International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 05 | May - 2024 SJIF Rating: 8.448 ISSN: 2582-3930

VisualTruth: Deepfake Video Detection System Using Deep Learning

1st Samruddhi S. Bagave
Computer Science and Technology
Usha Mittal Institute Of Technology
Mumbai, India
samruddhibagave@gmail.com

2nd Gauri M. Raut
Computer Science and Technology
Usha Mittal Institute Of Technology)
Mumbai, India
in.7gauri@gmail.com

3rd Surbhi D. Raut
Computer Science and Technology
Usha Mittal Institute Of Technology
Mumbai, India
surbhi0424@gmail.com

4rd Prof. Kumud Wasnik
HOD Of Computer Science and Technology
Usha Mittal Institute Of Technology
Mumbai, India
kumud.wasnik@umit.sndt.ac.in

Abstract*—In recent years, the proliferation of free deep learning-based software tools has made it remarkably straightforward to create highly convincing "DeepFake" (DF) videos, which often exhibit few traces of manipulation. While manipulations in digital videos have been possible for decades through visual effects, recent advances in deep learning have dramatically heightened the realism and accessibility of creating fake content, popularly referred to as AI-synthesized media or DF. The process of generating DF using AI tools has become relatively simple. However, detecting these deepfakes presents a significant challenge, as training algorithms to spot them is a complex endeavor. To address this challenge, we have developed an innovative approach to DF detection by harnessing the power of Convolutional Neural Networks (CNN) and Recurrent Neural Networks (RNN). Our system employs a CNN to extract frame-level features from videos, which are subsequently used to train an RNN. This RNN learns to classify whether a video has undergone manipulation and is capable of identifying the temporal inconsistencies introduced by DF creation tools. Our research is underscored by the evaluation of our system's performance against a substantial dataset of fake videos collected from standard sources,demonstrating competitive results achieved with a straightforward architecture. In the current era, characterized by the ubiquity of Artificial Intelligence (AI), the ease with which multimedia data can be manipulated and fabricated has raised profound concerns. The emergence of AI-generated deceptive content, commonly known as Deepfakes, has introduced complex challenges for the academic community. Existing methodologies for Deepfake identification have exhibited limitations in terms of generalization. This research is centered on addressing the pressing issue of Deepfake detection, with a focus on developing an effective mechanism that can operate robustly in real-world contexts.

Keywords - Deeplearning, Machine Learning, Deepfake Detection, LSTM, ResNext.

I. Introduction

The term "Deepfake" comes from the combination of "Deep Learning (DL)" and "Fake", describing highly realistic video or image content created using DL techniques. It came to prominence after an anonymous Reddit user used deep learning methods in late 2017 to create photorealistic fake videos by replacing faces in pornographic content with different faces. This marked a significant moment in the development of deepfake technology, showing its potential for both creative and malicious purposes.

Deepfake detection plays a key role in today's digital environment and acts as a critical defense against the growing threats posed by manipulated multimedia content. With rapid advances in artificial intelligence and machine learning, the ability to identify and combat sophisticated fake videos, images, and audio has become imperative.

The importance of deepfake detection goes beyond mere technological advancement. It includes various critical aspects such as maintaining factual accuracy, preventing the spread of disinformation, protecting individuals and organizations from malicious misuse, enhancing digital forensics capabilities, and ensuring national security and political integrity.

Moreover, the introduction of effective deep fake video detection mechanisms not only supports ethical media practices, but also encourages continued technological progress and research efforts in areas essential to today's digital society. These mechanisms not only protect against malicious activities, but also contribute to the development of robust and trustworthy digital ecosystems and promote trust and reliability in digital media and content platforms.

II. LITERATURE SURVEY:

TABLE I
SUMMARY OF THE PAPERS ANALYZED

Paper Details	Description	Observation
Paper Title:Deepfake Detection: A Systematic Literature Review Authors:MD Shohel Rana, Mohammad nur nobi, Beddhu Murali ,AND Andrew H. Sung Publication: IEEE Access, vol. 10, pp. 25494-25513, 2022	This study undertakes a comprehensive systematic literature review (SLR) encompassing 112 pertinent articles published between 2018 and 2020, delineating an array of methodological approaches. These approaches are categorized into four distinct groups: deep learning-based techniques, classical machine learning-based methods, statistical methodologies, and blockchain-based techniques. The study further appraises the detection efficacy of these diverse methods across disparate datasets, ultimately discerning that deep learning-based approaches exhibit superior performance in the realm of Deepfake detection.	Deep learning methods, particularly those utilizing diverse datasets containing both real and Deepfake images, show promise for effective Deepfake detection. However, they may face computational challenges and potential misclassification of real images with manipulations beyond facial regions, impacting their reliability in certain scenarios.
Paper Title:The Emergence of Deepfake Technology Authors:Mika Westerlund	The paper highlights the advantages and risks of deepfake technology. It can create realistic training videos or personalized marketing messages, but it also poses threats like fake news and reputation damage through impersonation. Challenges in combating deepfakes include their evolving nature, making detection tough, and their widespread online sharing, making origin tracing challenging..	The paper introduces a new approach to natural language generation. It recognizes the challenges in combating deepfakes and notes the proposed method's improvements but also its limitations compared to human-written text.
Paper Title: A Convolutional LSTM based Residual Network for Deepfake Video Detection Authors: Shahroz Tariq. Sangyup Lee, Simon S. Woo	The paper introduces CLRNet, a new method for deepfake detection using convolutional LSTM to analyze sequential images. Transfer learning enhances its adaptability, but potential effects of noise on its real-world performance are not addressed.	The paper emphasizes CLRNet's strengths in deepfake detection, such as superior generalization and robust performance. However, its computational complexity may limit deployment on resource-constrained devices, posing implementation challenges for less experienced deep learning practitioners.
Paper Title: The Face Deepfake Detection Challenge Authors:Luca Guarnera , Oliver Giudice, Francesco Guarnera , Alessandro Ortis , Giovanni Puglisi , Antonino Paratore , Linh M. Q. Bui , Marco Fontani , Davide Alessandro Coccomini, Roberto Caldelli , Fabrizio Falchi, Claudio Gennaro , Nicola Messina, Giuseppe Amato, Gianpaolo Perelli,	This paper talks about how to spot fake images and recreate the real ones in a challenge. They used computer tricks and smart math to do this. They found that the computer tricks worked really well, even when the pictures were messed up on purpose. They also came up with a new idea to bring back the real pictures from the fake ones, which no one had tried before. But, they couldn't make this new idea work as they expected.	The observation from the paper is that deep learning methods are highly effective for deepfake detection, particularly when applied to diverse datasets containing both real and deepfake images. However, there are challenges such as the computational cost of these methods and potential misclassification issues when relying on certain representations like Discrete Cosine Transform (DCT), which may lead to false positives or negatives in detection results.

III. PROPOSED SYSTEM

The study addresses the critical requirement to discern between deepfake and legitimate videos, using artificial intelligence to combat this deception. Unlike previous tools like FaceApp and Face Swap, which use pre-trained neural networks like GANs or Autoencoders, our method analyzes temporal sequencing using a Long Short-Term Memory (LSTM) neural network. Additionally, we use a pre-trained ResNext CNN to extract frame-level features, which improves the accuracy of our LSTM-based recurrent neural network for video categorization. While tools for creating deepfakes abound, resources for identifying them are limited. Our method attempts to bridge this gap by providing a user-friendly web platform for video authentication, with the potential to evolve into a browser plugin for automated deepfake detection. This project makes major contributions to reducing the spread of misleading.

A. Creation of Deepfake

Understanding the creation process of deepfake videos is crucial for their detection. Most tools, like GANs and autoencoders, take a source image and target video, replacing faces in each frame and enhancing video quality. We employ a similar approach for detection. While deepfakes appear realistic to the naked eye, they often leave subtle traces or artifacts. Our paper aims to identify these imperceptible cues to distinguish between deepfake and authentic videos.

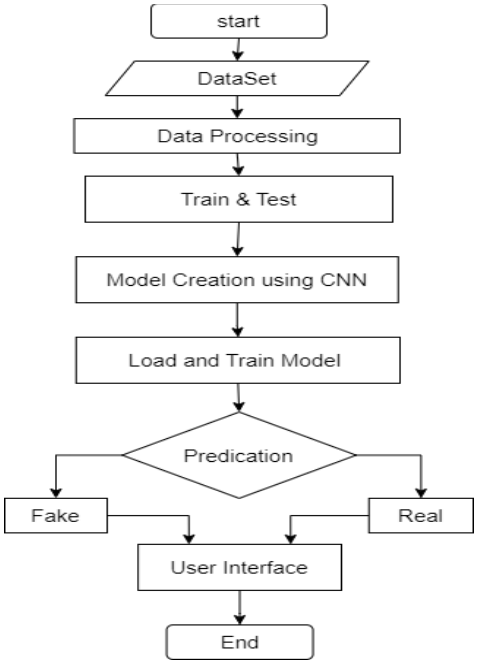
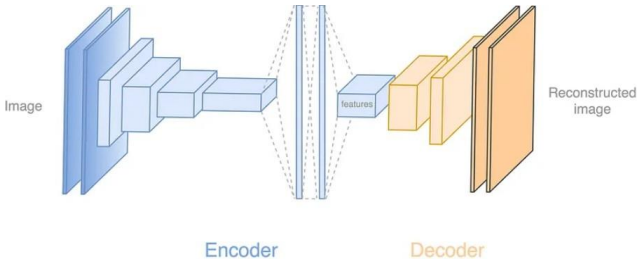


Fig. 1. Proposed System Flow



B. Data-set Gathering

- We collected data from diverse sources, such as YouTube and the Deepfake Detection Challenge (DFDC), to construct a comprehensive dataset of synthetic or manipulated videos. This extensive dataset enables precise and real-time detection across various video types, with an equal distribution of genuine and fake videos to mitigate training bias.
- After preprocessing our collected dataset to remove audio-altered videos, we obtained 1500 genuine and 1500 forged videos. Furthermore, we included 1000 genuine and 1000 fake videos from the Deepfake Detection Challenge (DFDC) dataset.

C. Pre-processing

- The videos undergo preprocessing to remove noise, focusing on detecting and cropping faces in each frame. This involves splitting the video into frames, identifying faces, and cropping them. The resulting frames containing faces are combined into a new video, creating a dataset of face-only videos, while frames without recognized faces are excluded.
- To ensure consistency and accommodate computational constraints, a threshold of 150 frames, based on the mean total frame count of each video, is set. This limitation allows for efficient processing within the GPU capabilities of the experimental environment, with only the first 150 frames of each video being saved to the new dataset.
- Frames for the new dataset are chosen sequentially, demonstrating the use of LSTM. The resulting videos are saved at 30 fps and a resolution of 112 × 112, suitable for analysis and model training.

D. Data-set split

- The dataset is divided into train (4,200) and test (1,800) videos in a 70:30 ratio. The train and test split is balanced, with 50% actual and 50% false videos.

E. Model Architecture

- Our model combines a pretrained ResNext CNN for feature extraction at the frame level with an LSTM network for video classification as either deepfake or pristine. The

ResNext model, specifically resnext50_32x4d , is used due to its high performance on deep neural networks. We fine-tune the ResNext by adding necessary layers and adjusting the learning rate for optimal convergence during training. The 2048-dimensional feature vectors from the ResNext's last pooling layers serve as input for the LSTM.

- The LSTM layer, comprising 2048 latent dimensions and 2048 hidden layers with a 0.4 dropout rate, processes the frames sequentially to analyze the video temporally. This allows for comparison between frames at different time intervals. The model incorporates a Leaky ReLU activation function, followed by a linear layer with 2048 input features and 2 output features for learning the correlation between input and output. An adaptive average pooling layer ensures consistent output image size, while a Sequential Layer facilitates sequential frame processing. Training is performed with a batch size of 4, and a SoftMax layer provides confidence scores during prediction.

F. Hyper-parameter tuning

To ensure optimum accuracy, numerous iterations are used while picking appropriate hyperparameters. For our dataset, we found that utilizing the Adam optimizer with a learning rate of 1e-5 (0.00001) and weight decay of 1e-3 produced the best results. This classification problem was solved using cross entropy loss, with batch training utilizing a batch size of four to ensure efficient computation in our setting. The User Interface was created using the Django framework, which ensures scalability. The index.html page allows visitors to upload videos, which are subsequently fed into the model for prediction. The model outputs whether the video is real or false, as well as the confidence level, which are presented in predict.html alongside the playing video.

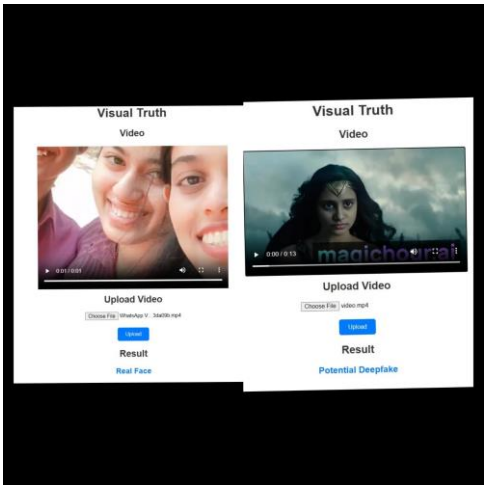
IV. Result and Analysis

In our project, we developed a model employing Resnext and LSTM networks. Additionally, we created a user interface enabling users to upload videos, with the system providing outputs indicating whether the video is Real or Potentially deepfake.

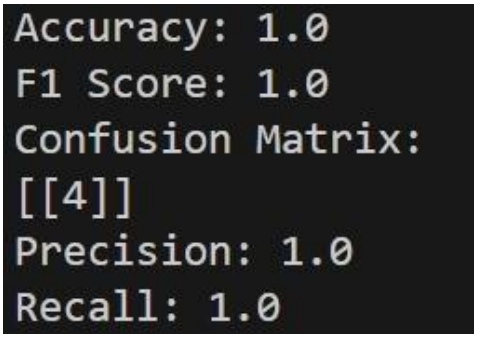
In the image below, we've uploaded a video on our project's user interface, which we created using an application. Upon analysis, the software flagged it as **'Potentially deepfake'**. Additionally, we tested it with a video featuring the well-known character Joey from Friends and received a classification of **'Real'**.

Conclusion and Future Scope

During the Data Exploration phase, we thoroughly analyzed both the training and testing datasets, examining metadata files and video content. This involved visualizing individual frames from real and fake videos and watching select videos in full for a deeper understanding.



Subsequently, we computed accuracy and F1 scores to evaluate the performance of our model.



In the Preprocessing phase, we utilized Python's 'glob' library to import videos efficiently and calculated a mean frame count of 150 frames as a threshold. We meticulously split the videos into frames, cropping each frame to isolate facial regions for focused analysis.

The cropped frames were combined to create new videos, standardized to 30 frames per second with a resolution of 112 x 112 pixels in mp4 format. This standardization ensured

consistency and compatibility across the dataset.

For optimal temporal analysis with the Long Short-Term Memory (LSTM) component, we included the first 150 frames from each video to capture essential temporal patterns, setting the stage for effective model development and training in our Deepfake Video Detection project.

The future scope of the project ”Deepfake Detection System using Deep Learning” with both Flask and Tkinter apps comprises various potential routes of development and enhancement:

- Improving Accuracy: Constantly improving the deep learning models used for deepfake detection to obtain greater accuracy in discriminating between authentic and altered videos.
- Real-time detection involves integrating real-time video analysis capabilities into the system to detect deepfake content as it is made or uploaded.
- Scalability refers to the system’s ability to process huge amounts of video data efficiently, providing smooth performance even as user traffic or dataset size increases.
- Expansion into Audio Deepfake Detection:Another possible future direction for the project is to expand the detection skills to include deepfake audio detection. This would entail creating algorithms and models to recognize modified or synthetic audio information to supplement the existing video-based deepfake detection method.
- Cloud Integration:connect our system with cloud services to store, process, and analyze video data. This will make it easier for multiple users to collaborate and use resources efficiently.
- Advanced Features: add more advanced tools like automatic training and updating of models, detecting unusual patterns, and preventing the creation of new deepfakes. This keeps us ahead of new tricks deepfake creators might use.
- Collaborative Research: We’ll work with other research groups to share ideas and findings about deepfake detection. This helps us stay updated with the latest developments and contribute to improving detection methods.

REFERENCES

[1] M. S. Rana, M. N. Nobi, B. Murali and A. H. Sung, ”Deepfake Detection: A Systematic Literature Review,” in IEEE Access, vol. 10, pp. 25494-25513, 2022.
[2] Westerlund, Mika. ”The emergence of deepfake technology: A review.” Technology innovation management review 9, no. 11 (2019).
[3] Tariq, S., S. Lee, and S. Woo. ”A convolutional lstm based residual network for deepfake video detection. arXiv 2020.” arXiv preprint arXiv:2009.07480.
[4] Nguyen, H. H., J. Yamagishi, and I. Echizen. ”Use of a capsule network to detect fake images and videos. arXiv 2019.” arXiv preprint arXiv:1910.12467.
[5] Guera, David, and Edward J. Delp. ”Deepfake video detection using recurrent neural networks.” In 2018 15th IEEE international conference on advanced video and signal based surveillance (AVSS), pp. 1-6. IEEE, 2018.
[6] Yogesh Patel 1, Sudeep Tanwar 1, (Senior Member, IEEE),Rajesh Gupta 1, (Member, IEEE), Pronaya Bhatttacharya 2, (Member, IEEE),INNOCENT EWEAN DAVIDSON 3,4 , (Senior Member, IEEE), Royi Nyameko 3,4,Srinivs Aluvala 5, AND Vrince Vimal 6,7, (Member, IEEE). ”Deepfake detection using deep learning methods: A systematic and comprehensive review.” In 20 November 2023.
[7] Rimsha Rafique, Rahma Gantassi, Rashid Amin, Jaroslav Frnda, Aida Mustapha & Asma Hassan Alshehri. ”Deep fake detection and classification using error-level analysis and deep learning.” 13, Article number: 7422 (2023).
[8] Thanh Thi Nguyena , Quoc Viet Hung Nguyenb , Dung Tien Nguyena, Duc Thanh Nguyena, Thien HuynhThec, Saeid Nahavandid, Thanh Tam Nguyene, Quoc-Viet Phamf, Cuong M. Nguyeng”Deep Learning for Deepfakes Creation and Detection: A Survey.”
[9] Preeti a, Manoj Kumar b, Hitesh Kumar Sharma c. ”A GAN-Based Model of Deepfake Detection in Social Media.” In 31 January 2023, Version of Record 31 January 2023.
[10] 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.
[11] Ramadhani, Kurniawan Nur, and Rinaldi Munir. ”A comparative study of deepfake video detection method.” In 2020 3rd International Conference on Information and Communications Technology (ICOIACT), pp. 394-399. IEEE, 2020.
[12] Yu, Peipeng, Zhihua Xia, Jianwei Fei, and Yujiang Lu. ”A survey on deepfake video detection.” Iet Biometrics 10, no. 6 (2021): 607-624.
[13] Kharbat, Faten F., Tarik Elamsy, Ahmed Mahmoud, and Rami Abdullah. ”Image feature detectors for deepfake video detection.” In 2019 IEEE/ACS 16th International Conference on Computer Systems and Applications (AICCSA), pp. 1-4. IEEE, 2019.
[14] Chadha, Anupama, Vaibhav Kumar, Sonu Kashyap, and Mayank Gupta. ”Deepfake: an overview.” In Proceedings of Second International Conference on Computing, Communications, and Cyber-Security: IC4S 2020, pp. 557-566. Springer Singapore, 2021.
[15] Hitawala, Saifuddin. ”Evaluating resnext model architecture for image classification.” arXiv preprint arXiv:1805.08700 (2018).
[16] https://www.kaggle.com/competitions/deepfake-detection-challenge/data