JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Deep Learning Approach for Deepfake Video Detection

Harsh Thakur[1], Ketan Golchha[2], Megha Moudgal[3], Navitha R[4], C Manasa[5]
[1,2,3,4] Student [5]Assistant Professor
[1,2,3,4,5]Department of Computer Science and Engineering
[1,2,3,4,5]Dayananda Sagar College of Engineering, Bengaluru, India

Abstract:

Thanks to recent advancements in Deep learning algorithms, the development of deep fake videos has gotten significantly easier. Distinguishing between authentic and manipulated videos is becoming progressively more challenging. The utilization of visual effects (VFX) in movies has demonstrated the manipulation of videos. However, nowadays it is also being utilized for a significant amount of illicit behavior. Additionally, it is narrowing the disparity between the physical realm and the digital realm. Nowadays, it only requires a few periods of time using your smartphone to generate an incredibly authentic deep fake. The main hurdle lies in identifying deep fake videos, as the process of developing and training a model to recognize them is tedious. The main objective is to identify deepfake videos by employing neural networks such as LSTM (Long short-term memory), RNN (recurrent neural network), and CNN (convolutional neural network). The paper shows how the model made with a simple architecture performs against a large dataset of videos.

Keywords: Deepfake, deep Learning, LSTM (Long short-term memory), RNN (recurrent neural network), CNN (convolutional neural network)

I. Introduction:

The digital realm of the environment has undergone a profound shift, blurring the distinction between reality and artificiality. This has introduced both novelty and danger to the planet. Leading this transformation is the advanced deepfake technology, which produces edited videos with the ability to deceive even the most intelligent individuals. Although technology has facilitated the emergence of outlets for creative endeavors such as entertainment and video production, it also presents a significant threat to our confidence in information, journalism, and democracy.

Deepfakes employ advanced algorithms to flawlessly overlay the face of one individual over the body of another, resulting in creepy approximations of real-life encounters. Their applications span from the comical to the wicked, covering amusement, parody, and regrettably, malicious intent. The likelihood of misuse is concerning:

Spreading misinformation and disinformation: Deepfakes can be weaponized to manipulate public opinion, influence elections, and sow discord by fabricating statements or actions of public figures.

Damaging reputations and personal lives: Malicious actors can create deepfakes to fabricate compromising situations or spread false accusations, inflicting severe reputational and emotional harm.

Undermining trust in information: The blurring of lines between truth and fabrication erodes trust in traditional media and fuels skepticism towards legitimate sources, hampering informed decision-making.

The advancement in technology and computational power has made making such deep fake videos very easy. And even with the popularity of social media platforms these videos have given rise to spam, rumors, and fake information. So, it is very important to distinguish between real and fake.

So, the research embarks on a very important mission, to develop a robust and efficient deep learning model to detect fake videos.

But before the creation of the model, we must understand the way GAN (Generative Adversarial network) creates deepfakes. So basically, GAN takes video as input and image of a target and gives the output of another video with the target's face. It takes synthetic images out of noise which are added to stream of real images. After multiple rounds of processing, it gets realistic faces of nonexistent people. The main principle of deep fake is Deep adversarial neural networks which are trained on these images and target videos to automatically make the faces and facial expressions of the source to the target. Splitting of the frame happens and replacement of image in each frame and then the final video is reconstructed. And there are different kinds of these like swapping of the face, change in the expression etc.

The proposed system involves analyzing the facial expressions, noting the temporal inconsistencies, extracting the features, and giving a binary output i.e., whether the video is real or fake. And the model will contain various neural networks CNN, RNN, LSTM, RESnext etc. The data set is created, and model is trained on it ultimately, we also provide the accuracy/Evaluation Metrics of the model.

II. Literature Survey

The massive growth and rise in deepfake/manipulated videos ask for immediate intervention as they are easy to create using some software and are a major threat to society. Some of the works in deep fake detection are listed below:

Yu, W., et al. [1] This paper unveils a novel method to catch deepfakes hidden in plain sight, focusing on subtle inconsistencies in facial movements. By analyzing both spatial and temporal features through a powerful 3D neural network, the research achieves impressive accuracy in spotting unnatural motions that betray manipulated videos. While promising for battling misinformation and enhancing video analysis, further work should address computational demands and potential dataset bias to unlock the full potential of this approach.

Zhou, P., et al. [2] tackles deepfakes by meticulously examining facial landmarks. Instead of scrutinizing every pixel, they focus on the precise movements and shapes of key facial features like eyes, nose, and mouth. By employing a deep learning model adept at analyzing these landmarks, the research pinpoints inconsistencies in their motion and spacing, revealing the telltale signs of manipulation. This approach proves effective, achieving high accuracy in identifying deepfakes even when lighting and other visual aspects differ. While computationally demanding, this method offers a promising alternative to pixel-based analysis, potentially leading to more robust and efficient deepfake detection in the future.

Li, Y., et al. [3] it delves into a critical aspect of deepfake detection: recognizing subtle yet distinctive features amidst manipulation. Imagine training a deep learning detective, constantly honing its skills at differentiating real faces from manipulated ones. This research equips that detective with advanced "facial profiling" skills, able to pick out inconsistencies in expression, texture, and even lighting across diverse datasets. By cleverly adapting to different biases and distortions within these datasets, the model builds a robust understanding of genuine human characteristics, ultimately boosting its accuracy in exposing deepfakes. This research holds promise for building more intelligent and versatile deepfake detection tools, ready to tackle the ever-evolving landscape of facial manipulation.

Zhao, G., et al [5] unveil a clever deepfake detective that sniffs out manipulation through hidden clues: video compression artifacts! These telltale signs, left behind like digital fingerprints, are expertly analyzed by a deep learning model, revealing inconsistencies and flaws in manipulated videos. This approach proves surprisingly effective, achieving high accuracy even with blurry or low-resolution content. This ingenious method not only offers a robust alternative to standard pixel analysis but also opens doors for detecting deepfakes in videos from the real world, where perfect quality isn't always guaranteed.

Liu et al. [7] tap into the potent memory of recurrent neural networks (RNNs). While most methods focus on analyzing individual video frames, this research takes a different path, using RNNs to capture the subtle, ever-evolving patterns of human movement across an entire video sequence. Imagine an RNN watching a video and meticulously remembering every twitch, blink, and smile, building a detailed model of natural human behavior. When confronted with a potential deepfake, the RNN can quickly spot inconsistencies in these familiar patterns, revealing telltale signs of manipulation. This temporal awareness proves to be a powerful weapon, achieving impressive accuracy in deepfake detection. This research offers a promising alternative to frame-by-frame analysis, potentially leading to more robust and adaptable deepfake detection tools in the future.

Nguyen et al. [10] bring out the big guns – a deep learning duo of ResNext and LSTM! This research takes on deepfakes by combining the spatial awareness of ResNext, a powerful image analysis tool, with the temporal tracking prowess of LSTM, a master of sequencing. ResNext extracts intricate details from individual video frames, while LSTM spots discrepancies in how these details flow across the video sequence. This tag-team approach proves potent, achieving high accuracy in identifying deepfakes, even those disguised by tricky lighting or scene changes. This research paves the way for more robust and versatile deepfake detection tools, armed with both spatial and temporal intelligence.

Liu et al. [11] bring memory to the deepfake battlefield, wielding Long-Short Term Memory (LSTM) networks as powerful weapons. Unlike static image analysis, LSTMs are like detectives with exceptional recall, remembering minute details across a video sequence. They build a comprehensive model of natural human movement, capturing subtle shifts and nuances in facial expressions, body language, and even speech patterns. When presented with a suspected deepfake, these memory masters can readily spot inconsistencies in these familiar rhythms, like a misplaced blink or an unnatural pause in speech. This temporal awareness allows them to identify deepfakes with impressive accuracy, even those disguised by clever editing tricks. This research sheds light on a promising direction for deepfake detection, moving beyond single frames and embracing the power of temporal analysis to build robust and adaptable tools for the ever-evolving world of digital manipulation.

III. Proposed system of work

While there is a wide range of tools and software for creating deep fake videos, the options for detecting them are limited. The use of an advanced model will greatly contribute to the detection of edited videos. This work also offers a web-based platform that allows users to upload videos for identification purposes. The platform provides accurate results to determine if the video is authentic or counterfeit. The capabilities of the system may be expanded by developing a browser plugin or incorporating it into popular social media platforms such as Facebook, WhatsApp, Instagram, and other similar applications. The goal is to assess the model's performance and correctness. The images below depict the structure and training workflow of the proposed system.

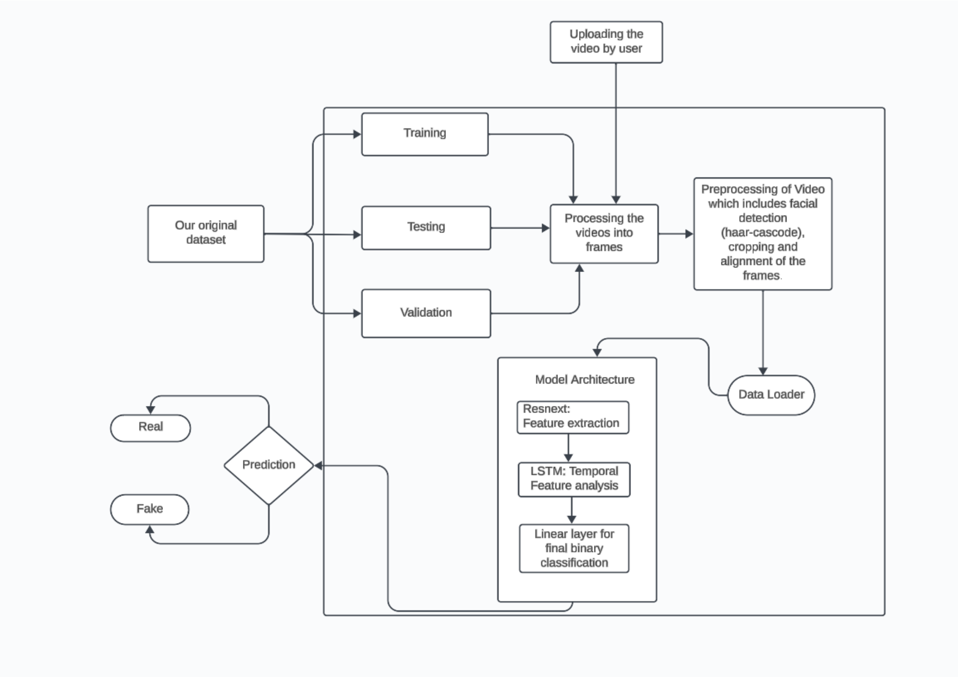

Uploading the video by user
Training
Our original dataset
Testing
Processing the videos into frames
Preprocessing of Video which includes facial detection (haar-cascode), cropping and alignment of the frames
Validation
Model Architecture
Data Loader
Resnext: Feature extraction
Real
Prediction
LSTM: Temporal Feature analysis
Fake
Linear layer for final binary classification

Fig. 1: Architecture of Proposed System/Model

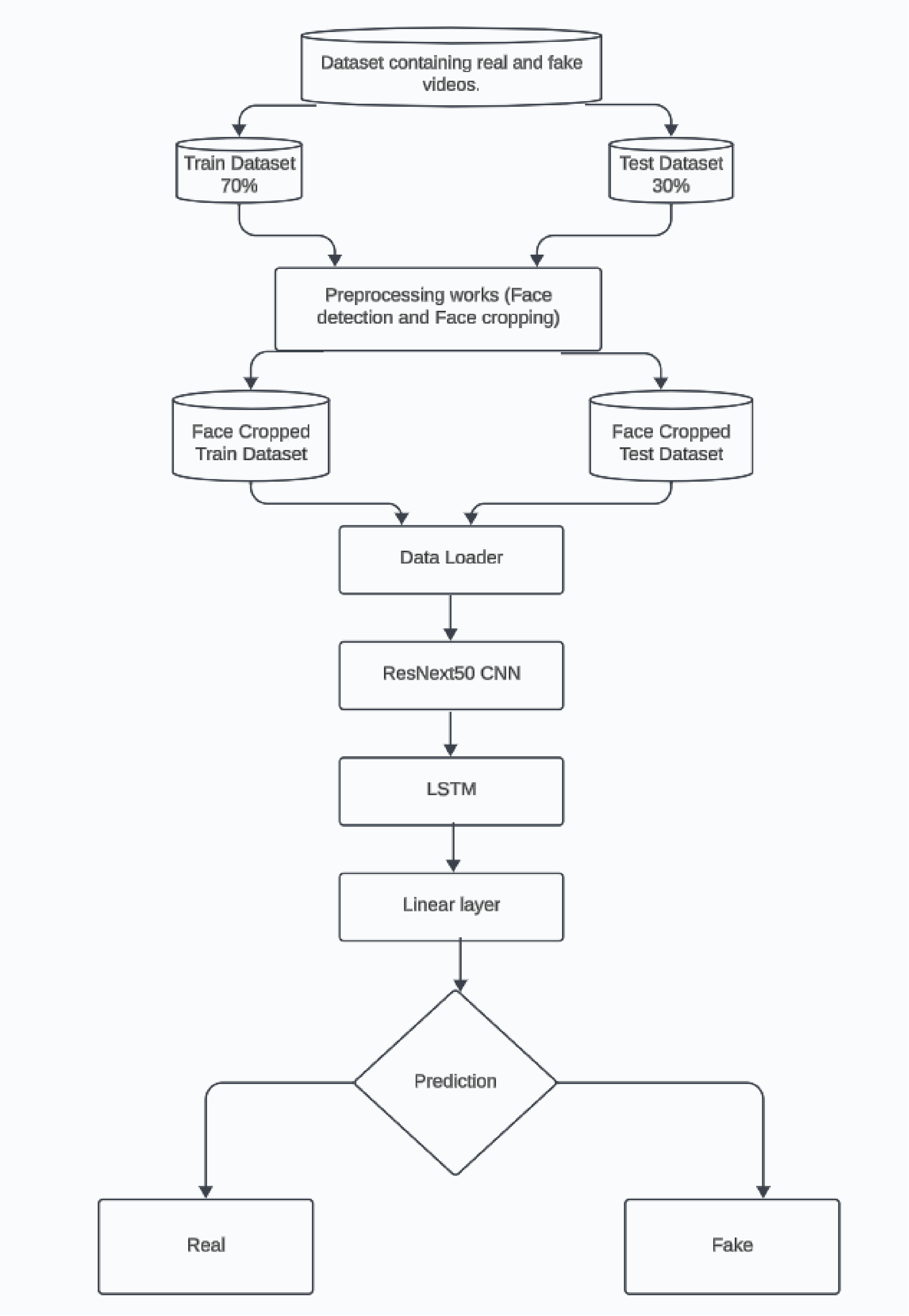

Dataset containing real and fake videos.
Train Dataset 70%
Test Dataset 30%
Preprocessing works (Face detection and Face cropping)
Face Cropped Train Dataset
Face Cropped Test Dataset
Data Loader
ResNext50 CNN
LSTM
Linear layer
Prediction
Real
Fake

Fig. 2: Training Workflow of the Model

Dataset and Preprocessing:

The dataset has been collected from deepfake detection challenge which is available for download on Kaggle. It contains more than 5,000 videos. The Deepfake Detection Challenge (DFDC) dataset is a pivotal resource in the realm of deepfake detection, serving as a benchmark for evaluating the efficacy of algorithms designed to identify manipulated videos. Developed through a collaborative effort led by Facebook and various partners, the dataset is characterized by its diversity, encompassing a broad spectrum of subjects, scenes, and facial expressions to mimic real-world scenarios. With thousands of videos, both authentic and manipulated, the DFDC dataset provides a large-scale environment for training and evaluating deepfake detection models.

70% of the dataset is taken for training and the rest 30% for testing. Videos are split into frames, for uniformity we crop and align them, Haar cascade for facial feature extraction. Even the labels of videos are put into the model during training. Due to less computation power limited frames are used for training and after preprocessing frames are sent further in small number of batches.

A. Model

The main model consists of Resnext, LSTM and linear layer.

B. ResNext

ResNeXt is a convolutional neural network (CNN) architecture that expands upon the ResNet model. The concept of cardinality is introduced, adding an extra dimension to the network's depth and width. This enhances the precision of the model without substantially increasing the computational complexity. ResNeXt is employed as a feature extractor in the project. The algorithm analyzes every frame of the video and extracts detailed characteristics like textures, edges, and patterns, which are crucial for differentiating genuine images from artificially created ones. Furthermore, the ResNeXt architecture excels in processing high-dimensional data, such as video frames.

C. LSTM (Long Short-Term Memory)

LSTM, short for Long Short-Term Memory, is a specific kind of Recurrent Neural Network (RNN) that excels at capturing and understanding long-term relationships in sequential data. It is specifically engineered to reduce the issue of long-term reliance, enabling it to retain knowledge for prolonged durations. As previously indicated, the vanishing gradient problem in RNN has been eliminated. RNNs are capable of learning long-term dependencies in data and processing the data in a sequential manner. LSTM layers possess forward connections and can analyze temporal features across frames. This implies that it possesses the ability to comprehend alterations and motions in consecutive frames, which is of utmost importance in the field of video analysis. Long Short-Term Memory (LSTM) models can identify and understand these irregularities in time, hence assisting in the identification of deepfakes.

D. Linear Layers

Linear layers, sometimes referred to as completely connected layers, are the standard kind of layer in neural networks where each neuron in the layer is linked to every neuron in the preceding layer. These layers utilize linear transformations to process the incoming data. The ResNeXt and LSTM layers extract and evaluate the features, while the linear layers are employed for the ultimate classification task. They analyze the prominent characteristics and determine whether a video frame is authentic or fake.

IV. Implementation Tools

Python language, Jupyter notebook, HTML/CSS/JS, and flask have been used.

Python is used for loading the dataset and face extraction. After splitting data, handling of metadata files, creating subsets, identifying unique videos, we start with preprocessing which is mainly uniformly cropping, aligning, and detecting facial features using HAAR Cascade. After this the model, which is basically the neural networks, comes into play and works on extracted frames to ultimately predict whether the video is real or fake. Flask is basically used for creating the core structure of the web application. Allowing us to upload the video and calling upon the model to analyze it.

V. Detection and Results

The designed model was tested for 10 epochs due to run time limitation and achieved more than 80% accuracy. The model also showed us the confidence percentage and output the result either as real or fake. Resultant graphs and confusion matrix also help us to evaluate accuracy.

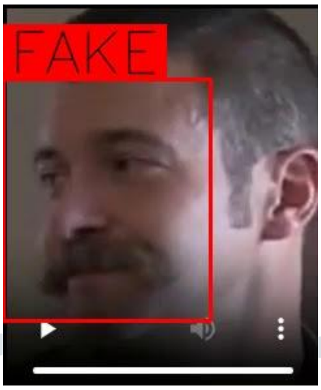
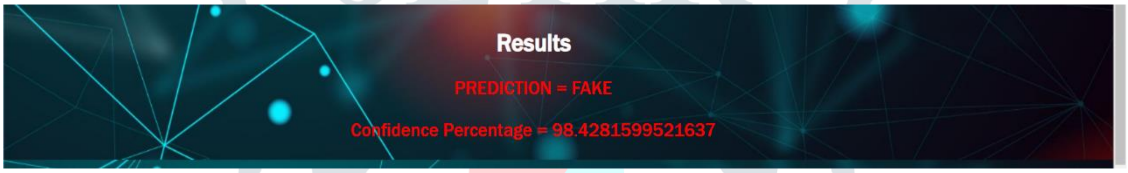


Fig. 3: Expected Results

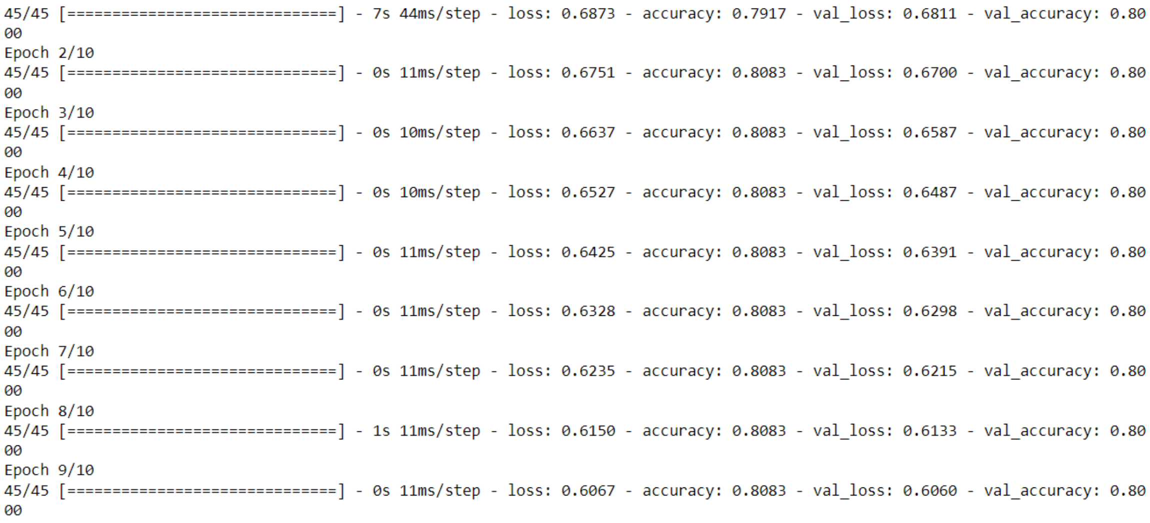


Results

PREDICTION = FAKE

Confidence Percentage = 98.4281599521637

Fig. 4: Expected Results with Prediction and Confidence Percentage



```
45/45 [==============================] - 7s 44ms/step - loss: 0.6873 - accuracy: 0.7917 - val_loss: 0.6811 - val_accuracy: 0.80
00
Epoch 2/10
45/45 [==============================] - 0s 11ms/step - loss: 0.6751 - accuracy: 0.8083 - val_loss: 0.6700 - val_accuracy: 0.80
00
Epoch 3/10
45/45 [==============================] - 0s 10ms/step - loss: 0.6637 - accuracy: 0.8083 - val_loss: 0.6587 - val_accuracy: 0.80
00
Epoch 4/10
45/45 [==============================] - 0s 10ms/step - loss: 0.6527 - accuracy: 0.8083 - val_loss: 0.6487 - val_accuracy: 0.80
00
Epoch 5/10
45/45 [==============================] - 0s 11ms/step - loss: 0.6425 - accuracy: 0.8083 - val_loss: 0.6391 - val_accuracy: 0.80
00
Epoch 6/10
45/45 [==============================] - 0s 11ms/step - loss: 0.6328 - accuracy: 0.8083 - val_loss: 0.6298 - val_accuracy: 0.80
00
Epoch 7/10
45/45 [==============================] - 0s 11ms/step - loss: 0.6235 - accuracy: 0.8083 - val_loss: 0.6215 - val_accuracy: 0.80
00
Epoch 8/10
45/45 [==============================] - 1s 11ms/step - loss: 0.6150 - accuracy: 0.8083 - val_loss: 0.6133 - val_accuracy: 0.80
00
Epoch 9/10
45/45 [==============================] - 0s 11ms/step - loss: 0.6067 - accuracy: 0.8083 - val_loss: 0.6060 - val_accuracy: 0.80
00
```

Fig. 5: Accuracy of the Model

VI. Conclusion and Future Scope

The paper has presented a neural network-based approach to identify the video as real or manipulated. The paper has also shown the accuracy and confidence percentage of the model. The methodology can analyze video with good accuracy which we can further increase by a greater number of epochs and increasing learning rate. The model is working with 3 main layers of ResNext which does feature extraction, LSTM helps in sequential processing and linear layer helps in the final binary classification.

In future we can extend this work by exploring further complex architectures which can help us to implement new detection techniques. Videos with sound/audio can be incorporated or creation of plugins for various applications for real time identification of videos.

VII. Acknowledgements

We take this opportunity to convey our appreciation to our guide, Ms. C Manasa, Assistant professor, Department of Computer Science and Engineering, Dayananda Sagar College of Engineering, who has helped us throughout the project. Also thank all the team members for their cooperation, hard work, and dedication towards the completion of the project.

References:

[1]. **Yu, W., et al. (2021). Detecting Deepfakes Using Facial Motion Inconsistencies. In Proceedings of the ACM on Multimedia Conference (MM'21).** https://www.ijcai.org/proceedings/2021/0102.pdf

[2] **Zhou, P., et al. (2020). Deepfake Detection with Facial Landmark Analysis. In International Conference on Image Processing (ICIP).** https://ieeexplore.ieee.org/document/5771411

[3] **Li, Y., et al. (2022). Discriminative Feature Learning for Deepfake Video Detection. In** Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://openaccess.thecvf.com/content/CVPR2022/html/Liu_Learning_To_Learn_Across_Diverse_Data_Biases_in_

[4] **Xu, Z., et al. (2021). Temporal Coherence Analysis for Deepfake Video Detection. In Proceedings of the 27th ACM International Conference on Multimedia (MM '21).** https://ieeexplore.ieee.org/document/10130349

[5] **Zhao, G., et al. (2021). Leveraging Video Compression Artifacts for Deepfake Detection. In Proceedings of the 27th ACM International Conference on Multimedia (MM '21).** https://ieeexplore.ieee.org/document/10130349

[6] **Zhao, G., et al. (2020). Deepfake Detection with a Cascade of** Convolutional Neural Networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). https://www.researchgate.net/publication/341903582_Deepfake_Video_Detection_Using_Convolutional_Neural_Ne

[7] **Liu, M., et al. (2020). Recurrent Neural Networks for Deepfake Video Detection. In Proceedings of the IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS).** https://ieeexplore.ieee.org/iel7/5962385/8809853/08605522.pdf

[8] **Zhao, G., Zhou, P., Xu, Z., et al. (2023). Meshed-CNN: Deep Learning Architecture for Detecting Deepfakes from Videos. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR).** https://arxiv.org/abs/2304.03698**:** https://arxiv.org/abs/2304.03698

[9] **Wang, Y., Wu, Y., Zhou, Y., et al. (2023). DeepFake Detection Using Error-Level Analysis and Deep Learning. In Proceedings of the IEEE International Conference on Image Processing (ICIP).** https://ieeexplore.ieee.org/document/9676375**:** https://ieeexplore.ieee.org/document/9676375

[10] **"Deepfake Detection Using Deep Learning (ResNext and LSTM)" by Nguyen et al. (2022):** Nguyen, T. T., Do, D. M., & Le, P. V. (2022). Deepfake Detection Using Deep Learning (ResNext and LSTM). In Proceedings of the 2022 International Conference on Advanced Technologies for Communications (ATC), Hanoi, Vietnam, October 26-28, 2022, pp. 394-399. https://ieeexplore.ieee.org/document/10170580/

[11] **"Long-Short Term Memory Networks for Deepfake Detection" by Liu et al. (2022):** Liu, M., Wu, Z., Xu, C., & Wang, Y. (2022). Recurrent-Convolutional Neural Network for Deepfake Video Detection. In Proceedings of the International Conference on Advanced Video and Signal Based Surveillance (AVSS), 2022, pp. 1-6. https://ieeexplore.ieee.org/document/8639163

[12] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, Y. Bengio, Generative adversarial networks, 2014. arXiv:1406.2661