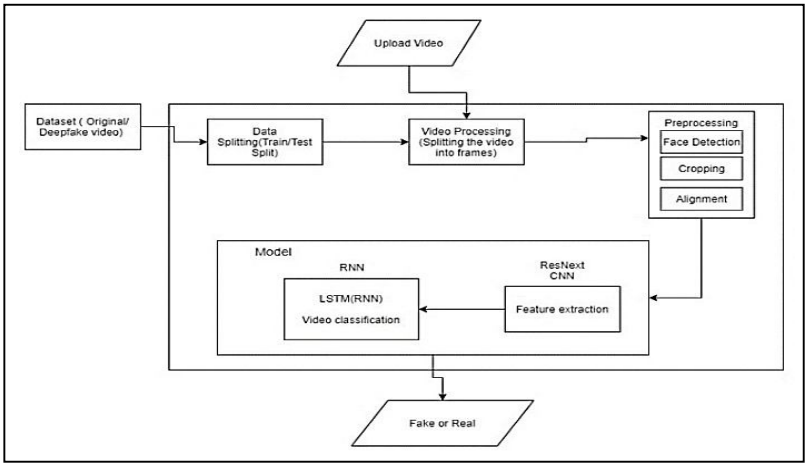


Figure 1: "System Architecture"

A. *Data Compilation*

Our dataset amalgamates videos sourced equitably from diverse platforms including YouTube, FaceForensics++, and the DeepFake Recognition Competition Collection. Comprising an equal distribution of genuine and manipulated content, this dataset is segregated into 70% training and 30% testing subsets.

B. *Pre-processing*

Initial pre-processing involves video segmentation into frames, subsequent face detection, and frame cropping to isolate detected faces. To ensure uniformity, frames are adjusted to be comparable the mean frame count. Frames devoid of facial features are excluded.

C. Model Framework

The proposed model employs a ResNext50_32x4d architecture followed by an LSTM layer. A Data Loader partitions pre-processed videos into learning and assessment sets, facilitating batch-wise model training and evaluation.

D. ResNext CNN

Figures Rather than crafting a new classifier, feature extraction leverages the 'ResNext CNN' to accurately capture frame-level attributes. Fine-tuning involves appending requisite layers and optimizing learning rates for effective convergence.

E. LSTM for Temporal Analysis

Addressing temporal dynamics, an LSTM unit processes frames sequentially, enabling contextual analysis by comparing current and past frames.

F. Prediction

Unfamiliar videos undergo pre-processing to align with the model's format before being forwarded for prediction, bypassing local storage in favour of direct frame-wise analysis.

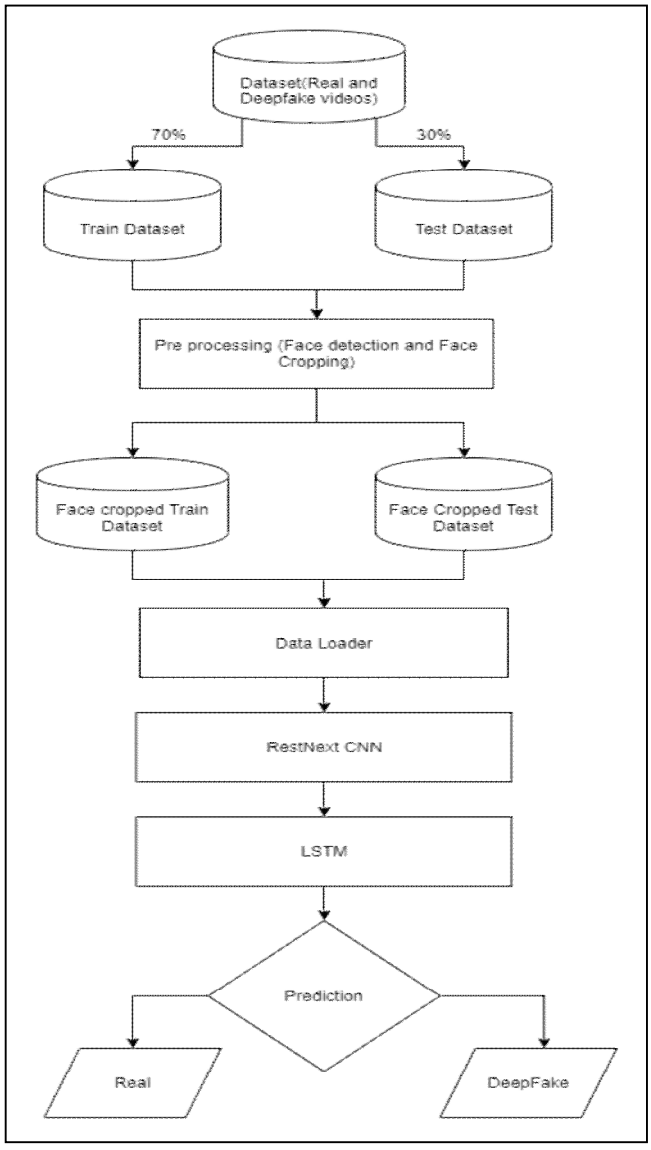


Figure 2: "Training Flow"

IV. RESULT

The model's output will determine if the flim is a deepfake or genuine, alongside the model's confidence level. An illustration of this is provided in 'Figure 3'.

Figure 3: "Expected Results"

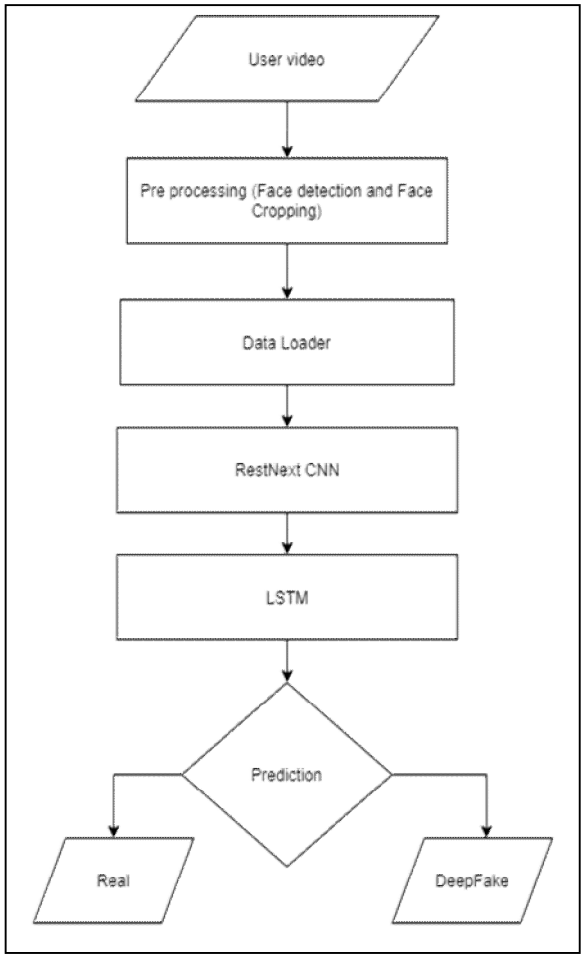


Figure 4: "Prediction flow"

V. CONCLUSIONS

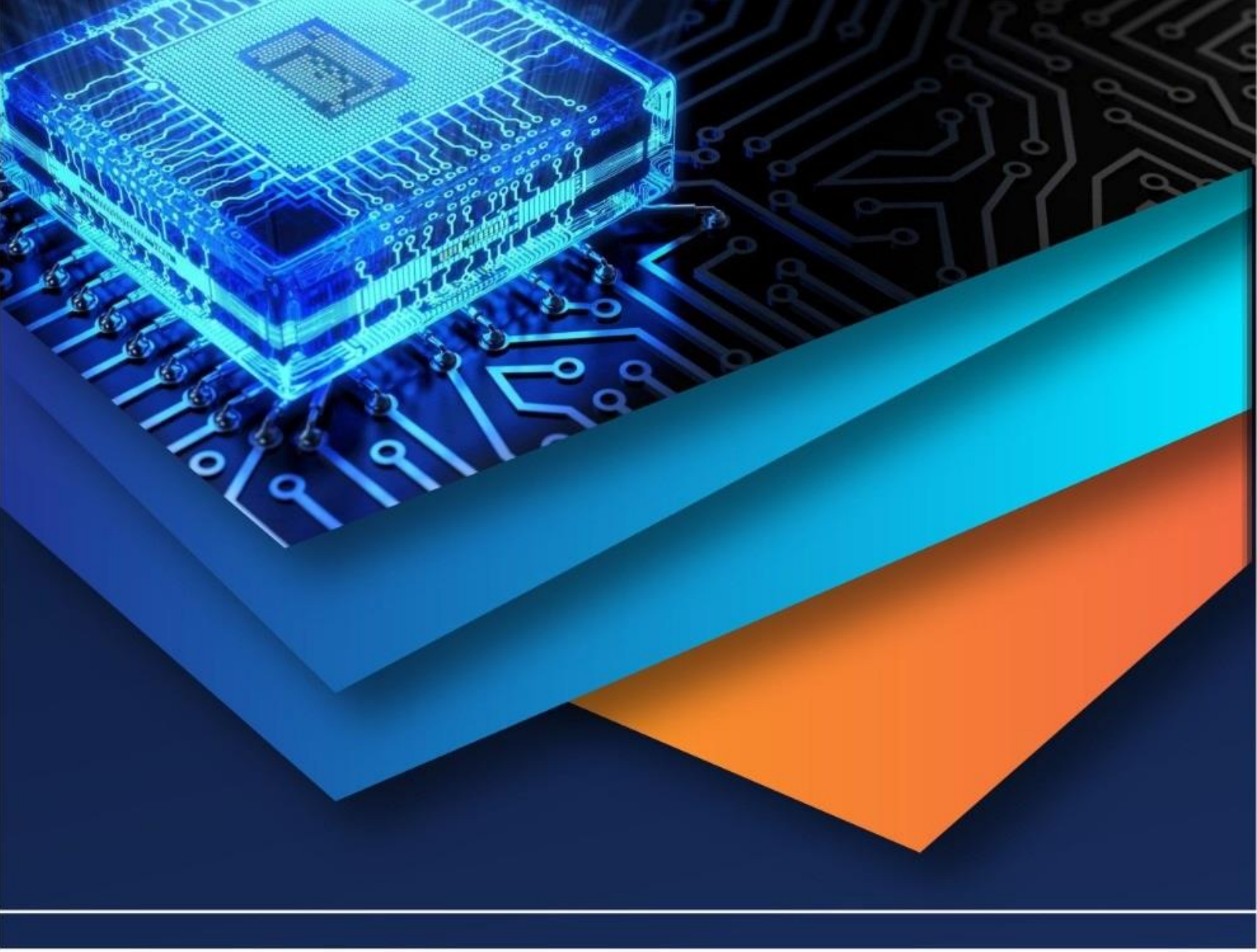
Our research introduces a neural network-driven technique for discerning between authentic and deep fake videos, incorporating a confidence metric for model assessment. Drawing inspiration from the synthesis process of deep fakes utilizing GANs and Autoencoders, our approach employs frame-level analysis utilizing ResNext CNN, coupled with video classification employing RNN and LSTM architectures. Our methodology demonstrates proficiency in identifying videos as genuine or deep fake, leveraging the specified parameters detailed in the paper. We anticipate achieving exceptional real-time accuracy with this model.

VI. ACKNOWLEDGMENT

The Without acknowledging the people who made the project possible, without whose unwavering support and guidance my efforts would be considered vain., the happiness that comes with its successful completion would be incomplete. As we completed our project on "Exposing Fake Faces Through Deep Neural Networks Combining Content and Trace Feature Extractor'" we felt delighted to thank and show thanks to everyone who helped us along the way. We appreciate all of the help and inspiration we received along the way from our guide, Mrs. Keerthi.P, Assistant Professor of 'CSE'. For her assistance and direction, we are appreciative of Dr. Dhananjaya V., Professor and Head of 'CSE'. We are grateful to the management and principal of Impact College of Engineering and Applied Sciences, Dr. JALUMEDI BABU, for providing us with the facilities and welcoming atmosphere that have allowed us to further our education. We'd also want to mention all of the teaching and non-teaching staff of the 'CSE'. We'd want to mention. our parents and friends for their gracious cooperation and support over the project's duration.

REFERENCES

[1] Yuezun Li, 'Siwei Lyu', "Exposing DF Videos ByDetecting 'Face Warping Artifacts'," in arxiv.
[2] Yuezun Li, 'Ming-Ching Chang' and 'Siwei Lyu' 'Exposing AI Created "Fake Videos" by Detecting Eye Blinking' in arxiv.
[3] Kaiming He, 'Xiangyu Zhang', 'Shaoqing Ren', and 'Jian Sun'. "Deep residual learning for image recognition". In CVPR, 2016.
[4] An 'Overview of "ResNet" and its Variants' :https://towardsdatascience.com/an-overview-of-resnet- and-its-variants-5281e2f56035
[5] "Long Short-Term Memory": From 'Zero to Hero' with "Pytorch": https://blog.floydhub.com/long-short-term-memory-from-zero-to-hero-with-pytorch/
[6] Sequence 'Models and LSTM Networks' https://pytorch.org/tutorials/beginner/nlp/sequence_models_tutorial.html
[7] https://discuss.pytorch.org/t/confused-about-the-image- preprocessing-in-classification/3965
[8] https://www.kaggle.com/c/deepfake-detection-challenge/data
[9] https://github.com/ondyari/FaceForensics
[10] Y. Qian et al. Recurrent color constancy.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY