International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395-0056
Volume: 10 Issue: 08 | Aug 2023 www.irjet.net p-ISSN: 2395-0072

A Neural Network Approach to Deep-Fake Video Detection

Prajwal Chunarkar[1], Vivek Upadhyay[1], Sagar Sanap[1] , Anirudh Talmale[1], Prof. Varshapriya J N [2]

[1]BTech Student, Dept of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India
[2]Associate Professor, Dept of Computer Engineering and IT, VJTI College, Mumbai, Maharashtra, India

---

Abstract - *Deep learning has been used to solve a variety of complex challenges, including big data analytics, computer vision, and human-level control. Deep learning developments, on the other hand, have been used to develop applications that pose a risk to anonymity, democracy, and national security. Deepfake is a recent example of a deep learning-powered technology. Deepfake algorithms can generate fake photographs and videos that are difficult to tell apart from real ones. As a result, the development of technology that can automatically identify and measure the integrity of digital visual media is critical.Using artificially intelligent software to build the DF is an easy process. However, detecting these DF poses a significant challenge. Since it is difficult to train the algorithm to detect the DF. Using Convolutional Neural Networks and Recurrent Neural Networks, we have made progress in detecting the DF. At the frame stage, the system employs a convolutional Neural network (CNN) to extract features. These characteristics are used to train a recurrent neural network (RNN), which learns to classify whether or not a video has been manipulated and can detect frame temporal inconsistencies caused by DF creation tools. Expected outcome as compared to a large number of fake videos gathered from a regular data set. We demonstrate how using a simple architecture, our system can achieve competitive results in this mission.*

Key Words: Deepfake Video Detection, convolutional, Neural network (CNN), recurrent neural network (RNN)

1.INTRODUCTION

Recent advances in artificial intelligence (AI) and cloud computing technology have resulted in rapid developments in audio, video, and image processing techniques. Deepfakes are the term for this kind of fake media material AI-based tools will also exploit media in more convincing ways, such as duplicating a public figure's voice or superimposing one person's face over another's body. Deep generative adversarial models that can exploit video and audio clips generate "DeepFake." The spread of the DF through social media channels has become very popular, resulting in spamming and the dissemination of incorrect information. This form of DF would be bad, and will threaten and confuse ordinary citizens.

The importance of DF identification in such a case cannot be overstated. As a result, we present a new deep learning-based method for distinguishing between AI-generated fake videos (DF Videos) and real videos. It's very important to develop technology that can identify fakes, so that the DF can be spotted and prevented from spreading over the internet.

2. LITERATURE REVIEW

Deep fake video's exponential development and illicit usage pose a serious threat to government, justice, and public trust. As a result, the market for fake video review, monitoring, and interference has risen. The following are some of the relevant studies in deep fake identification.

Exposing AI Created Fake Videos by Detecting Eye Blinking [1] explains a new approach for exposing fake face videos produced by deep neural network models. The approach relies on the identification of eye blinking in videos, which is a physiological signal that isn't well shown in the fake videos. The method is tested on eye-blinking recognition datasets and shows positive results when it comes to detecting videos created with Deep Neural Network based software DF.

Their system relies solely on the absence of blinking as an identification clue. However, other factors such as teeth enchantment, lines on the forehead, and so on must be noticed when detecting a deep fake. All of these criteria are taken into account by our system.

Using capsule networks to detect forged images and videos [2] uses a method that uses a capsule network to detect forged, manipulated images and videos in different scenarios, like replay attack detection and computer-generated video detection.

In their method, they have used random noise in the training phase which is not a good option. Still the model performed beneficial in their dataset but may fail on real time data due to noise in training. Our method is proposed to be trained on noiseless and real time datasets.

Luo et al. [3] introduces JPEG error analysis to determine when bitmap images have been compressed previously, approximate quantization steps, and detect the quantization table in a JPEG image. Their ability to detect JPEG blocks as small as 8x8 assists in the identification of tampered areas in an image. Bianchi et al.

J. H. Bappy et al. [4] introduce a method for manipulation localization that uses an LSTM network and an encoder-decoder system to construct a mapping from low resolution activations to pixel-wise predictions. CNNs have

been successfully used to identify fake photographs in recent years.

The Biological Signals Method for Detecting Synthetic Portrait Videos [5] extracts biological signals from facial regions on authentic and fake portrait video pairs. Train a probabilistic SVM and a CNN using transformations to compute spatial coherence and temporal consistency, capture signal characteristics in feature sets and PPG maps, and capture signal characteristics in feature sets and PPG maps. The aggregate authenticity probabilities are then used to determine whether the video is genuine or not.

3. DATA PREPARATION

Basic concept- Deepfakes have a very basic definition. Let's presume we want to add person A's face to a video of person B.

To begin, we gather hundreds or thousands of photographs for both individuals. Using a deep learning CNN network, we built an encoder to encode all of these images. The picture is then reconstructed using a decoder. This auto encoder (encoder and decoder) has over a million parameters, but is not even close enough to remember all the pictures.

Intuitively, the encoder detects face angle, skin tone, facial expression, lighting, and other data needed to recreate individual A.

Fig -I: Original vs Deepfake

We draw individual B but with the background of A when we use the second decoder to reconstruct the picture. The reconstructed image in the image below has Trump's facial characters while keeping the goal video's facial expression.

The task of Deep-Fake detection is framed as a binary classification task, with real and fake images as the two classes. We use supervised learning to train a Convolutional Neural Network on sets of real and false samples to estimate the likelihood of test images being influenced by deep-fake manipulations. The various Deep-Fake video datasets used in this method are described in this section, as well as the process for curating image datasets for training and testing.

Fig -II :Dataset Distribution

A. Deep-Fake Datasets

This research makes use of three of the most recent Deep-Fake video datasets that were recently made public –

1. DeepFakeDetectionChallenge (DFDC) Preview [8]- Facebook AI has compiled the most recent entry in the field of Deep-Fake datasets. To prevent cross-set face swaps, a pool of 66 paying actors was divided into train and test sets, and their filmed sequences were used to produce manipulated videos internally. There are 5214 videos in total, with 78.125 percent of them being manipulated. They achieved high manipulation efficiency by selecting pairs of identical appearances, and visual quality is also high.

2. FaceForensics is a forensics dataset made up of 1000 original video sequences that were edited using four different automatic face simulation methods: Deepfakes, Face2Face, FaceSwap, and NeuralTextures. The data came from 977 YouTube images, all of which have a trackable largely frontal face with no occlusions, allowing automatic tampering methods to create practical forgeries.

3. Celeb-DF [7] - The Celeb-DF dataset contains 590 origin-al YouTube videos of subjects of various ages, ethnic groups, and races, as well as 5639 DeepFake videos.

Table -I: Sample Table format

Dataset	Real	Fake
DFDC	1566	1727
FF	1000	1000
Celeb-DF	580	588
Combined	3146	3315

4. PROPOSED SYSTEM

There are many resources available for generating the DF, but there are few tools available for detecting the DF. Our method for detecting the DF can make a significant contribution to preventing the DF from percolating. Detection of all forms of DF, including replacement, retrenchment, and interpersonal DF.

The model is made up of resnext50 32x4d and one LSTM layer. The Data Loader loads the preprocessed face cropped videos and divides them into two sets, one for training and one for testing.

We propose that instead of rewriting the classifier, we use the ResNext CNN classifier for extracting features and accurately detecting frame level features.

We need to fix the de- sign of a model to recursively process a sequence in a meaningful manner, so we use LSTM for Sequence Processing.

We're also using a transfer learning approach to boost our model's generalizability over a wide variety of datasets.

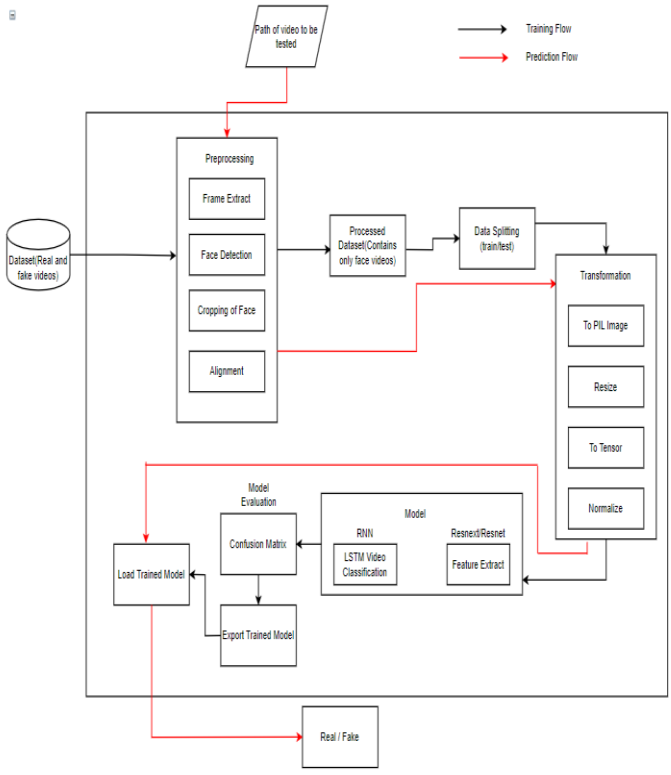

Fig -III: System Architecture

A. Dataset:

We are using a mixed dataset which consists of equal amount of videos from different dataset sources like YouTube, FaceForensics++[9,] Deep fake detection challenge dataset[8]. Our newly prepared dataset contains 48.69% of the original video and 51.31% of the manipulated deepfake videos. The dataset is split into 70% train and 30% test set.

B. Preprocessing:

Splitting the video into frames is part of the dataset preprocessing. Following that, face detection is done, and the frame containing the detected face is cropped. To keep the number of frames constant, the video dataset's mean is calculated, and a new processed face cropped dataset with frames equal to the mean is created. Frames that do not contain faces are ignored during preprocessing.

It will take a lot of computing power to process a 20-second video at 25 frames per second, or 500 frames in total. So, for the sake of experimentation, we recommend training the model with only the first 100 frames.

C. Model:

The model is made up of resnext50 32x4d/ resnet50 and one LSTM layer. The Data Loader loads the preprocessed face cropped videos and divides them into two sets, one for training and one for testing. In addition, the frames from the processed videos are transferred to the model in mini batches for training and testing.

D. ResNext CNN/ ResNet CNN for Feature Extraction

Instead of writing the rewriting the classifier, we are proposing to use the ResNext CNN /ResNet CNN classifier for extracting the features and accurately detecting the frame level features. Following, we will be fine-tuning the network by adding extra required layers and selecting a proper learning rate to properly converge the gradient descent of the model. The 2048-dimensional feature vectors after the last pooling layers are then used as the sequential LSTM input.

E. LSTM for Sequence Processing

Assume a 2-node neural network with the probabilities of the series being part of a deep fake video or an untampered video as input and a sequence of ResNext CNN feature vectors of input frames as output. The main problem we must solve is the design of a model that can recursively process a sequence in a meaningful way. We propose using a 2048 LSTM unit with a 0.4 probability of dropping out to solve this problem, which is capable of achieving our goal. The LSTM is used to process the frames in a sequential manner such that the video can be temporally analyzed by comparing the frame at 't' second to the frame at 't-n' seconds. Before t, n can be any number of frames.

F. Training:

We choose one of the 2 types of Convolutional Neural Network architecture Resnet50 and Resnext50_32x4d which

are trained on ImageNet [10] dataset which contains over 14 million images and can classify over 1000 categories. A lot of low-level features can be learned from these models which are going to help learning algorithm do better. The Adam optimization algorithm is used, along with a cross-entropy loss function and a mini-batch size of four videos. All the training is performed on the NVIDIA Tesla P100GPUs. The transfer learning aided models were initialized with ImageNet weights. The NVIDIA Tesla P100GPUs are used for all of the training. ImageNet weights were used to initialize the transfer learning aided models. This process is referred as fine-tuning. The model is trained with a slightly lower learning rate. As compared to a model trained without Transfer Learning, this results in a substantial increase in validation accuracy after only one epoch.

G. Predict:

The learned model is given a new video to predict. A new video is also preprocessed to incorporate the trained model's format. The video is split into frames, then face cropped, and instead of storing the video locally, the cropped frames are transferred directly to the qualified model for detection.

It is inevitable.

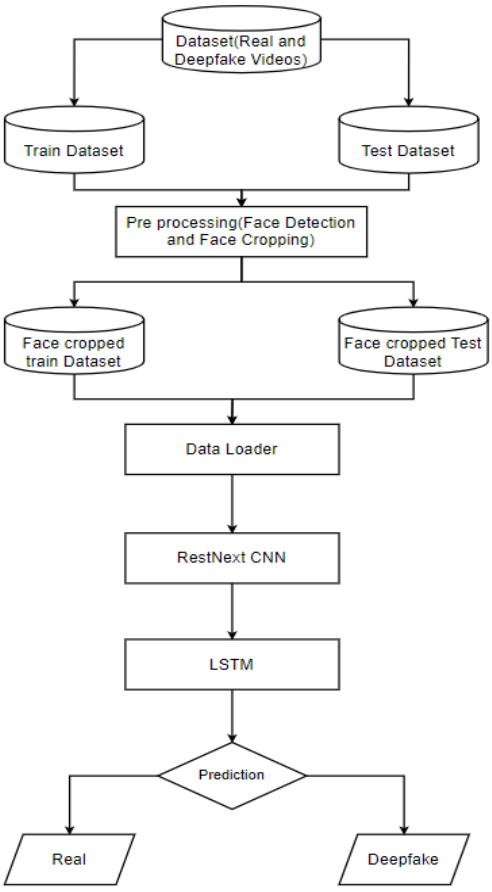

Fig -IV: Training Flow

5. Result

Table II can be used to draw a lot of inferences. Throughout the table, Transfer Learning has had a significant effect on accuracies in both Resnet50 with LSTM and Resnext50_32x4d with LSTM models. With the exception of one block when we used ResNext with Transfer Learning with LSTM sequence length 60, the strong influence of increasing sequence length can be seen invariably throughout the table.

Fig -V: Expected Results

Table -II: Sample Table format

Sequence Length of LSTM	MODELS			
	RESNET50 with LSTM		RESNEXT50 with LSTM	
	With TL	Without TL	With TL	Without TL
20	82.97	62.44	84.46	65.62
40	86.54	65.16	89.96	66.01
60	89.11	67.65	87.71	67.80
80	89.96	68.04	92.61	68.94
100	91.67	70.13	93.24	69.56

6. Conclusion

The results show that the intuition behind the effectiveness of Transfer Learning for Deep-Fake Detection is accurate. Another interesting observation is that ResNext CNN with LSTM is giving better results in terms of accuracies than ResNet CNN with LSTM keeping sequence length same for both , with the exception of only two blocks having accuracies higher in Resnet CNN with LSTM than ResNext CNN with LSTM . The accuracy increase rate with increase in

sequence length decreases with increase in sequence length, but the computational cost to train the model increases tremendously so it is not feasible to just keep increasing the sequence length.

We introduced a neural network-based method for classifying videos as deep fake or true, as well as the model's trust level. Our system uses ResNext CNN for frame level identification and RNN and LSTM for video classification. Based on the criteria mentioned in the paper, the proposed method will detect whether a video is a deep fake or not. It can, we think, have extremely high precision on real-time data.

REFERENCES

[1] Yuezun Li, Siwei Lyu, "Exposing DF Videos By Detecting Face Warping Artifacts," in arXiv:1811.00656v3. M. Young, The Technical Writer's Handbook. Mill Valley, CA: University Science, 1989.

[2] Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen "Using capsule networks to detect forged images and videos".

[3] Luo, W., Huang, J., & Qiu, JPEG Error Analysis and Its Applications to Digital Image Forensics, IEEE Transactions on Information Forensics and Security, 5(3), pp 480 491, 2010.

[4] J. H. Bappy, C. Simons, L. Nataraj, B. S. Manjunath and A. K. Roy-Chowdhury, "Hybrid LSTM and Encoder-Decoder Architecture for Detection of Image Forgeries," IEEE Transactions on Image processing. 28, no. 7, pp. 3286-3300, 2019.

[5] Umur Aybars Ciftci, ˙Ilke Demir, Lijun Yin "Detection of Synthetic Portrait Videos using Biological Signals" in arXiv:1901.02212v2.

[6] He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages

[7] Y. Li, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celb-DF: A new dataset for deepfake forensics," arXiv 1909.12962, 2019.

[8] B. Dolhansky, R. Howes, B. Pflaum, N. Baram, C. Ferrer, "The Deepfake detection Challenge (DFDC) Preview, Dataset", arXiv:1910:08854, 2019.

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

[10] Deng J, Dong W, Socher R, Li L-J, Li K, Fei-Fei L. Imagenet: A large-scale hierarchical image database. In: 2009 IEEE conference on computer vision and pattern recognition. 2009.

[11] M. Torres, T. Raeder, R. Alaiz-RodiGuez, N. Chawla, F. Herrera, "A unifying view on dataset shift in classification", Pattern Recognition, Elsevier, 2012.

[12] 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.