IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

Deep Fake Video Detection

Raj Parulekar[1], Kaushal Partole[2], Rahul Bhalerao[3], Om Budhawant[4], Dr. Nita Patil[5]
Students, Department of Information Technology[1,2,3,4]
Faculty, Department of Information Technology[5]
KC College of Engineering, Thane, India

Abstract: *Deep learning algorithms have become so strong due to increasing processing power that it is now quite easy to produce an identical human-synthesized video, sometimes known as a "deep fake.". Scenarios where these realistic face swapped deep fakes are used to create political distress, fake terrorism events, revenge porn, blackmail peoples are easily envisioned. In this work, we describe a deep learning-based method that can effectively distinguish AI-generated fake videos from real videos. Our technique can recognize the replacement and reenactment automatically. deepfakes. We are trying to use Artificial Intelligence(AI) to fight Artificial Intelligence(AI). Our system uses a Res-Next Convolution n e u r a l network to extract the frame-level features and these features and further used to train the Long Short Term Memory(LSTM) based Recurrent Neural Network(RNN to classify whether the video is subject to any kind of manipulation or not, ie whether the video is deep fake or real video. In order to improve the model's performance on real-time data and replicate real-world circumstances, we assess our approach using a sizable, well-balanced, and diverse set of prepared data by mixing the various available data-set like FaceForensic++, Deepfake detection challenge, and Celeb-DF. We also show how our system achieved competitive result using very simple and robust approach*

Keywords: Deepfake Video Detection , Convolutional Neural Network(CNN) , recurrent neural network (RNN) , Res- Next Convolution Neural Network , long short term memory(LSTM) , Generative Adversarial Network (GAN) , DF- Deepfake

I. INTRODUCTION

In the world of ever growing Social media platforms, Deepfakes are considered as the major threat of the Al. It becomes very important to spot the difference between the deepfake and Real video. We are using AI to fight AI Deepfakes are created using tools like Face-App and Face Swap, which using pre-trained neural networks like Generative Adversarial Network (GAN) or Auto encoders for these deepfakes creationOur approach processes the sequential temporal analysis of the video frames using an LSTM-based artificial neural network, then extracts the frame-level features using a pre-trained Res-Next CNN.. Res-Next Convolution neural network extracts the frame level characteristics, which are then utilized to train an artificial Recurrent Neural Network based on Long Short Term Memory to distinguish between Deepfake and actual videos. To emulate the real time scenarios and make the model perform better on real time dataTo enhance customer usability, we've created a front-end application where users can upload their videos, making them ready for immediate use.The video will be processed by the model and the output will be rendered back to the user with the classification of the video as deepfake or real. The rising sophistication of smartphone cameras and widespread access to high-quality internet connections globally have significantly expanded the reach of social media and media sharing platforms, making the creation and sharing of digital videos easier than ever. Advancements in computational power have also empowered deep learning techniques to an extent previously deemed impossible.

II. LITERATURE SURVEY

The rapid proliferation of deep fake videos and their illicit utilization poses a significant threat to democracy, justice, and public trust. Consequently, there has been a surge in the Dr Nita Patil dept of information technology Mumbai university Mumbai, India nita.patil@kccemsr.edu.in demand for analysis, detection, and intervention measures to address this issue. Some of the related word in deep fake detection are listed below:

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

Yuezun Li, Siwei Lyu et al. [1] employed an approach to detect artifacts by utilizing a dedicated Convolutional Neural Network model to compare the generated face areas with their surrounding regions. This method focused on identifying two types of Face Artifacts. Their methodology was founded on the observation that existing
DeepFake algorithms typically produce images of restricted resolutions, necessitating further transformations to align the replaced faces with those in the source video.

Yuezun Li et al. [2] describes a new method to expose fake face videos generated with deep neural network models. The method is based on detection of eye blinking in the videos, One physiological signal that is often inadequately represented in synthesized fake videos is the subtle variations in facial microexpressions, which convey genuine emotions and reactions. The method is evaluated over benchmarks of eye-blinking detection datasets and shows promising performance on detecting videos generated with Deep Neural Network based software DF. The approach relies on identifying eye blinking in the videos, a physiological signal that is often poorly replicated in synthesized fake videos.The method is evaluated over benchmarks of eye-blinking detection datasets and shows promising performance on detecting videos generated with Deep Neural Network based software DF. While their method primarily relies on the absence of blinking as a detection clue, it's essential to consider additional parameters for deep fake detection, such as teeth alignment, facial wrinkles, and other subtle facial features. Our method is proposed to consider all these parameters.

Huy H. Nguyen et al. [3] employs a capsule network approach to identify forged and manipulated images and videos across various scenarios, including replay attack and computer-generated video detection. However, their method utilizes random noise during the training phase, which may not be optimal. Despite this, the model demonstrated effectiveness on their dataset; however, it could potentially falter when faced with real-time data due to the inclusion of noise in training. In contrast, our proposed method aims to be trained on noiseless and real-time datasets, potentially enhancing its performance in practical applications.

Umur Aybars Ciftci et al. [4] approach extract biological signals from facial regions on authentic and fake portrait video pairs. Utilize transformations to calculate spatial coherence and temporal consistency, capturing signal characteristics within feature sets and PPG maps. Then, employ training of a probabilistic Support Vector Machine (SVM) and a Convolutional Neural Network (CNN) to further analyze and classify the data. Then, the aggregate authenticity probabilities to decide whether the video is fake or authentic. Each of these research summaries explores different methodologies for ECG arrhythmia classification, ranging from deep learning techniques like CNNs and RNNs to traditional methods like SVM and clustering algorithms, highlighting their performance, advantages, and limitations.

III. SYSTEM DESIGN

A. System Architecture

There are numerous tools available for creating deepfakes, but there's a noticeable scarcity in tools for detecting them.Our approach to detecting deepfakes could significantly stem their proliferation across the internet. We're developing a web-based platform allowing users to upload videos for classification as authentic or manipulated. This project has the potential to expand into a browser plugin for automatic deepfake detection. Major applications like WhatsApp and Facebook could integrate this technology for seamless pre- detection of deepfakes before sharing. A key objective is to assess its performance and acceptance, focusing on security, user-friendliness, accuracy, and reliability. Our method targets various types of deepfakes, including replacement, retrenchment, and interpersonal variations.

Dataset Gathering

For making the model efficient for real time prediction. We have gathered the data from different available. data-sets like Face Forensic ++(FF),Deepfake detection challenge(DFDC), and Celeb-DF. Moreover, we amalgamated the collected datasets to construct a new, comprehensive dataset. This amalgamation aims to facilitate precise and real-time detection across various types of videos. To avoid the training bias of the model we have considered 50% Real and 50% fake videos total.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

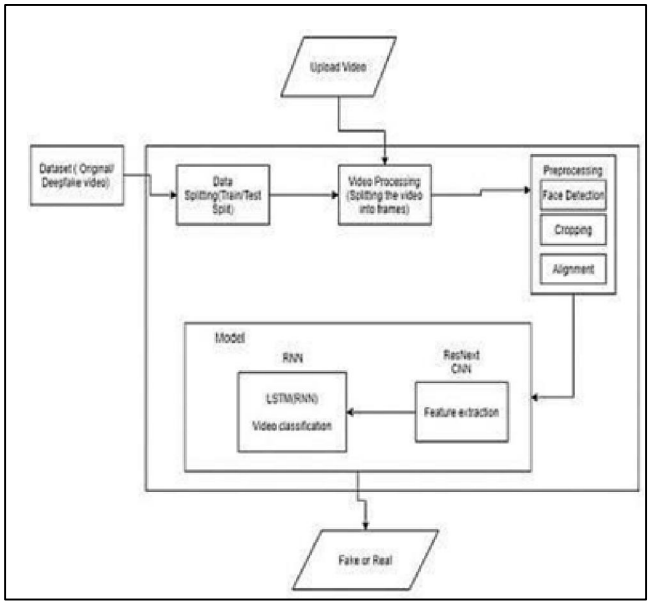

Fig 1 . system architecture

Pre-processing

The initial preprocessing step involves splitting the video into individual frames. Subsequently, facial detection is performed on each frame, and the frame is cropped to isolate the detected face. Following this, the cropped frames are reassembled into a new video by combining each processed frame sequentially.The face cropped frames are again written to new video using Video Writer. This process is repeated for each video, resulting in the creation of a processed dataset comprising videos containing only faces. To maintain the uniformity in the number of frames the mean of the dataset video is calculated and the A new processed dataset is generated, comprising cropped face frames that are standardized to a mean frame count. The frames that doesn't have faces in it are ignored during preprocessing. CV2 VideoCapture is used to read the videos and get the mean number of frames in each video. OpenCV an computer vision library is used during Preprocessing.

Model Details -

The model consists of following layers:

Res-Next CNN:

The Residual Convolutional Neural Network (CNN) pre- trained model is employed for the task.The model name is resnext50 32x4d0[22]. This model consists of 50 layers and 32 × 4 dimensions

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

stage	output	ResNeXt-50 (32×4d)
conv1	112×112	7×7, 64, stride 2
conv2	56×56	3×3 max pool, stride 2 [1×1, 128 3×3, 128, C=32 1×1, 256] ×3
conv3	28×28	[1×1, 256 3×3, 256, C=32 1×1, 512] ×4
conv4	14×14	[1×1, 512 3×3, 512, C=32 1×1, 1024] ×6
conv5	7×7	[1×1, 1024 3×3, 1024, C=32 1×1, 2048] ×3
	1×1	global average pool 1000-d fc, softmax
# params.		$\mathbf{25.0}\times10^6$

Fig No. 2 Res-Next Architecture

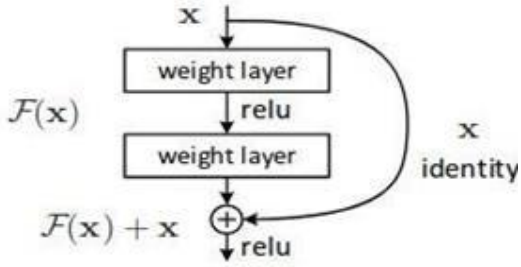

Fig No.3 Res-Next Working

Sequential Layer:

Sequential serves as a container for stacking modules together, enabling them to be executed sequentially.Sequential layer is used to store feature-vector returned by the Res-Next model in a ordered way. So that it can be passed to the LSTM sequentially.

LSTM Layer

LSTM is used for sequence processing and spot the temporal change between the frames.2048-dimensional feature vectors is fitted as the input to the LSTM. We are using 1LSTM layer with 2048 latent dimensions and 2048 hidden layers along with 0.4 chance of dropout, which is capable to do achieve our objective.

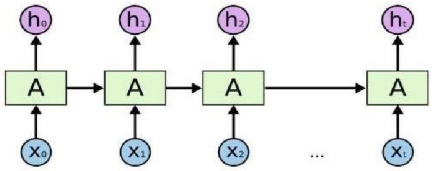

Fig No. 4 Overview of LSTM Architecture

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

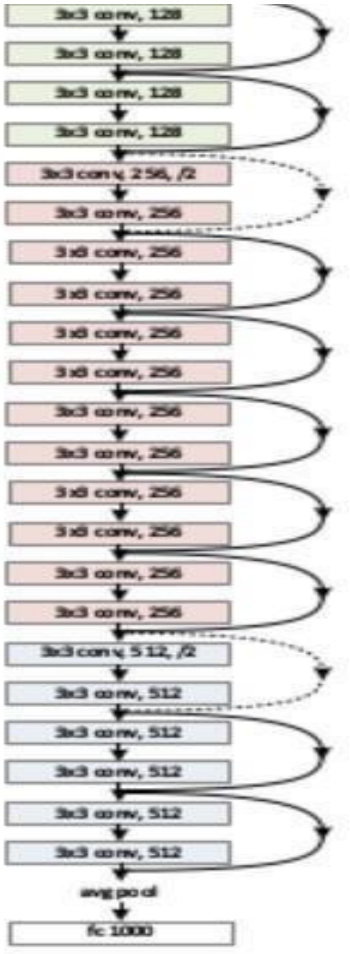

Fig .5 Overview Of Res-Next Architecture

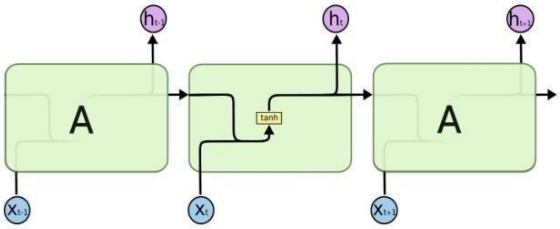

Fig No. 6 Internal LSTM Architecture

Model Prediction Details –

- The model is loaded in the application
- The new video for prediction is preprocessed and passed to the loaded model for prediction
- The trained model performs the prediction and return if the video is a real or fake along with the confidence of the prediction
- Adaptive Average Pooling Layer

It is used to reduce variance, reduce computation complexity and extract low level features from neighborhood. 2-dimensional Adaptive Average Pooling Layer is used in the model.

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

Model Training Details –

- Train Test Split: The dataset is split into train and test dataset with a ratio of 70% train videos and 30% test videos. The train and test split is a balanced split 1. 50% of the real and 50% of fake videos in each split.
- Data Loader: It is used to load the videos and their labels with a batch size of 4.

Export Model –

After the model is trained, we have exported the model. So that it can be used for prediction on real time data

IV. EVALUATION PARAMETERS

- Blinking Of Eyes
- Teeth Enchantment
- Bigger Distance For Eyes
- Moustaches
- Double Edges, Eyes, Ears, Nose
- Wrinkles On Face
- Inconsistent Head Pose
- Face Angle

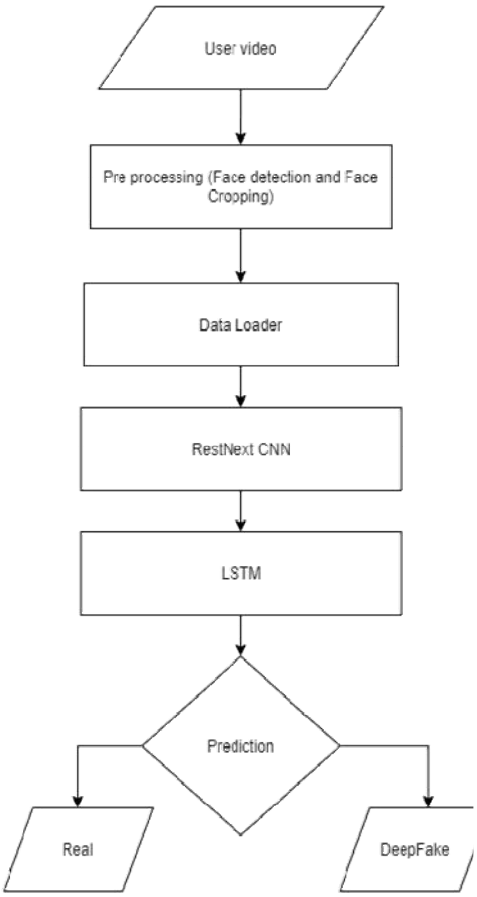

Fig 7. Flowchart

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

V. RESULT & ANALYSIS

The model's output will indicate whether the video is a deepfake or a real video, accompanied by the model's confidence level

Fig 8

The training is done for 20 epochs with a learning rate of le- 5 (0.00001),weight decay of le-3 (0.001) using the Adam optimizer.

Model Name	Dataset	No. of videos	Sequence length	Accuracy
model_90_acc _20_frames_ FF_data	Face Forensic++	2000	20	90.95477
model_95_acc _40_frames_ FF_data	Face Forensic++	2000	40	95.22613
model_97_acc _60_frames_ FF_data	Face Forensic++	2000	60	97.48743
model_97_acc _80_frames_ FF_data	Face Forensic++	2000	80	97.73366
model_97_acc _100_frames_ FF_data	Face Forensic++	2000	100	97.76180
model_93_acc _100_frames_ celeb_FF_data	Celeb-DF + FaceForensic++	3000	100	93.97781
model_87_acc _20_frames_ final_data	Our Dataset	6000	20	87.79160
model_84_acc _10_frames_ final_data	Our Dataset	6000	10	84.21461
model_89_acc _40_frames_ final_data	Our Dataset	6000	40	89.34681

Accuracy

Precision, Recall, F1 Score, Accuracy, Specificity

Models	Precision	Recall	F1 Score	Accuracy	Specificity
Model_40Frames	0.94	0.95	0.94	0.95	0.94
Model_20Frames	0.89	0.89	0.89	0.90	0.92
Model_60Frames	0.95	0.96	0.96	0.97	0.95
Model_10Frames	0.83	0.79	0.82	84.21	0.81

IJARSCT
ISSN (Online) 2581-9429
International Journal of Advanced Research in Science, Communication and Technology (IJARSCT)
International Open-Access, Double-Blind, Peer-Reviewed, Refereed, Multidisciplinary Online Journal
Impact Factor: 7.53
Volume 4, Issue 1, May 2024

VI. CONCLUSION

We introduced a neural network-based method for classifying videos as either deepfake or real, providing confidence scores generated by the proposed model. The proposed method is inspired by the way the deep fakes are created by the GANs with the help of Autoencoders. Our approach conducts frame-level detection utilizing ResNext CNN, while video classification is performed using a combination of RNN and LSTM. The proposed method is capable of detecting the video as a deep fake or real based on the listed parameters in paper. We believe that, it will provide a high accuracy on real time data

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 "Exposing AI Created Fake Videos by Detecting Eye Blinking" in arxiv.
[3]. Huy H. Nguyen , Junichi Yamagishi, and Isao Echizen " Using capsule networks to detect forged images and videos ".
[4]. Umur Aybars Ciftci, ˙Ilke Demir, Lijun Yin "Detection of Synthetic Portrait Videos using Biological Signals" in arXiv:1901.02212v2.
[5]. 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.
[6]. David G uera and Edward J Delp. Deepfake video detection using recurrent neural networks. In AVSS, 2018.
[7]. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, 2016.
[8]. An Overview of ResNet and its Variants : https://towardsdatascience.com/an-overview-of-resnet- and-its-variants-5281e2f56035
[9]. Long Short-Term Memory: From Zero to Hero with Pytorch: https://blog.floydhub.com/long- short-term-memory-from-zero-to-hero-with- pytorch/
[10]. Sequence Models And LSTM Networks https://pytorch.org/tutorials/beginner/nlp/sequenc e_mod els_tutorial.html