International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Face Morphing Identification Using LSTM and ResNext

[1]Mr. A. Yugandhara Rao,[2]P. Anupama, [3]K. Ashok Raj, [4]L. Teja, [5]R. Venkat Kiran.

[1]Associate Professor, [2,3,4,5]Student -8th Semester
Department of CSE, Lendi Institute of Engineering & Technology, Vizianagaram, India

ABSTRACT:

People will believe in what they see rather than hidden facts. In recent times as technology is improving and it became very easy to change a particular person in an existing image or video with someone else. So, spreading these types of modified videos causes spamming and peculating wrong information over social media and this will cause the great extent of misleading and threatening to the common people. To overcome these tedious situations, we are building a deep learning model that will identify the genuinity of the video or image in terms of the origin of its source. In our proposed system we are using Long Short-Term Memory and ResNext.

Keywords: **ResNext, Convolution neural network, Recurrent Neural Network (RNN), Long Short- Term Memory (LSTM), Computer vision, Deepfake Video Detection.**

1. INTRODUCTION

Morphing is a technique for human image synthesis based on neural network tools like GAN (Generative Adversarial Network) or Auto Encoders etc. These tools super impose target images onto source videos using a deep learning technique and create a realistic looking morphed video. These morphed videos are so real that it becomes impossible to spot difference by the naked eyes. Spreading of the DF over the social media platforms have become very common leading to spamming and peculating wrong information over the platform. These types of the Deepfakes will be terrible, and lead to threating, misleading of common people. In this work, we describe a new deep learning-based method that can effectively distinguish AI-generated fake videos from real videos. We are using the limitation of the morphing creation tools as a powerful way to distinguish between the pristine and duplicate videos. During the creation of the morphing the current morphing creation tools leave some distinguishable artifacts in the frames which may not be visible to the human being but the trained neural networks can spot the changes.

To discover the Deepfake it's veritably important to understand the way Generative Adversarial Network (GAN) creates the Deepfake. GAN takes as input a videotape and an image of a specific existent (‘target’), and labors another videotape with the target’s faces replaced with those of another existent (‘source’). The backbone of Deepfake is deep adversarial neural networks trained on face images and target videos to automatically collude the faces and facial expressions of the source to the target. With proper post-processing, the performing videos can achieve a high position of literalism. The GAN resolves the videotape into frames and replaces the input image in every frame. Further, it reconstructs the videotape. This process is generally achieved by using autoencoders. We describe a new deep learning-based method that can effectively distinguish Deepfake videos from real ones. Our system is grounded same process that's used to produce the Deepfake by GAN. The system is grounded on the properties of the Deepfake videos, due to the limitation of computation resources and production time, the Deepfake algorithm can only synthesize face images of a fixed size, and they must undergo an affinal warping to match the configuration of the source’s face. This warping leaves some distinguishable artifacts in the output deepfake videotape due to the resolution inconsistency between the depraved face area and the surrounding context. Our system detects similar artifacts by comparing the generated face areas and their surrounding regions by disassociating the videotape into frames and extracting the features with a ResNext Convolutional Neural Network (CNN) and using the Recurrent Neural Network (RNN) with Long Short-Term Memory (LSTM) capture the temporal inconsistencies between frames introduced by GAN during the reconstruction of the Deepfake. To train the ResNext CNN model, we simplify the process by simulating the resolution inconsistency in affine face wrappings directly.

2. LITERATURE SURVEY

The incendiary growth of deep fake video and its unlawful use is a major threat to democracy, justice, and legitimacy. Due to this, there is an increased demand for fake video analysis, detection, and intervention. Some of the related words in deep fake detection are listed below:

[1] Yuezun Li, Siwei Lyu “Exposing Deepfake Videos by Detecting Face Warping Artifacts” used an approach to detect artifacts by comparing the generated face areas and their girding regions with a dedicated Convolutional Neural Network model. In this work there were two-fold of Face Artifacts.

Their approach is grounded on the compliances that current Deepfake algorithm can only induce images of limited resolutions, which are also demanded to be further converted to match the faces to be replaced in the source videotape.

[2] Yuezun Li, Ming-Ching Chang and Siwei Lyu "Exposing AI Created Fake Videos by Detecting Eye Blinking" describes a new system to expose fake face videos generated with deep neural network models. The system is grounded on discovery of eye blinking in the videos, which is a physiological signal that isn't well presented in the synthesized fake videos. The system is estimated over marks of eye- blinking finding datasets and shows Deep Neural Network based software Deepfake.

Their approach only uses the lack of blinking as an indication for detection. Nevertheless, certain other parameters must be considered for discovery of the deep fake like teeth enchantment, wrinkles on faces etc. Our approach is advanced to consider all these parameters.

[3] Huy H. Nguyen , Junichi Yamagishi, and Isao Echizen "Using capsule networks to discover forged images and videos" uses a system that uses a capsule network to descry forged, manipulated images and videos in different scenarios, like iteration attack detection and computer- generated video detection.

In their methodology, they've used arbitrary noise in the training phase which isn't a good option. Still, the model performed salutary in their dataset but may fail on real- time data due to noise in training. Our methodology is proposed to be trained on noiseless and real- time datasets.

[4] Umur Aybars Ciftci, ˙Ilke Demir, Lijun Yin "Discovery of Synthetic portrayal Videos using the Biological Signals" approach excerpts biological signals from facial regions on authentic and fake portrayal video pairs. Apply transformations to compute the spatial coherence and temporal consistency, capture the signal characteristics in property sets and PPG maps, and train a probabilistic SVM and a CNN. Then, the aggregate authenticity chances to decide whether the video is fake or authentic.

Fake Catcher detects fake content with high delicacy, independent of the generator, content, resolution, and quality of the video. Due to the lack of a discriminator leading to the loss in their findings to save biological signals, articulating a differentiable loss function that follows the proposed signal processing way isn't straight forward process.

3. PROPOSED SYSTEM

While there are numerous tools accessible for making Deepfakes, there are very few tools available for their detection. In order to prevent the Deepfake from spreading over the internet, our method for recognizing it will be a big help. We'll offer a web-based platform where users may upload videos and mark them as bogus or authentic. From creating a web-based platform to a browser plugin for automatic Deepfake detections, this project can be expanded up. Even large applications like Facebook's WhatsApp can incorporate this project into their software for simple Deepfake pre-detection before sending to another user. Evaluating its performance and acceptance in terms of security, user friendliness, accuracy, and reliability is one of the key goals. figure.1 represents the simple system architecture of the proposed system:

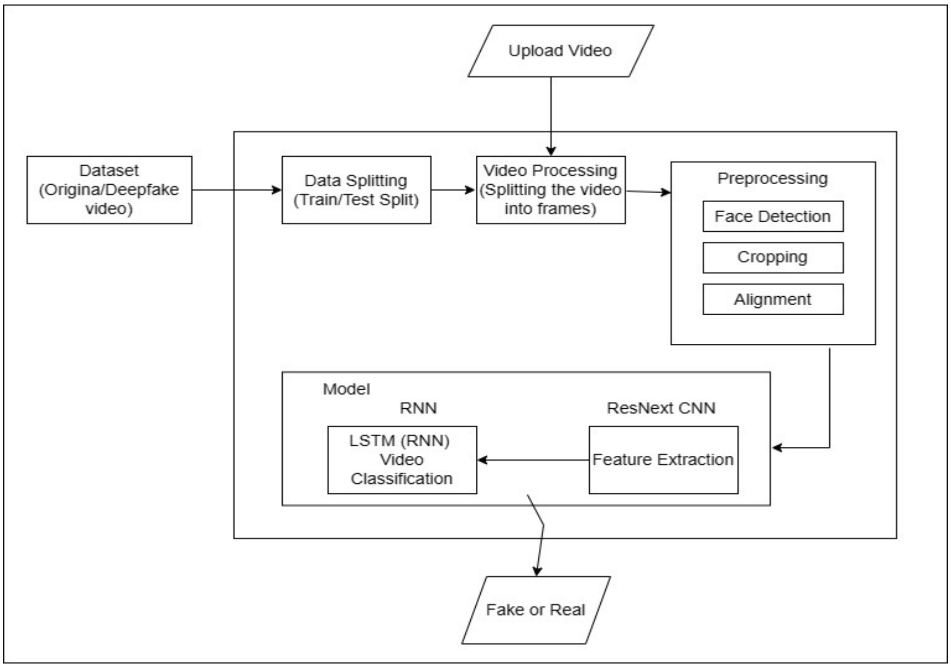

Upload Video
Dataset (Origina/Deepfake video)
Data Splitting (Train/Test Split)
Video Processing (Splitting the video into frames)
Preprocessing
Face Detection
Cropping
Alignment
Model
RNN
ResNext CNN
LSTM (RNN) Video Classification
Feature Extraction
Fake or Real

Fig. 1: System Architecture

A. Dataset:

We're utilizing a mixed dataset made up of an equal number of videos from various dataset sources, including YouTube, FaceForensics++[14], and the Deep fake detection challenge dataset [13]. 50% of the original video and 50% of the altered deepfake videos are included in our recently created dataset. The dataset is divided into a 30% test set and a 70% train set.

B. Preprocessing:

The video is divided into frames as part of the dataset preprocessing procedure. Face detection and cropping the frame to include the found face come next. The mean of the dataset video is determined in order to maintain consistency in the number of frames, and a new processed face-cropped dataset is constructed using the frames that make up the mean. Preprocessing ignores the frames that don't contain any faces.

It will take a lot of computing power to process the 300 frames in a 10-second video at 30 frames per second. Therefore, we are suggesting that for experimental purposes, the model be trained using only the first 150 frames.

C. Model:

The model comprises one LSTM layer followed by resnext50_32x4d. The preprocessed face-cropped videos are loaded by the data loader, which partitions them into a train set and a test set. Additionally, the model receives the frames from the edited videos for training and testing in small batches.

D. ResNext CNN for Feature Extraction

We suggest using the ResNext CNN classifier for properly recognizing the frame level features rather than constructing the classifier from scratch in order to extract the features. The network will then be fine-tuned by adding any additional necessary layers and choosing an appropriate learning rate to properly converge the gradient descent of the model. The sequential LSTM input is then taken from the 2048-dimensional feature vectors that remain after the last pooling layers.

E. LSTM for Sequence Processing

Consider a 2-node neural network using a sequence of ResNext CNN feature vectors of input frames as input, along with probability that the sequence is either a deep fake video or an unaltered video. The design of a model to recursively process a sequence in a meaningful way is the main issue that needs to be addressed. We suggest using a 2048 LSTM unit with a 0.4 likelihood of dropout for this task to accomplish our goal. By comparing the frame at second 't' with the frame at second 't-n', LSTM is used to sequentially process the frames to do a temporal analysis on the video. In which n is the number of frames before to t.

F. Predict:

The trained model receives a new video for prediction. Additionally, a fresh video is preprocessed to incorporate the trained model's format. The video is divided into frames, followed by face cropping, and the cropped frames are immediately sent to the trained model for detection rather than being stored locally.

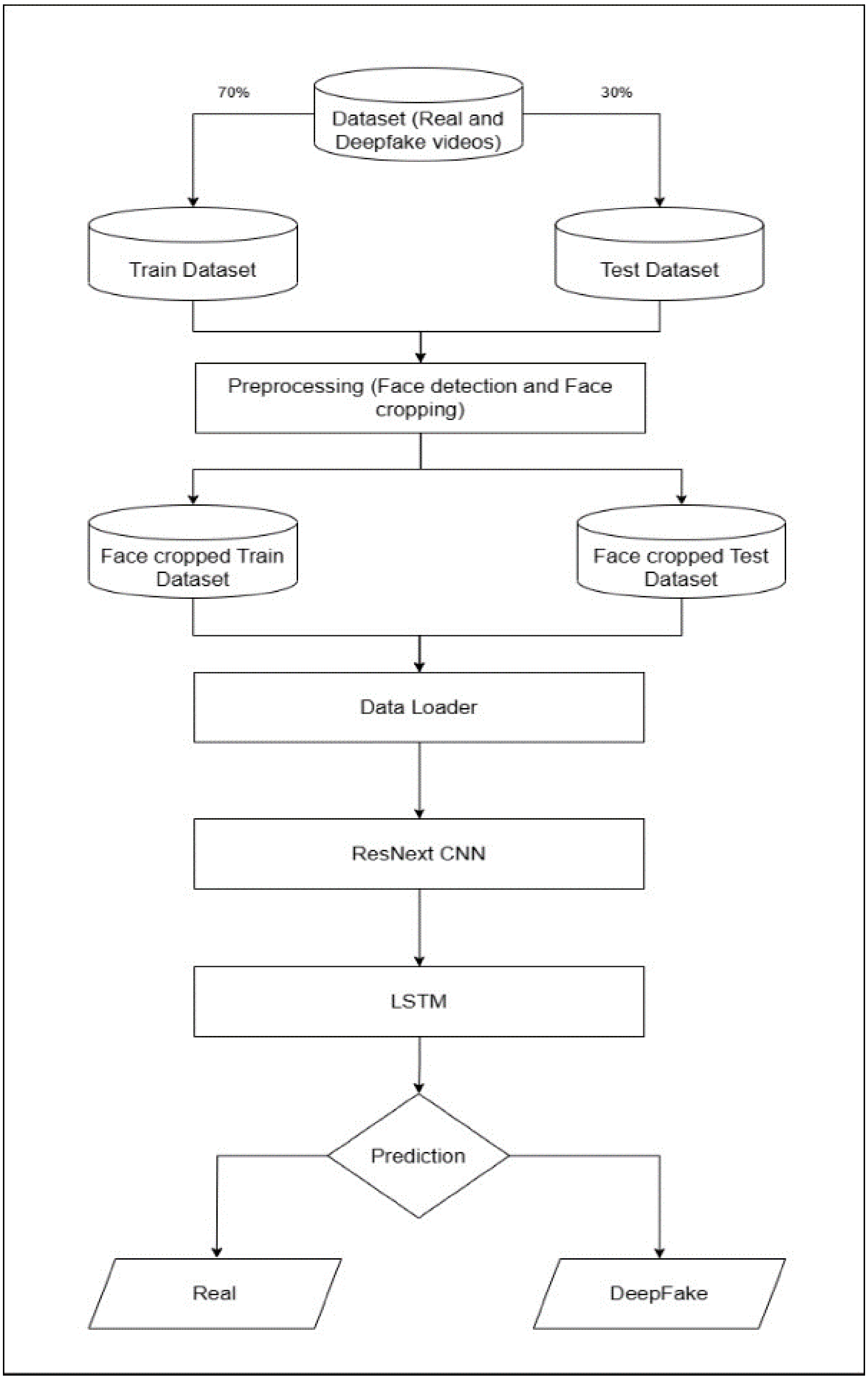

70%
30%
Dataset (Real and Deepfake videos)
Train Dataset
Test Dataset
Preprocessing (Face detection and Face cropping)
Face cropped Train Dataset
Face cropped Test Dataset
Data Loader
ResNext CNN
LSTM
Prediction
Real
DeepFake

Fig. 2: Training Flow

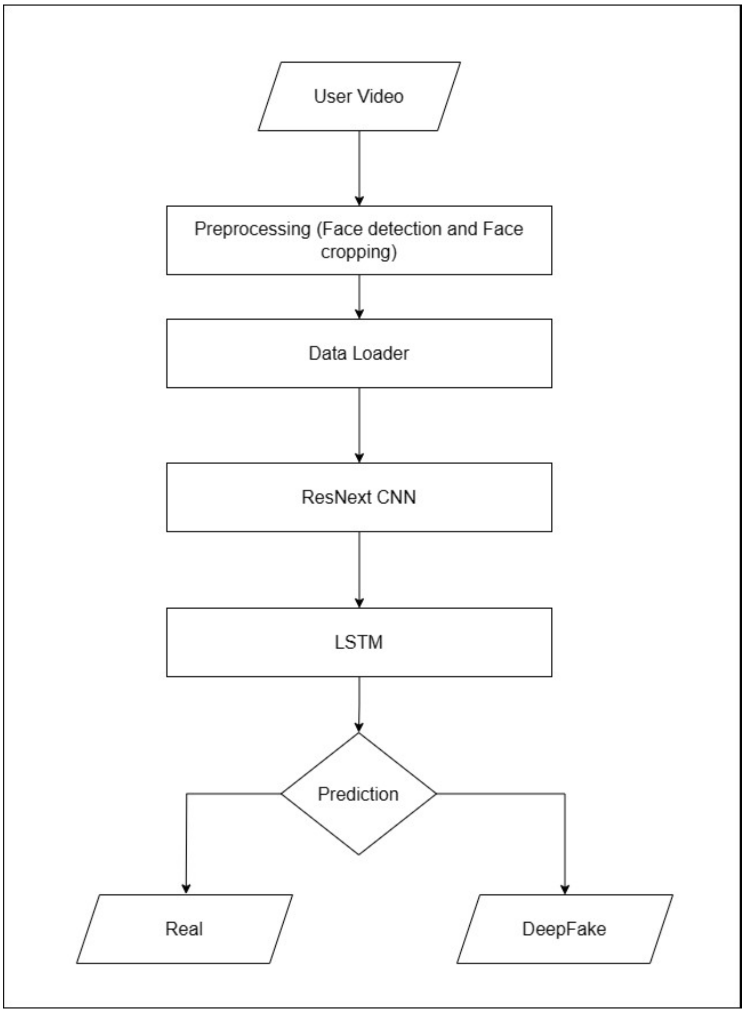
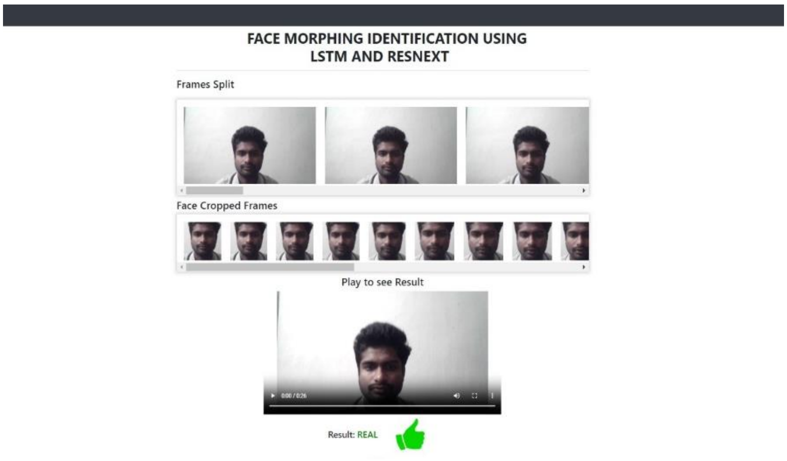


Fig. 3: Prediction Flow

4. RESULT

The model's output will include the model's confidence level and a determination of whether the video is authentic or a deep fake. Figure 4 presents one instance.



FACE MORPHING IDENTIFICATION USING LSTM AND RESNEXT

Frames Split

Face Cropped Frames

Play to see Result

Result: REAL

Fig. 4: Expected Results

5. Conclusion

We provided a neural network-based method for determining if a video is a deep fake or the real thing, along with the model's level of confidence. The deep fakes produced by GANs with the aid of Autoencoders serve as an inspiration for the suggested strategy. Our approach uses ResNext CNN for frame level detection and RNN and LSTM for video classification. Based on the factors stated in the study, the suggested method can determine if a video is a deep fake or real. We think it will deliver real-time data with extremely high accuracy.

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] Hyeongwoo Kim, Pablo Garrido, Ayush Tewari and Weipeng Xu "Deep Video Portraits" in arXiv:1901.02212v2.

[5] Umur Aybars Ciftci, ˙Ilke Demir, Lijun Yin "Detection of Synthetic Portrait Videos using Biological Signals" in 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. In NIPS, 2014.

[7] David G¨uera and Edward J Delp. Deepfake video detection using recurrent neural networks. In AVSS, 2018.

[8] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In CVPR, 2016.

[9] An Overview of ResNet and its Variants: https://towardsdatascience.com/an-overview-of-resnet-and-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_mod els_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

[15] Y. Qian et al. Recurrent color constancy. Proceedings of the IEEE International Conference on Computer Vision, pages 5459–5467, Oct. 2017. Venice, Italy.

[16] P. Isola, J. Y. Zhu, T. Zhou, and A. A. Efros. Image-to-image translation with conditional adversarial networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

[17] 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.

[18] Tiago de Freitas Pereira, Andr´e Anjos, Jos´e Mario De Martino, and S´ebastien Marcel, "Can face anti spoofing countermeasures work in a real-world scenario?,"in ICB. IEEE, 2013.

[19] Nicolas Rahmouni, Vincent Nozick, Junichi Yamagishi, and Isao Echizen, "Distinguishing computer graphics from natural images using convolution neural networks," in WIFS. IEEE, 2017.

[20] F. Song, X. Tan, X. Liu, and S. Chen, "Eyes closeness detection from still images with multi-scale histograms of principal oriented gradients," Pattern Recognition, vol. 47, no. 9, pp. 2825–2838, 2014.

[21] D. E. King, "Dlib-ml: A machine learning toolkit,"JMLR, vol. 10, pp. 1755–1758, 2009.