A Project/Dissertation Report

on

DEEPFAKE DETECTION IN VIDEO

Submitted in partial fulfillment of the

requirement for the award of the degree of

COMPUTER SCIENCE AND ENGINEERING

Under The Supervision of

Surendra Singh Chauhan

(Assistant Professor)

Submitted By

ROHAN KAPOOR

(19SCSE1140010/19021140008)

PRIYANKA YADAV

(19SCSE10109021/9021012027)

SCHOOL OF COMPUTING SCIENCE AND ENGINEERING

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

GALGOTIAS UNIVERSITY,GREATER NOIDA

INDIA

CANDIDATE'S DECLARATION

I/We hereby certify that the work which is being presented in the project, entitled "**DEEPFAKE DETECTION IN VIDEO**" in partial fulfillment of the requirements for the award of the Bachelor Degree submitted in the School of Computing Science and Engineering of Galgotias University, Greater Noida, is an original work carried out during the period of JULY2021 to DECEMBER2021, under the supervision of **Surendra Singh Chauhan** AP, Department of Computer Science and Engineering of School of Computing Science and Engineering , Galgotias University, Greater Noida

The matter presented in the project has not been submitted by me/us for the award of any other degree of this or any other places.

ROHAN KAPOOR- 19SCSE1140010

PRIYANKA YADAV-19SCSE1010902

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Surendra Singh Chauhan

Assistant Professor

CERTIFICATE

The Final Project Viva-Voce examination of PRIYANKA YADAV (19SCSE1010902) ROHAN KAPOOR (19SCSE1140010) has been held on _____ and his/her work is recommended for the award of B.TECH.

Signature of Examiner(s) **Signature of Supervisor(s)**

Signature of Project Coordinator **Signature of Dean**

Date: December, 2021

Place: Greater Noida

ACKNOWLEDGEMENT

I am overwhelmed in all humbleness and gratefulness to acknowledge my depth to all those who have helped me to put these ideas, well above the level of simplicity and into something concrete. I would like to express my special thanks of gratitude to my project guide who gave me the golden opportunity to do this wonderful project on the topic “DEEPFAKE DETECTION IN VEDIO” which also helped me in doing a lot of Research and i came to know about so many new things. I am really thankful to them.

Any attempt at any level can‘t be satisfactorily completed without the support and guidance of my friends.

I would like to thank my friends who helped me a lot in gathering different information, collecting data and guiding me from time to time in making this project. Despite their busy schedules, they gave me different ideas in making this project unique.

ABSTRACT

Deep learning has been successfully applied to solve various complex problems ranging from big data analytics to computer vision and human-level control. Deep learning advances however have also been employed to create software that can cause threats to privacy, democracy and national security. One of those deep learning-powered applications recently emerged is deepfake. Deepfake algorithms can create fake images and videos that humans cannot distinguish them from authentic ones. The proposal of technologies that can automatically detect and assess the integrity of digital visual media is therefore indispensable. This paper presents a survey of algorithms used to create deepfakes and, more importantly, methods proposed to detect deepfakes in the literature to date. We present extensive discussions on challenges, research trends and directions related to deepfake technologies. By reviewing the background of deepfakes and state-of-the-art deepfake detection methods, this study provides a comprehensive overview of deepfake techniques and facilitates the development of new and more robust methods to deal with the increasingly challenging deepfakes.

Deep faux videos are AI-generated movies that look actual however are fake. Deep faux films are normally created with the aid of face-swapping techniques. It started out as amusing however like every technology, it is being misused. inside the starting, these motion pictures could be recognized with the aid of human eyes. However, due to the improvement of machine getting to know, it has become less difficult to create deep fake films. It has almost come to be indistinguishable from actual motion pictures. Deep faux motion pictures are generally created by the use of GANs (Generative hostile network) and different deep gaining knowledge of technology. The chance of this is that era may be used to make humans believe something is real when it isn't. cellphone desktop applications like FaceApp and pretend Apps are constructed in this method. Those videos can have an effect on a person’s integrity. So, identifying and categorizing these movies has come to be a necessity. This paper evaluates strategies of deepfake detection and discusses how they can be combined or changed to get greater accurate outcomes. With a bit of luck, we will be able to make the internet a safer location.

Contents

CHAPTER-1 Introduction

1.1 Introduction

Deep Face detection is turning into a mile more popular topic amongst nowadays computer vision international. Deep Fakes talk over when a performance by an actor is superimposed onto a photograph or video of a target person to make it appear like the target is appearing the actions that the actor is doing. The introduction of deepfakes has been enabled through current AI/ML advances and cutting-edge deepfakes are clearly imperceptible from actual people to human eyes. This era is devastating to humans focused by using them, as politicians can be made to offer speeches, they in no way could have, archive photos can be doctored. It's far therefore important that there exist robust algorithms to distinguish real photos or photos from deep fakes. Detecting deep fakes is thrilling, as they're swiftly becoming greater widespread in these days internationally, have severe capacity for damage, and are an exceptionally difficult mission for humans to perform unaided.

A developing disquiet has settled around the emerging deepfake that makes it viable to create proof of scenes that have by no means ever taken place. Celebrities and politicians are those who're drastically affected by this. Deepfake can optimally stitch anybody right into a video or photograph that they in noway have real expertise with. Nowadays due to the fact technologies are elevating broadly the structures can synthesize photographs and motion pictures extra quick. A writer would first teach a neural network on many hours of real video photos to provide it a realistic know-how of what she or he seems like on many angles or lights in order to create a deep fake video of a person. Then they could integrate the trained network into graphics techniques to superimpose a replica of person into exclusive one. Creative use of artificial voice and video can beautify overall success and learning outcomes with scale and limited expenditure. Deepfakes can democratize the VFX era as a robust tool for unbiased storytellers. Deepfakes additionally has disadvantages which have an effect on extraordinary businesses of our society. It is getting used to revenge porn to defame famous personalities, developing fake news and propaganda and many others. As quickly as these faux films cross viral humans consider initially, and keep on sharing with others makes the focused individual embarrassed watching this fake stuff.

Keywords: *Deepfake, deepfake detection, deep learning, detection techniques, eye blinking*

1.2 Formulation of Problem

• To save you from hoaxes, financial frauds, faux information, etc.

• To prevent the terror activities as faux picture graphs can be utilized in passports and other authorities identification-cards so to save you terror activities.

• Deepfake detection can also be used to make social media bills as the images may be used as the profile snapshots in the social media.

• To examine new things as it's far an AIML primarily based undertaking so we will study more about pandas, NumPy's, and so on.

1.3 Tools & Technologies

- Programming Languages
  - Python3
  - JavaScript
- IDE
  - Google colab
  - Visual Studio Code
- Cloud Services
  - Google Cloud Platform

WHAT ARE DEEPFAKES?

Deepfakes- Deepfakes are synthetic media in which someone in an existing image or video is transformed into person else's likeness. The act of injecting a faux character in a photograph is not new. However, recent Deepfakes strategies commonly leverage the recent improvements of effective GAN models, aiming at facial manipulation.

In general, facial manipulation is usually conducted with Deepfakes and can be categorized in the following categories:

· **Face synthesis**- In this category, the goal is to create non-existent practical faces for the usage of GANs. The most popular approach is style GAN. in short, a brand new generator structure learns separation of excessive-stage attributes (e.g., pose and identity while trained on human faces) without supervision and stochastic variant inside the generated pictures (e.g., freckles, hair), and it permits intuitive, scale-particular control of the synthesis.

· **Face swap**- Face swap is the most popular face manipulation class in recent times. The aim here is to discover whether an image of a person is faux after swapping its face. The most popular database with faux and real snapshots is Face Forensic. The fake images in this dataset were made using computer pics (Faceswap) and deep mastering strategies (Deepfake). The Face switch app is written in Python and makes use of face alignment, Gauss-Newton optimization, and photograph mixing to change the face of a person seen by using the digicam with a face of a person in a supplied photograph.

· **Facial attributes and expression**- Facial attributes and expression manipulation consist of enhancing attributes of the face together with the colour of the hair or the pores and skin, the age, the gender, and the expression of the face with the aid of making it glad, unhappy, or angry. The most famous instance is the Face app cellular software that recently came into existence. The majority of those approaches adopt GANs for image-to-image translation. One of the first-class performing strategies is megastar Gan that makes use of a single version skilled across multiple attributes' domain names in place of education of multiple turbines for every area.

Literature Survey

In this section, we are going to discuss the various literature works in the Deepfake creation and detection domain.

Xin Yang, et. al. [1] have proposed a system to detect Deepfake using inconsistent headposes. Algorithms used in the previous model create the face of different persons without changing the original expressions hence creating mismatched facial landmarks. The landmark locations of few false faces often vary from those of the real faces, as a consequence of interchanging faces in the central face region in the DeepFake process. The difference in the distribution of the cosine distances of the two head orientation vectors for real and Deepfakes suggest that they can be differentiated based on this cue. It uses the DLib package for face detection and to extract 68 facial landmarks. The standard facial 3D model is created with OpenFace2, and then difference is calculated. The proposed system uses UADFV dataset. Trained SVM classifier with Radial basis function (RBF) kernels on the training data is used. Area Under ROC (AUROC) of 0.89, is achieved by the SVM classifier on the UADFV dataset. The crucial point that can be inferred from this paper is the focus on how the Deepfakes are generated by splicing a synthesized face region into the original image, and how it can also use 3D pose estimation for detecting synthesized videos.

Rohita Jagdale, et. al. [2] have proposed a novel algorithm NA-VSR for Super resolution. The algorithm initially reads the low resolution video and converts it into frames. Then the median filter is used to remove unwanted noise from video. The pixel density of the image is increased by bicubic interpolation technique. Then Bicubic transformation and image enhancement is done for mainly resolution enhancement. After these steps the design metric is computed. It uses the output peak signal-to-noise ratio (PSNR) and structural similarity index method (SSIM) to determine the quality of image. Peak signal-to-noise ratio and structural similarity index method parameters are computed for NA-VSR and compared with previous methods. Peak signal to noise ratio (PSNR) of the proposed method is improved by 7.84 dB, 6.92 dB, and 7.42 dB as compared to bicubic, SRCNN, and ASDS respectively.

Siwei Lyu,[3] has surveyed various challenges and also discussed research opportunities in the field of Deepfakes. One critical disadvantage of the current DeepFake generation methods is that they cannot produce good details such as skin and facial hairs. This is due

to the loss of information in the encoding step of generation. Head puppetry involves copying the source person's head and upper shoulder part and then pasting it on the target person's body , so that target appears to behave in a similar way as that of the source. The second method is face swapping which swaps only the face of the source person with that of the target. It also keeps the facial expressions unchanged. The third method is Lip syncing which is used to create a falsified video by only manipulating the lip region so that the target appears to speak something that she/he does not speak in reality. The detection methods are formulated as frame level binary classification problems. Out of the three widely used detection methods, the first category considers inconsistencies exhibited in the physical/physiological aspects in the DeepFake videos. The second algorithm makes use of the signal-level artifacts. Data driven is the last category of Detection in this, it directly employs multiple types of DNNs trained on genuine and Fake videos but captures only explicit artifacts. It also sheds some light on the limitations of these methods such as quality of deepfake datasets, social media laundering, etc.

Digvijay Yadav, et. al. [4] have elaborated the working of the deepfake techniques along with how it can swap faces with high precision. The Generative Adversarial Neural Networks (GANs) contain two neural networks, the first is generator and other is discriminator. Generator neural networks create the fake images from the given data set. On the other hand, discriminator neural networks evaluate the images which are synthesized by the generator and check its authenticity. Deepfake are harmful because of cases like individual character defamation and assassination, spreading fake news, threat to law enforcement agencies. For detection of Deepfakes blinking of eyes can be considered as a feature. The limitations for making Deepfakes are the requirement of large datasets, training and swapping is time consuming, similar faces and skin tones of people, etc. Deepfake video detection can be done using recurrent neural networks. CNN is best known for its visual recognition and if it is combined with LSTM it can easily detect changes in the frames and then this information is used for detecting the DeepFakes. The paper suggests that Meso-4 and Mesoinception-4 architectures are capable of detecting the Deepfake video with the accuracy of 95% to 98% on Face2Face dataset.

Irene Amerini, et. al. [5] have proposed a system to exploit possible inter-frame dissimilarities using the optical flow technique. CNN classifiers make use of this clue as a feature to learn. The optical flow fields calculated on two consecutive frames for an original video and the corresponding Deepfake one are pictured and it can be noticed that the motion vectors around the chin in the real sequence are more vociferous in comparison with those of the altered video that appear much smoother. This is used as a clue to help neural networks learn properly. FaceForensics++ dataset was used, in that 720 videos were

used for training, 120 videos for validation, and 120 videos for testing. They used two neural networks VGG16 and ResNet 50. For Face2Face videos, VGG gives detection accuracy of 81.61 % and ResNet50 gives detection accuracy of 75.46 %. The uniqueness of this paper is the consideration of inter-frame dissimilarities, unlike other techniques which rely only on intra-frame inconsistencies and how to overcome them using the optical flow based CNN method.

XTao, et. al. [6] have proposed a system that emphasizes the fact that to achieve better results, proper frame alignment and motion compensation needs to be done. The authors have introduced a sub-pixel motion compensation layer (SPMC) layer in a CNN framework. Along with FlowNet-S CNN, frame alignment and motion compensation is achieved using motion compensation transformer (MCT) module. Also, they have collected 975 sequences from high-quality 1080p HD video clips publically available on the internet and downsampled the original frames to 540 × 960 pixels. The proposed method has a Peak signal-to-noise ratio (PSNR) of 36.71 and a Structural Similarity Index (SSIM) value of 0.96 which is better than that of the previously proposed SRCNN. This paper provides an insight into how to organize multiple frame inputs for getting better results. Also, it gives a foreknowledge about how data is to be sampled before feeding it to the CNN model.

David guera, et. al. [7] have demonstrated how Deepfake videos are created and how they can be detected using CNN and LSTM. GAN's are used for better quality deepfake videos. For the generation process, the encoder of the original image is used but for swapping faces with the target image, the decoder of the target image is used. They tried various techniques over Deepfake videos for devising accurate detection systems and came to the final conclusion that the best accuracy was found when the video was split into 80 frames per second along with a combination of CNN and LSTM. The maximum accuracy which was acquired was around 97.1%. But the accuracy which was acquired was on a set of high-resolution images. The above paper illustrates in great detail how the Deepfake videos are generated.

Jiangang Yu, et. al. [8] have proposed a system that focuses on one drawback of the super resolution algorithm, which is low accuracy for videos with facial changes. For super resolution, first the video is broken down into multiple frames and then CNN is applied over each frame differently. Authors found out that when the video has facial expression changes, it is very difficult to produce higher accuracies for super resolution systems. To overcome this problem, the system proposed how handling of a facial image in a non-rigid manner can be done. The system proposed in the paper works in three steps 1) global

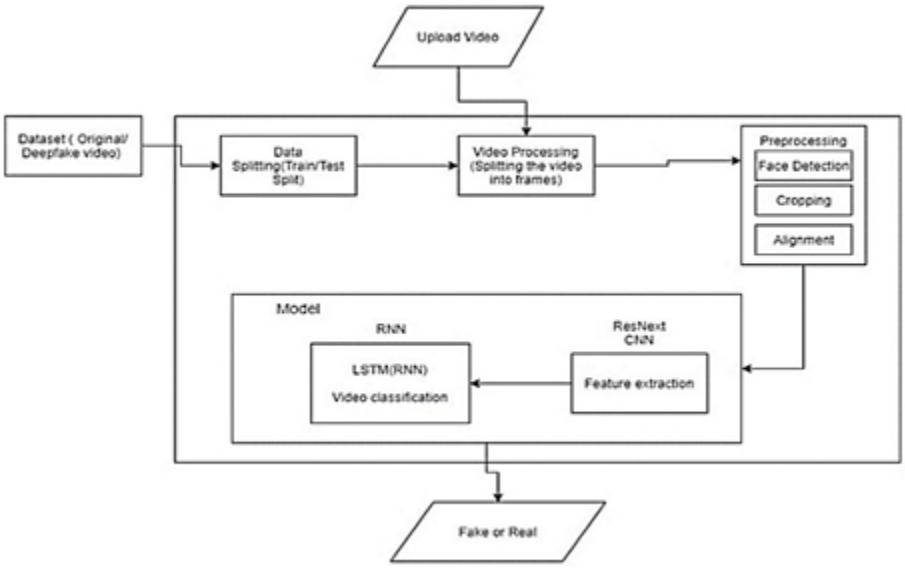
tracking 2) local alignment 3) Super Resolution algorithm. For performance measurement of the system, authors recorded 10 video sequences of 10 people lasting for about 3 minutes. The PSNR value was found to be 26.6261 which is a betterment from previous PSNR value 20.6261 using a global only approach.The above system has given insights on the problems faced by super resolution algorithm when there are facial changes. If the facial expression changes hamper the accuracy of super resolution, it will affect the system very badly because the output of the super resolution phase is given as input to CNN stage in our system. To overcome this problem, the paper has given a solution of using the handling of facial image in a non-rigid way. The PSNR value is also increased using this approach.

Analysis Table

Sr. No	Title of Paper	Techniques used	Dataset used	Accuracy
1	Deepfake: A Survey on Facial Forgery Techniqu Using Generative Adversarial Network[4]	1.Convolutional Net Network (CNN) 2.L Short-Term Memory LSTM)	Face2Face, Redd user deepfakes	95%
2	Deepfake Video Detectio through Optical Flow ba CNN[5]	1.Convolutional Net Network (CNN)	Face2Face	VGG16 81.61% ResNet50 75.4
3	Detail-revealing Deep Video Super-resolution[	1.FlowNet-S CNN With a sub-pixel motion compensation layer SPMC) layer	975 sequences fro high-quality 108 HD video clips	Method(F3) 36.71/0.96, Method (F5) 36.62/0.96
4	Deepfake Video Detectio Using Recurrent Neural Network[7]	1.Convolution Neura Networks (CNN) 2.Long Short-Term Memory (LSTM)	1.HOHA dataset	Conv-LSTM(2 frames) 96.7%,Conv-LS M(40 frames) 97.1%

The various algorithms and features used for the Deepfake detection are analyzed in the above table. It includes the Machine Learning and Deep Learning based techniques. From the analysis table above, it can be seen that CNN algo with the LSTM gives better results and accuracy which can be further increased by using the Concept of Super resolution.

Proposed System



1. Dataset:

We are using a mixed dataset which consists of equal amounts of videos from different dataset sources like YouTube, FaceForensics++, Deep fake detection challenge dataset.Our newly prepared dataset contains 50% of the original video and 50% of the manipulated deep fake videos. The dataset is split into two parts 70% train and 30% test set.

2. Preprocessing:

Dataset preprocessing includes splitting of the video into frames. It is followed by face detection and cropping the frame with detected face. To maintain the equality in the number of frames the mean of the dataset video is calculated, and the new processed face dataset is created containing the frames equal to the mean. The frames that do not have faces in it are ignored during preprocessing. As processing the 10 second video at 40 frames per second i.e total 400 frames will require a lot of computational power. So, for experimental purposes we are proposing to use only the first 100 frames for training the model.

3. Model:

The model consists of resnext50_32x4d with one Long-Short Term Memory layer. The Data Loader loads the preprocessed face cropped videos and splits the videos into train and testset. Further the frames from the processed videos are passed to the model for training and testing in mini batches.

4. ResNext CNN for Feature Extraction

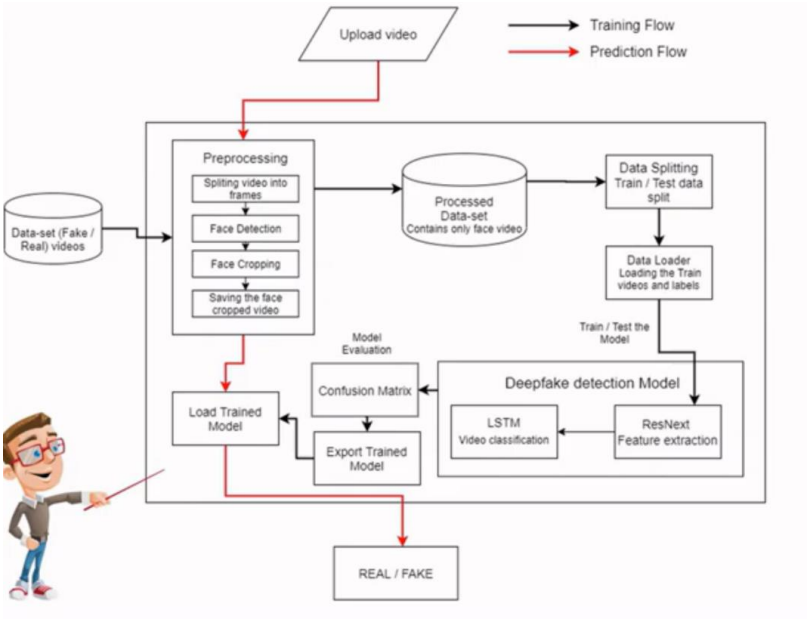
Instead of rewriting the classifier, we are proposing to use the ResNext CNN classifier for extracting the features and accurately detecting the frame level features. Following, we will be tuning the network by adding required layers and selecting a proper learning rate to properly converge the gradient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

5. LSTM for Sequence Processing

Let us assume a sequence of ResNext CNN feature vectors of input frames as input and a node neural network with the probabilities of the sequence being part of a deep fake video or an untampered video. The key challenge that we need to address is the de- sign of a model to recursively process a sequence in a meaningful manner. for the problem we are proposing to the use of a 2048 LSTM unit with 0.4 chance of dropout, which is capable of achieving our objective. LSTM is used to process the frames in a sequential manner so that the temporary analysis of the video can be made, by comparing the frame at ‘t’ second with the frame of ‘t-n’ seconds. Here n(frames) can be any number of frames before t(seconds).

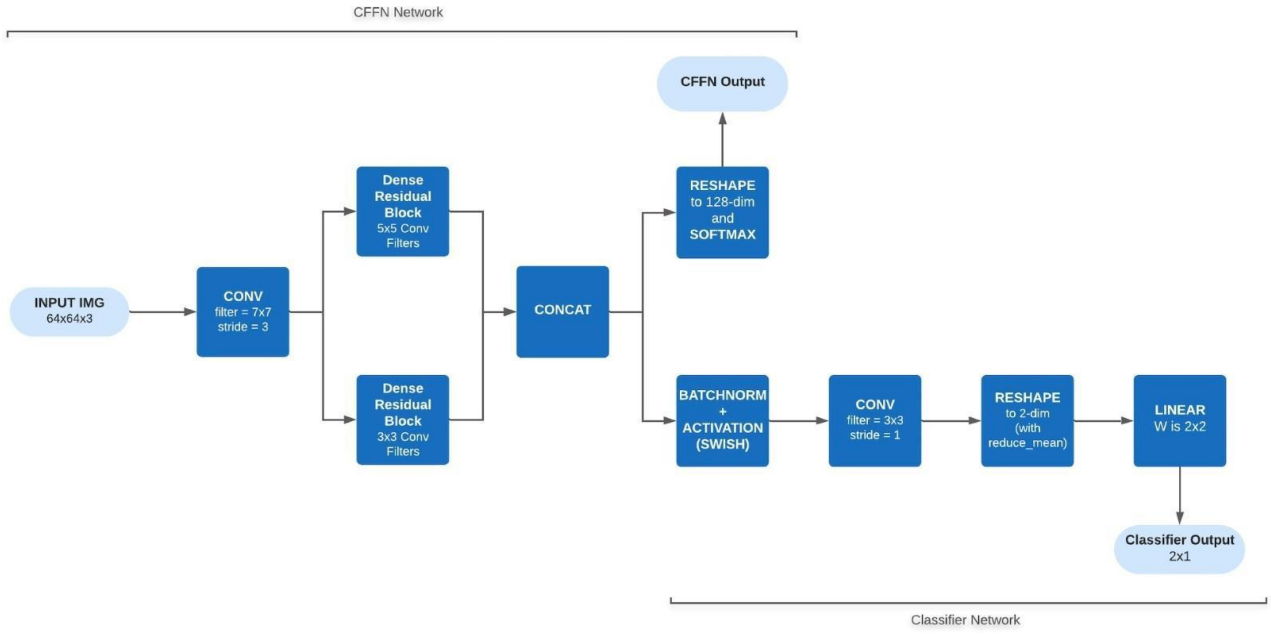
6. Predict:

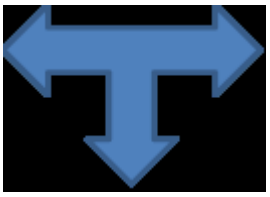
The new video is passed to the trained model for the prediction. The new video is preprocessed to bring the new video in the format of the trained model. The video is split into frames followed by face cropping and instead of storing the video into local storage the cropped frames are directly passed to the trained model for detection.



Result & Conclusiom

● The output of the model is going to be whether the photo is deepfake or a real photo along with the confidence of the model.





CONCLUSION

Deep learning can be used as a deepfake creation, and detection methods. Deepfake creates forged images or videos that persons cannot differentiate from real images or videos. Deepfakes are created using generative adversarial networks, in which two machine learning models exit. One model trains on a dataset and the other model tries to detect the deepfakes. The forger creates fakes until the other model can't detect the forgery. Deepfakes creating fake news, videos, images, and terrorism events that can cause social, and financial fraud. It is increasing affects religions, organizations, individuals and communities', culture, security, and democracy. When deepfake videos and images increase on social media people will ignore to trust the truth. In this study, the available datasets, deepfake creation tools, deepfake challenges, fake video detection techniques and detect fake video by using eye blinking were discussed. Also, the detection models trained on the datasets and the total and the eye-blink detection accuracy results were computed. Deepfake detection is a method to detect real and fake images or videos. In this study, the CNN to extract frame feature and to classify the eye states, and LSTM for temporal sequence analysis have been used. Also, the eye aspect ratio, used for eye blinking rate classification and the CNN and eye aspect ratio detect the eye blinking intervals. The detection models have been trained on UADFV publically available real and fake videos. The deepfake detection methods detect the deepfakes by eye blinking. In the experiment, the eye blinking detection accuracy result on real videos is 91.59% and eye blinking detection accuracy on fake videos 90.27%. Furthermore, the overall detection accuracy results on real videos is 93.23% and the overall detection accuracy on fake videos is 98.30%. In the eye blinking detection, when the person moves his/her head quickly and when the eye focus on the area below them the eyelids cover the eye and the eye detected as blink or closed this affects the accuracy of the model.

REFRENCES

[1] Xin Yang, Yeuzen Li and Siwei Lyu, “EXPOSING DEEP FAKES USING INCONSISTENT HEAD POSES”, ICASSP 2019 - 2019 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP).

[2] Rohita Jagdale and Sanjeevani Shah, “A Novel Algorithm for Video Super-Resolution”, Proceedings of ICTIS 2018, Volume 1, Information and Communication Technology for Intelligent Systems (pp.533-544).

[3] Siwei Lyu,”DEEPFAKE DETECTION: CURRENT CHALLENGES AND NEXT STEPS”,2020 IEEE International Conference on Multimedia & Expo Workshops(ICMEW)

[4] Digvijay Yadav, Sakina Salmani, “Deepfake: A Survey on Facial Forgery Technique Using Generative Adversarial Network”, Proceedings of the International Conference on Intelligent Computing and Control Systems (ICICCS 2019).IEEE Xplore Part Number: CFP19K34-ART; ISBN: 978-1-5386-8113-8.

[5] Irene Amerini, Leonardo Galteri, Roberto Caldelli, Alberto Del Bimbo, “Deepfake Video Detection through Optical Flow based CNN”, 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW).

[6] Xin Tao, Hongyun Gao, Renjie Liao, Jue Wang, Jiaya Jia, “Detail-revealing Deep Video Super-resolution”, 2017 IEEE International Conference on Computer Vision (ICCV).

[7] David Guera, Edward J. Delp, “Deepfake Video Detection Using Recurrent Neural Networks”, 2018 15th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS).

[8] Jiangang Yu and Bir Bhanu,“Super-resolution of Facial Images in Video with Expression Changes”,IEEE 5th Conference on Advanced Video and Signal based Surveillance,2018.