e-ISSN: 2582-5208
International Research Journal of Modernization in Engineering Technology and Science
(Peer-Reviewed, Open Access, Fully Refereed International Journal)
Volume:07/Issue:05/May-2025 Impact Factor- 8.187 www.irjmets.com

DEEPFAKE VIDEO DETECTION USING NEURAL NETWORKS

Nikhil Shinde[*1], Jaydeep Nigade[*2], Yashkumar Bagal[*3], Rohan Avatade[*4], Rutuja Taware[*5]

[*1,2,3,4]Student, Department Of Computer Engineering, SVPM's College Of Engineering, Malegaon (Bk), Maharashtra, India.

[*5]Professor, Department Of Computer Engineering, SVPM's College Of Engineering, Malegaon (Bk), Maharashtra, India.

DOI : https://www.doi.org/10.56726/IRJMETS75444

ABSTRACT

In recent times, the emergence of free deep learning-based software tools has made it easier to generate highly realistic face-swapped videos, commonly known as "DeepFakes" (DF). While video manipulation has been possible for decades using traditional visual effects, recent breakthroughs in deep learning have significantly enhanced the realism of such content and lowered the barrier to creating it. These AI-generated videos, also known as AI-synthesized media, have become increasingly convincing and difficult to detect.

Although generating DeepFakes has become relatively straightforward with the help of artificial intelligence, detecting them remains a complex and challenging task. Training models to reliably identify manipulated content requires sophisticated techniques. In this work, we propose a DeepFake detection approach that leverages both Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs). The CNN is used to extract spatial features from individual video frames, which are then passed to an RNN that captures temporal dependencies and inconsistencies across frames caused by DeepFake generation tools.

Our system is evaluated on a large, publicly available dataset of manipulated videos. The experimental results demonstrate that our approach, despite using a relatively simple architecture, achieves competitive performance in detecting DeepFakes.

Keywords: Deepfake Video Detection, Convolutional Neural Network (CNN), Recurrent Neural Network (RNN).

I. INTRODUCTION

With the rapid advancement of smartphone camera technology and widespread access to high-speed internet, the creation and sharing of digital videos has become easier than ever. This, combined with the rise of social media platforms, has significantly amplified the spread of video content across the globe. At the same time, increasing computational power has fueled the progress of deep learning, enabling applications that were once considered impossible. However, like many emerging technologies, these advancements bring new challenges—one of the most pressing being the rise of "DeepFakes."

DeepFakes are synthetic videos generated using deep learning techniques, particularly Generative Adversarial Networks (GANs), which can manipulate both video and audio to produce highly realistic but fake media. The ease with which DeepFakes can now be created and shared has led to a surge in misinformation, harassment, and deception on digital platforms. This makes the detection of DeepFakes critical to preserving the integrity of information online.

To address this issue, we propose a novel deep learning-based approach for detecting AI-generated fake videos. Identifying DeepFakes requires a thorough understanding of how GANs generate these videos. Typically, GANs take a video and an image of a 'target' person and produce a new video where the target's face is replaced with that of a 'source' individual. This process is powered by deep neural networks trained on facial images and target videos to accurately map the expressions and movements of the source onto the target. The video is processed frame by frame, and the modified frames are then stitched back together, often using autoencoders to achieve a high degree of realism.

However, DeepFake generation is not without limitations. Due to constraints in computational resources and time, the GAN-generated face images are usually of a fixed resolution and must be warped to align with the target face's configuration. This affine transformation introduces noticeable artifacts—specifically, inconsistencies between the resolution of the synthesized face area and the surrounding regions.

Our detection method is designed to exploit these artifacts. We analyze each frame of the video, using a ResNeXt-based Convolutional Neural Network (CNN) to extract visual features, focusing on the contrast between the manipulated face and the background. To further enhance accuracy, we employ a Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) units to identify temporal inconsistencies between frames—irregularities often introduced during DeepFake video reconstruction.

To train the ResNeXt model, we simulate the resolution mismatches typically found in DeepFakes by applying affine transformations to facial images, helping the model learn to recognize these subtle discrepancies. This approach allows our system to effectively distinguish between real and AI-generated content, providing a promising solution for combating the spread of DeepFakes online.

II. LITERATURE SURVEY

The rapid rise in the creation and misuse of DeepFake videos poses a significant threat to democracy, justice, and public trust. This growing concern has led to an increasing demand for advanced methods in fake video analysis, detection, and intervention. Several existing approaches in DeepFake detection are highlighted below:

1. Face Warping Artifact Detection

The method presented in "Exposing DeepFake Videos by Detecting Face Warping Artifacts" \[1] focuses on identifying inconsistencies by analyzing differences between synthetically generated facial areas and their surrounding regions using a dedicated Convolutional Neural Network (CNN). Their approach is grounded in the observation that current DeepFake generation techniques typically produce facial images at limited resolutions, which must then be scaled and adjusted to fit the source video, often introducing detectable artifacts.

2. Eye Blinking Analysis

The work titled "Exposing AI-Created Fake Videos by Detecting Eye Blinking" \[2] introduces a novel approach to identify fake videos based on the absence of natural eye blinking—a physiological cue often missing in AI-generated videos. The method is validated using eye-blinking detection benchmark datasets and has shown promising results. However, relying solely on blinking may not be sufficient. Additional facial attributes like teeth detail, skin texture, and wrinkle patterns are crucial, and our proposed method incorporates such features for a more comprehensive analysis.

3. Capsule Network-Based Detection

In "Using Capsule Networks to Detect Forged Images and Videos" \[3], the authors employ capsule networks to distinguish manipulated media across various scenarios, including replay attacks and AI-generated videos. Although their model performs well on controlled datasets, the use of random noise during training could hinder its effectiveness on real-world data. Our proposed method addresses this limitation by training on clean, real-time datasets for improved generalization.

4. Biological Signal-Based Detection

The study "Detection of Synthetic Portrait Videos Using Biological Signals" \[5] explores the use of biological signals—such as those captured from facial blood flow patterns (e.g., via PPG signals)—to detect fake videos. The approach involves extracting features related to spatial and temporal consistency from real and synthetic video pairs. These features are then used to train a CNN and probabilistic SVM for classification. The "FakeCatcher" system from this research achieves high detection accuracy across various conditions, independent of the video’s source, content, or resolution. However, the absence of a discriminator component and the complexity in designing a differentiable loss function to preserve biological signals pose limitations.

III. PROPOSED SYSTEM

While numerous tools exist for creating DeepFakes (DF), there is a significant lack of reliable and accessible solutions for detecting them. Our proposed approach aims to fill this gap by offering an effective method for identifying DeepFake content, thereby helping to prevent its spread across the internet. A key component of our solution is a user-friendly web-based platform where users can upload videos to determine whether they are authentic or manipulated.

This project has strong potential for scalability—from a basic web application to a browser extension that can automatically detect DeepFakes in real time. Furthermore, popular messaging and social media platforms such as WhatsApp and Facebook could integrate this solution to enable automatic detection of DeepFake content before it is shared with others.

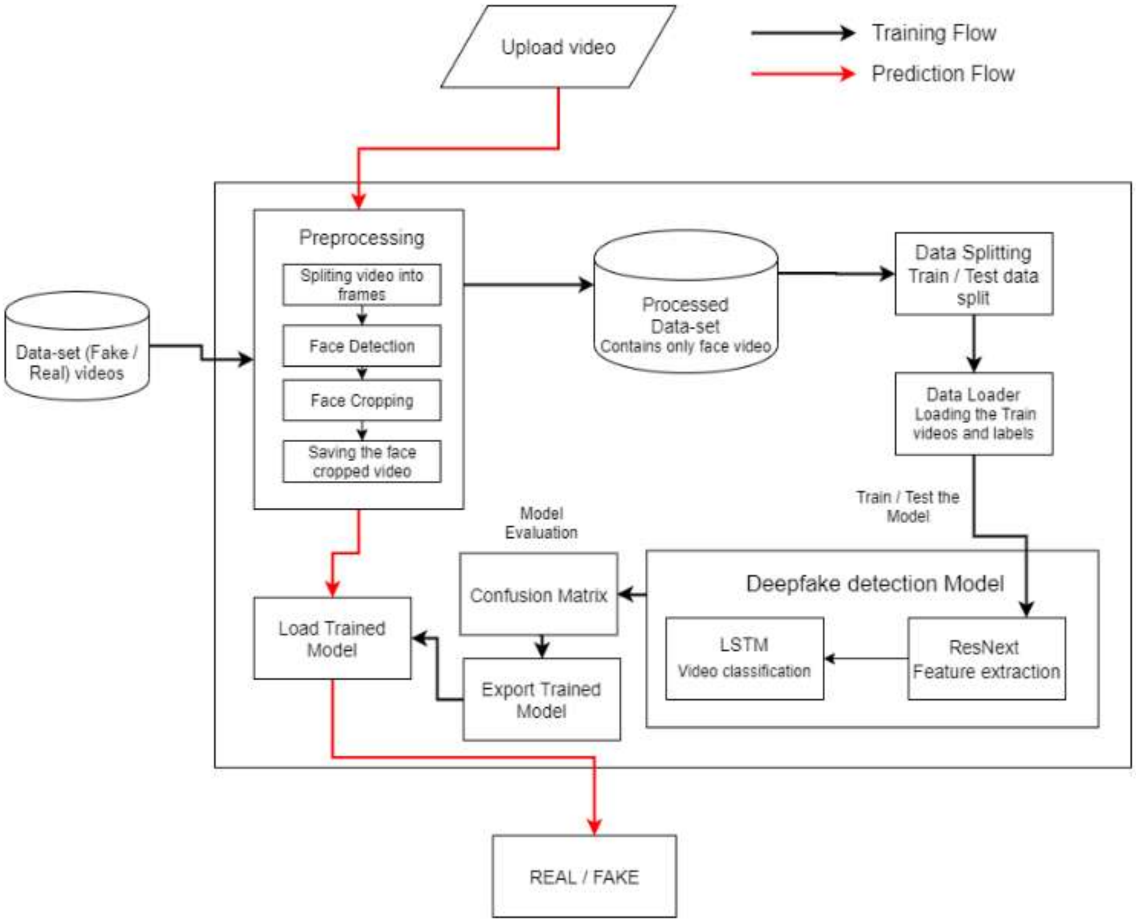

Upload video
Training Flow
Prediction Flow
Data-set (Fake / Real) videos
Preprocessing
Spliting video into frames
Face Detection
Face Cropping
Saving the face cropped video
Processed Data-set Contains only face video
Data Splitting Train / Test data split
Data Loader Loading the Train videos and labels
Train / Test the Model
Model Evaluation
Confusion Matrix
Deepfake detection Model
LSTM Video classification
ResNext Feature extraction
Load Trained Model
Export Trained Model
REAL / FAKE

Fig. 1: System Architecture

A. Dataset:

We utilize a combined dataset that includes an equal distribution of videos sourced from various publicly available datasets such as YouTube, FaceForensics++ \[14], and the DeepFake Detection Challenge dataset \[13]. The curated dataset is composed of 50% authentic (real) videos and 50% manipulated (DeepFake) videos. For model training and evaluation, the dataset is divided into two subsets: 70% for training and 30% for testing.

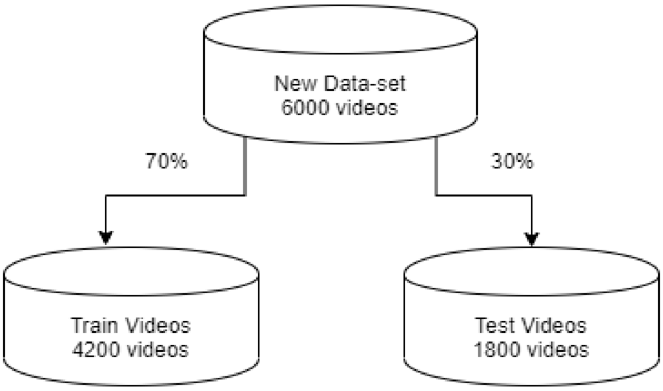


Fig. 2: Dataset

B. Preprocessing:

The dataset preprocessing involves several steps, starting with splitting each video into individual frames. Next, face detection is performed on these frames, and only the regions containing faces are cropped and retained for further processing. To ensure consistency in the number of frames across all videos, the mean number of frames per video in the dataset is calculated. A new, standardized dataset is then generated, where each video sample contains a number of face-cropped frames equal to this mean. Any frames that do not contain a detectable face are excluded during preprocessing.

Processing an entire 10-second video at 30 frames per second results in approximately 300 frames, which can be computationally intensive. To manage resource constraints during experimentation, we limit the input to the first 100 frames per video for training the model.

C. Model:

The proposed model architecture comprises a ResNeXt-50 (32x4d) network followed by a single Long Short-Term Memory (LSTM) layer. The data pipeline begins with a Data Loader that loads the preprocessed, face-cropped video samples and splits them into training and testing sets. During model training and evaluation, frames from these videos are processed in mini-batches and passed through the model to learn spatial and temporal features for DeepFake detection.

D. ResNext CNN for Feature Extraction:

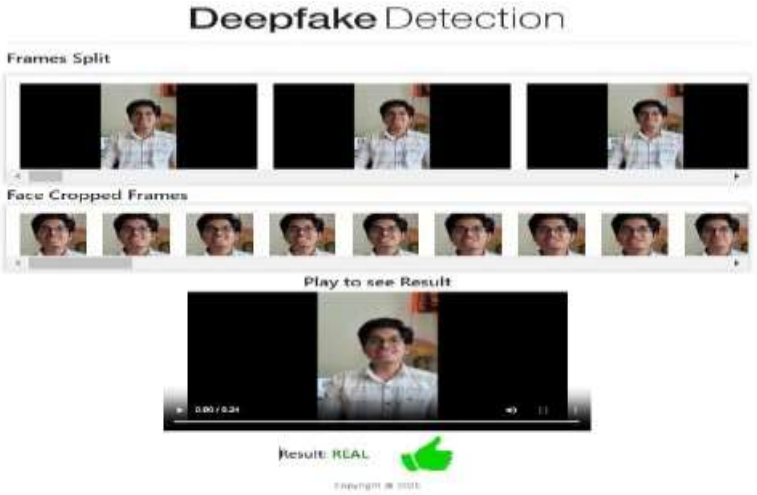
Rather than developing a classifier from scratch, we propose utilizing the ResNeXt CNN architecture for effective feature extraction and accurate detection of frame-level characteristics. To enhance performance, the network will be fine-tuned by adding necessary layers and selecting an appropriate learning rate to ensure stable and efficient convergence during gradient descent. The output from the final pooling layer of ResNeXt, which consists of 2048-dimensional feature vectors, serves as the input sequence for the subsequent LSTM layer.

E. LSTM for Sequence Processing:

Consider a sequence of feature vectors extracted from video frames using the ResNeXt CNN. These vectors are fed into a neural network with two output nodes representing the probabilities of the sequence belonging to either a DeepFake video or an authentic (untampered) video. A central challenge in this approach is designing a model that can effectively interpret the temporal sequence of frames. To address this, we propose using an LSTM layer with 2048 units and a dropout rate of 0.4. This configuration enables the model to perform sequential analysis, capturing temporal dependencies across frames. The LSTM processes each frame in order, allowing the model to compare the state of a frame at time `t` with one at time `t-n`, where `n` refers to the number of frames prior to `t`, thus enabling meaningful temporal context analysis.

IV. RESULT

The model's output will indicate whether the video is a DeepFake or authentic, along with the associated confidence level of the prediction.



Result: REAL

V. LIMITATIONS

Our method currently does not account for audio, meaning it is unable to detect audio-based DeepFakes. However, we plan to incorporate audio DeepFake detection in future iterations of the system.

VI. CONCLUSION

We have presented a neural network-based approach for classifying videos as either DeepFake or authentic, along with the confidence level of the model's prediction. The proposed method draws inspiration from the generation of DeepFakes using GANs and Autoencoders. Our approach employs ResNeXt CNN for frame-level feature extraction and utilizes an RNN with LSTM for video classification. The system is designed to accurately determine whether a video is a DeepFake or real, based on the parameters outlined in this paper. We believe that this approach will achieve high accuracy when applied to real-time data.

VII. REFERENCES

[1] Yuezun Li, Siwei Lyu. “Exposing DeepFake Videos by Detecting Face Warping Artifacts,” arXiv:1811.00656v3.

[2] Yuezun Li, Ming-Ching Chang, and Siwei Lyu. “Exposing AI-Generated Fake Videos by Detecting Eye Blinking,” arXiv.

[3] Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen. “Using Capsule Networks to Detect Forged Images and Videos.”

[4] Hyeongwoo Kim, Pablo Garrido, Ayush Tewari, and Weipeng Xu. “Deep Video Portraits,” arXiv:1901.02212v2.

[5] Umur Aybars Ciftci, İlke Demir, Lijun Yin. “Detection of Synthetic Portrait Videos Using Biological Signals,” arXiv:1901.02212v2.

[6] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. “Generative Adversarial Nets,” NIPS, 2014.

[7] David Güera, Edward J. Delp. “DeepFake Video Detection Using Recurrent Neural Networks,” AVSS, 2018.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. “Deep Residual Learning for Image Recognition,” CVPR, 2016.

[9] An Overview of ResNet and Its Variants: https://towardsdatascience.com/an-overview-of-resnetand-its-variants-5281e2f56035

[10] Long Short-Term Memory: From Zero to Hero with Pytorch: https://blog.floydhub.com/long-short-term-memory-from-zero-to-hero-with-pytorch/

[11] Sequence Models and LSTM Networks:
https://pytorch.org/tutorials/beginner/nlp/sequence_models_tutorial.html

[12] https://discuss.pytorch.org/t/confused-about-the-image-preprocessing-in-classification/3965

[13] https://www.kaggle.com/c/deepfake-detection-challenge/data

[14] https://github.com/ondyari/FaceForensics