Deepfake Video Detection and Fake News Detection using Machine Learning

Abhishek Kumar Bharti[1] Mudit Singh[2] Sailesh Singh[3] Rishabh Shukla[4] Mr. Ketan Anand[5]
[1,2,3,4]*Student, Department of Computer Science, Axis Institute of Technology and Management, Kanpur, U.P, India*
[5] *Assistant Professor, Department of Computer Science, Axis Institute of Technology and Management, Kanpur, U.P, India*

***Abstract*— In recent months, free deep literacy-grounded software tools have eased the creation of believable face exchanges in videos that leave many traces of manipulation, in what they're known as" DF"(DF) videos. Manipulations of digital videos have been demonstrated for several decades through the good use of visual goods, recent advances in deep literacy have led to a drastic increase in the literalism of fake content and the availability in which it can be created. These so-called AI- synthesized media(popularly appertained to as DF). Creating the DF using the Instinctively intelligent tools are simple task. But, when it comes to discovery of these DF, it's major challenge. Because training the algorithm to spot the DF isn't simple. We've taken a step forward in detecting the DF using Convolutional Neural Network and intermittent neural Network. System uses a convolutional Neural network(CNN) to prize features at the frame position. These features are used to train a intermittent neural network(RNN) which learns to classify if a videotape has been subject to manipulation or not and suitable to descry the temporal inconsistencies between frames introduced by the DF creation tools. Anticipated result against a large set of fake videos collected from standard data set. We show how our system can be competitive result in this task results in using a simple armature.**

Keywords: **DF Video Detection, convolutional Neural network (CNN), recurrent neural network (RNN)**

I. INTRODUCTION

The adding complication of smartphone cameras and the vacuity of good internet connection each over the world has increased the ever- growing reach of social media and media participating doors have made the creation and transmission of digital videos easier than ever ahead. The growing computational power has made deep literacy so important that would have been allowed insolvable only a sprinkle of times agone Like any transformative technology, this has created new challenges. So- called" DF" produced by deep generative inimical models that can manipulate videotape and audio clips. Spreading of the DF over the social media platforms have come veritably common leading to spamming and peculating wrong information over the platform. These types of the DF will be terrible, and lead to threating, misleading of common people.

To overcome such a situation, DF discovery is veritably important. So, we describe a new deep literacy- grounded system that can effectively distinguish AI- generated fake videos (DF videos) from real videos. It's incredibly important to develop technology that can spot fakes, so that the DF can be linked and averted from spreading over the internet. For discovery the DF it's veritably important to understand the way Generative Adversarial Network (GAN) creates the DF. 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 DF are deep inimical 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 resolve 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 literacy- grounded system that can effectively distinguish DF videos from the real bones. Our system is grounded same

process that's used to produce the DF by GAN. The system is grounded on a parcel of the DF videos, due to limitation of calculation coffers and product time, the DF algorithm can only synthesize face images of a fixed size, and they must suffer an affinal screwing to match the configuration of the source's face. This screwing leaves some distinguishable vestiges in the affair DF videotape due to the resolution inconsistency between depraved face area and girding environment.

Our system detects similar vestiges by comparing the generated face areas and their girding regions by unyoking the videotape into frames and rooting the features with a ResNext Convolutional Neural Network (CNN) and using the intermittent Neural Network (RNN) with Long Short Term Memory (LSTM) prisoner the temporal inconsistencies between frames introduced by GAN during the reconstruction of the DF. To train the ResNext CNN model, we simplify the process by bluffing the resolution inconsistency in affine face wrappings directly.

II. LITERATURE SURVEY

The explosive growth in deep fake video and its illegal use is a major trouble to democracy, justice, and public trust. Due to this there is a increased the demand for fake video analysis, discovery and intervention. Some of the related word in deep fake discovery are listed below:

ExposingDF videos by Detecting Face Warping Vestiges used an approach to detects vestiges by comparing the generated face areas and their girding regions with a devoted Convolutional Neural Network model. In this work there were two-fold of Face Artifacts. Their system is predicated on the obedience that current DF algorithm can only induce images of limited judgments, which are also demanded to be further converted to match the faces to be replaced in the source video.

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 predicated on discovery of eye blinking in the videos, which is a physiological signal that is not well presented in the synthesized fake videos. The system is estimated over marks of eye- blinking discovery datasets and shows promising performance on detecting videos generated with Deep Neural Network grounded software DF. Its relative insensitivity to gap length is its advantage over other RNNs, hidden Markov models and other sequence literacy styles.Their system only uses the lack of blinking as a indication for discovery. still certain other parameters must be considered for discovery of the deep fake like teeth enchantment, wrinkles on faces etc. Our system is proposed to consider all these parameters. Using capsule networks to descry forged images and videos uses a system that uses a capsule network to descry forged, manipulated images and videos in different scripts, like renewal attack discovery and computer-generated videotape discovery. In their system, 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 system is proposed to be trained on quiet and real time datasets. Discovery of Synthetic portrayal videos using natural Signals approach excerpt natural signals from facial regions on authentic and fake portrayal videotape dyads. Apply metamorphoses to cipher the spatial consonance and temporal thickness, prisoner the signal characteristics in point sets and PPG charts, and train a probabilistic SVM and a CNN. also, the total authenticity chances to decide whether the videotape is fake or authentic. Fake Catcher detects fake content with high delicacy, independent of the creator, content, resolution, and quality of the videotape. Due to lack of discriminator leading to the loss in their findings to save natural signals, formulating a differentiable loss function that follows the proposed signal processing way isn't straight forward process.

III. PROPOSED SYSTEM

There are numerous tools available for creating the DF, but for DF discovery there's hardly any tool available. Our approach for detecting the DF will be great donation in avoiding the percolation of the DF over the world wide web. We'll be furnishing a web-grounded platform for the stoner to upload the videotape and classify it as fake or real. This design can be gauged up from developing a web- grounded platform to a cybersurfed plugin for automatic DF

findings. Indeed, big operation like WhatsApp, Facebook can integrate this design with their operation for easy pre discovery of DF before transferring to another person. One of the important ideals is to estimate its performance and adequacy in terms of security, person- benevolence, delicacy and trustability. Our system is fastening on detecting all types of DF like relief DF, retrenchment DF and interpersonal DF. figure.1 represents the simple system armature of the proposed system-

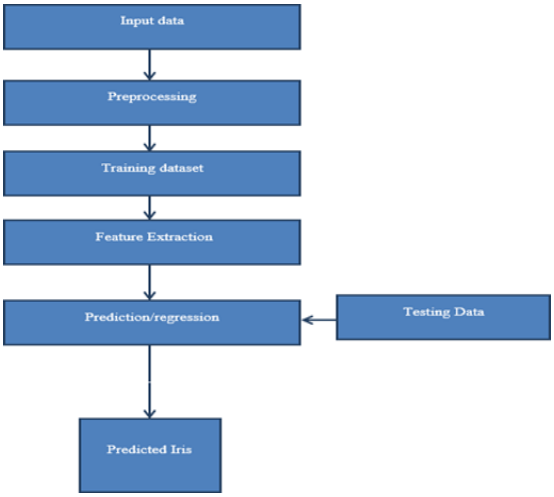

Fig 1. Data Flow Diagram

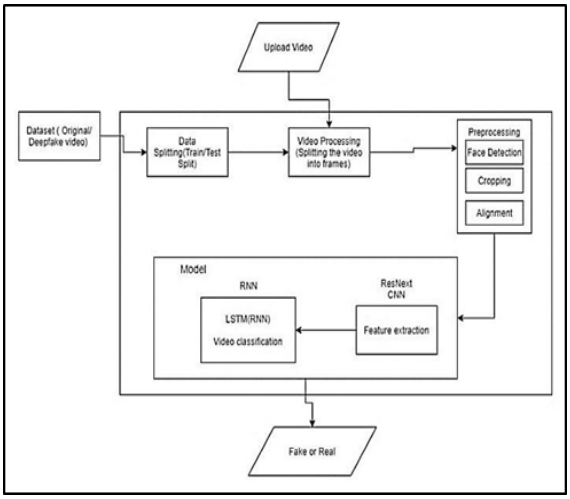

Fig. 2: System Architecture

A. Dataset

We're using a mixed dataset which consists of equal quantum of videos from different dataset sources like YouTube, FaceForensics, Deep fake discovery challenge dataset. Our recently prepared dataset contains 50 of the original videotape and 50 of the manipulated DF videos. The dataset is resolve into 70% train and 30% test set.

B. Preprocessing:

Dataset preprocessing includes the unyoking the videotape into frames. Followed by the face discovery and cropping the frame with detected face. To maintain the uniformity in the number of frames the mean of the dataset videotape is calculated and the new reused face cropped dataset is created containing the frames equal to the mean. The frames that do not have faces in it are ignored during preprocessing. As recycling the 10 alternate videotape at 30 frames per alternative total 300 frames will bear a lot of computational power. So, for experimental purpose we're proposing to used only first 100 frames for training the model.

C. Model:

The model consists ofresnext50_32x4d followed by one LSTM subcaste. The Data Loader loads the preprocessed face cropped videos and resolve the videos into train and test set. Further the frames from the reused videos are passed to the model for training and testing in mini batches.

D. ResNext CNN for Feature Extraction:

ResNext CNN for point birth rather of writing the rewriting the classifier, we're proposing to use the ResNext CNN classifier for rooting the features and directly detecting the frame position features. Following, we will be fine- tuning the network by adding redundant needed layers and opting a proper literacy rate to duly meet the grade descent of the model. The 2048- dimensional point vectors after the last pooling layers are also used as the successional LSTM input.

E. LSTM for Sequence Processing:

LSTM for Sequence Processing Let us assume a sequence of ResNext CNN point vectors of input frames as input and a 2- knot neural network with the chances of the sequence being part of a deep fake videotape or an untampered videotape. The crucial challenge that we need to address is the de- sign of a model to recursively reuse a sequence in a meaningful manner. For this problem, we're proposing to the use of a 2048 LSTM unit with0.4 chance of powerhouse, which is able to do achieve our ideal. LSTM is used to reuse the frames in a successional manner so that the temporal analysis of the videotape can be made, by comparing the frame at ‘ t ’ second with the frame of ‘ t- n ’ seconds. Where n can be any number of frames before t.

F. Predict:

A new videotape is passed to the trained model for vaticination. A new videotape is also preprocessed to bring in the format of the trained model. The videotape is resolve into frames followed by face cropping and rather of storing the videotape into original storehouse the cropped frames are directly passed to the trained model for detection.

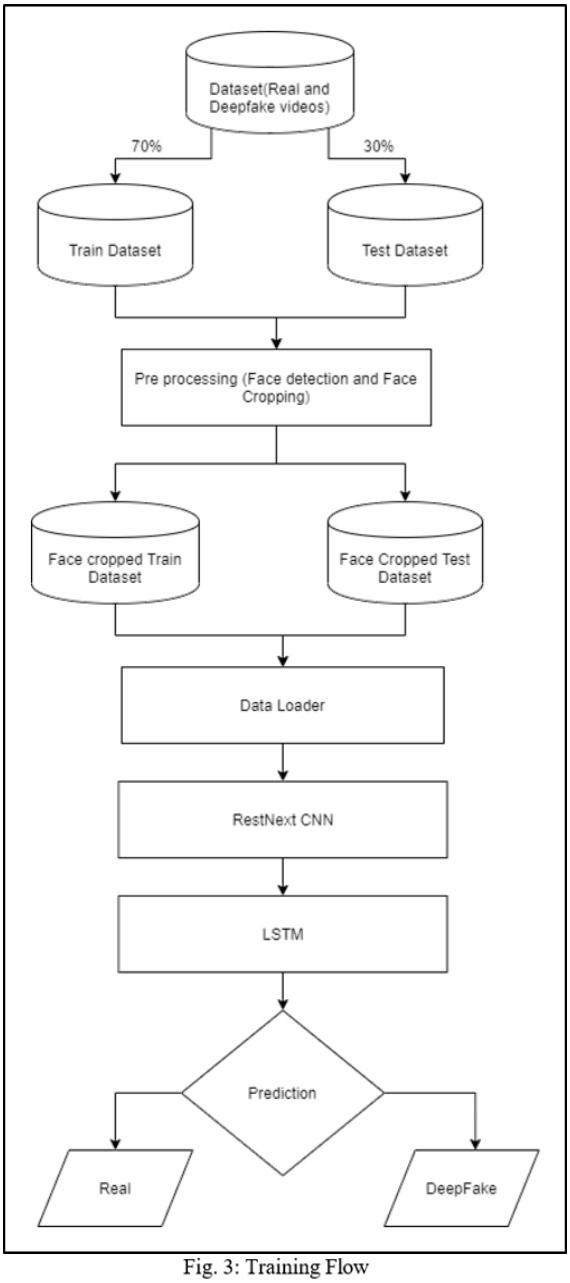

Fig. 3: Training Flow

IV. RESULT

The affair of the model is going to be whether the videotape is DF or a real videotape along with the confidence of the model. One illustration is shown in the figure 3.

Fig. 4: Expected Results

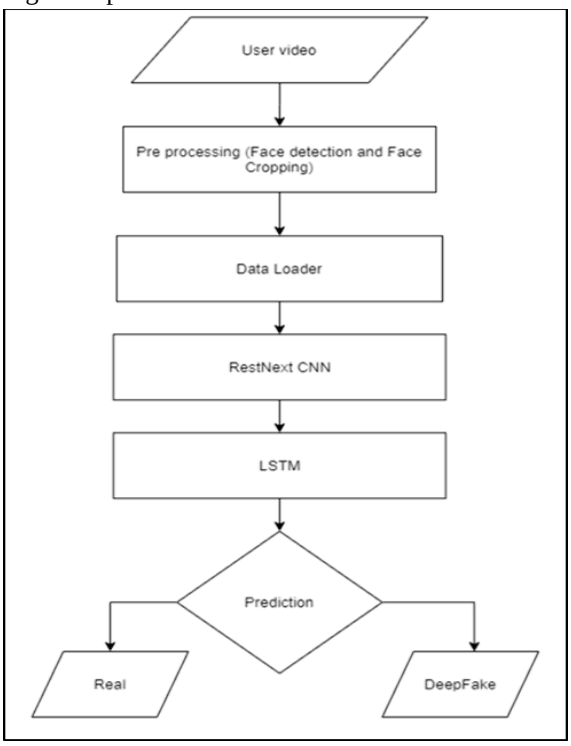

Fig 5. Prediction Flow

V. CONCLUSION

We presented a neural network-based approach to classify the video as deep fake or real, along with the confidence of proposed model. The proposed method is inspired by the way the deep fakes are created by the GANs with the help of Autoencoders. Our method does the frame level detection using ResNext CNN and video classification using RNN along with 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 very high accuracy on real time data.

IV. LIMITATIONS

Our method has not considered the audio. That's why our method will not be able to detect the audio deep fake. But we are proposing to achieve the detection of the audio deep fakes in the future.

REFERENCE

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

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

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

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

[9] https://www.kaggle.com/c/DF- detection-challenge/data

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

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

[12] 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, pages 5967– 5976, July 2017. Honolulu, HI.

[13] 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.