A Project/Dissertation ETE Report

on

DEEPFAKE DETECTION IN VIDEO

Submitted in partial fulfillment of the
requirement for the award of the degree of

COMPUTER SCIENCE AND ENGINEERING

UnderTheSupervision of
SurendraSinghChauhan
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 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 deep fakes 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 no way 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. It may want to give individuals new equipment for self-expression and amalgamation inside the on-line world. 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.

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
    - Jupyter Notebook
    - 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.

CHAPTER-2 Literature Survey

In this section, we are going to discuss the various literature works in the Deepfake creation and detection domain. In the paper [1], they give a comprehensive survey on the video content authentication techniques. These techniques are usually categorized as active and passive models and this paper gives a detailed survey on the various passive blind video content authentication with the main focus on forgery detection, video recapture, and phylogeny detection. Later with the advancement in Deep Learning, Deepfake has become popular among a wide range of users due to the high quality of the tampered videos being generated and the ease of use ability of the online Deepfake generation and face-swapping applications implemented using deep learning techniques. Deep Learning is well known for its ability in handling complex and high dimensional data. Deep learning models such as encoder-decoder or auto encoders were earlier used and were widely used in the computer vision domain to solve several problems. But these auto encoders had the following disadvantages when compared to the recently used convolutional or neural networks [2]. First disadvantage is the Lack of temporal awareness which is the basic source of multiple abnormalities in the auto encoders. Because the auto encoder uses a frame-by-frame generation for the analysis of the Deepfake videos and was completely unaware of any previously generated face that it could have created automatically. Next is the inconsistencies existing with the face encoder i.e. the Encoder is unaware of the skin tone or other background information. It is very common to have boundary effects when combining the new face image with the rest of the frame. The third disadvantage is the visual inconsistency that exists due to the use of multiple cameras, different lighting conditions, or simply the use of different video codecs which make it tough for the auto encoder to create very accurate and realistic videos under different conditions. Finally, it is the inconsistency in choosing the illuminates between the different backgrounds with frames. This usually leads to blinking in the face region in most of the Deepfake videos. So, to overcome these disadvantages over the auto encoder and another deep learning technique like the convolutional neural network (CNN), Recurrent Neural Network (RNN), Generative Adversarial Network (GAN), etc. were developed.

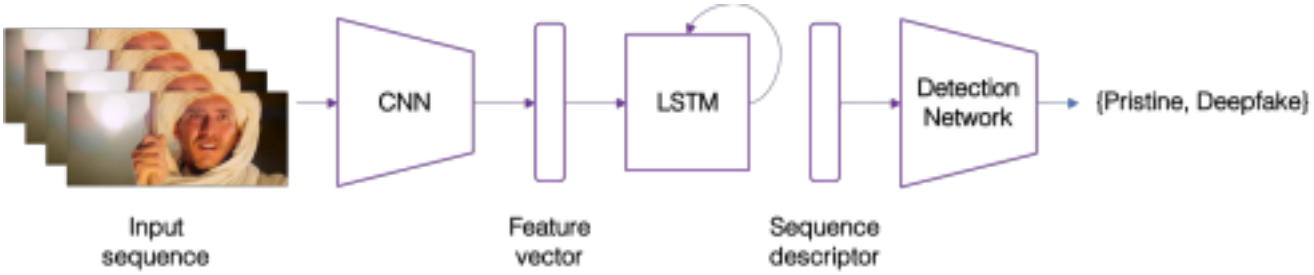
With the advancement in this convolutional net- work, there were many other schemes that developed in the creation and detection of Deepfake using Recurrent Neural Network (RNN), Long Short-Term Memory (LSTM), even the hybrid approaches of all the recent algorithm in the Deep Learning, etc. David Guera and Edward J Delp [2] bring up the first approach where the frame level features are extracted out after each processing in a convolutional neural network and these features are fed into the recurrent neural network as training samples and the output from this RNN is the classification result. Along with the combination of CNN and RNN, a set of encoder-decoders with shared weight for the encoder network is also used for dimensionality reduction and image compression in the training and generation phase and an LSTM network is used for the temporal sequential analysis. Another such approach was brought up by Ekraam Sabir in [3], the face manipulation detection using RNN strategies where they use a combination of variations in RNN models along with domain-specific face pre-processing techniques to obtain state-of-the-art performance on publicly available facial manipulation videos generated through FakeApp, Face2Face, and FaceSwap and this experimental evaluation shows an accuracy of 4.55%. A similar approach was stated in [4] for the swapped face detection using Deep Learning and Subjective Assessment. In this paper, they proposed a swapped face detection system which shows 96% positive result with few false alarms when compared with the other existing systems. Along with the detection of face-swapping, this model also evaluates the uncertainty in each prediction which is very much critical in the evaluation of the performance of a system. In order

to improve this predictability, they have set up a website to review the human response over the dataset by collecting pair to pair comparison of images over the videos on humans. Based on these comparisons, images are classified as real or fake. The output can be based on some kind of probability and this classification output is compared with the outputs from their automatic model which gives a very good but showing imperfect correspondence with linear correlations greater than 0.75. This experiment results show that this proposed model is much better when compared to the existing systems.

When it comes to the most advanced version of this deep neural network, a hybrid approach was implemented. A two-stream neural network was proposed in [5] where they train a GoogLeNet to detect the tampering artifacts like strong edges near lips, blurred areas on the forehead, etc. in image classification stream and in the second stream, a patch-based three-layer network is trained for capturing local noise components and camera characteristics. This network is designed to determine whether the obtained both patches come from the same image. It was found that the patches extracted out from the real samples around the face region seem very simple and have very small distance between them and while in the case of the tampered video, the patches from the face region will be different and will have a larger distance between them. Also, in the case of tampered video, the characteristics between the frames near the face region will be different when compared to the authentic or real one and here the classification based on these features are done by using an SVM classifier. They developed another dataset generated by two online face-swapping applications that consists of 2010 manipulated images, each of which contains a forged face for the performance evaluation. The experimental results show that this approach is able to learn both manipulated artifacts and hidden noise components.

Another concept in the hybrid model was pairwise learning [6] where a deep learning-based approach is used to identify manipulated images by combining the contrastive loss. First, the state-of-the-art GANs network will be used to generate a pair of fake and real images. Then, these pairs of image samples are fed into a common fake feature network (CFFN) to learn the distinguishing feature between the fake image and real image as a paired information. Then in the final stage, a small network will be used to combine these features to make the decision on whether it's fake or real. Experimental results show that the proposed method has high performance when compared to the existing state of the art image detection techniques. In the paper [7], a task-oriented GAN for PoISAR image classification and clustering techniques were used which consists of a Triplet network. Along with the generator and the discriminator, there is another network called the task network or T-net. The network in this proposed system basically has two task networks – one is called the classifier and another is called a clustered network. The first is the learning stage which has the two competing generator and discriminator networks which work hand in hand as in GANs network. In the second phase, the generator network is adjusted and oriented as a Task network where some samples from the training samples are assigned with a specific task that is the generation of the manipulated data. This takes up the advantage of a GAN network and also overcomes the disadvantage of the GAN network. After completing the learning phase, manipulated data are employed to take up the task to enhance the training sets and avoid overfitting among samples so that Task-Oriented GAN performs well even if the manual- labelled data are small. To verify the accuracy of the Task Network, a visual comparison is provided where some manipulated digits generated from Task-Oriented GAN in parallel with that from GAN as illustrated. The most important thing to be considered is that there is a greater difference between the PolSAR image dataset and this PolSAR image dataset used in this scheme as the image. The performance is evaluated through several experiments with this PolSAR image dataset and it shows that on three

PolSAR images, the proposed method shows high accuracy in dealing with PolSAR image classification and clustering.



HARDWARE REQUIREMENTS

- **A POWERFUL CPU**

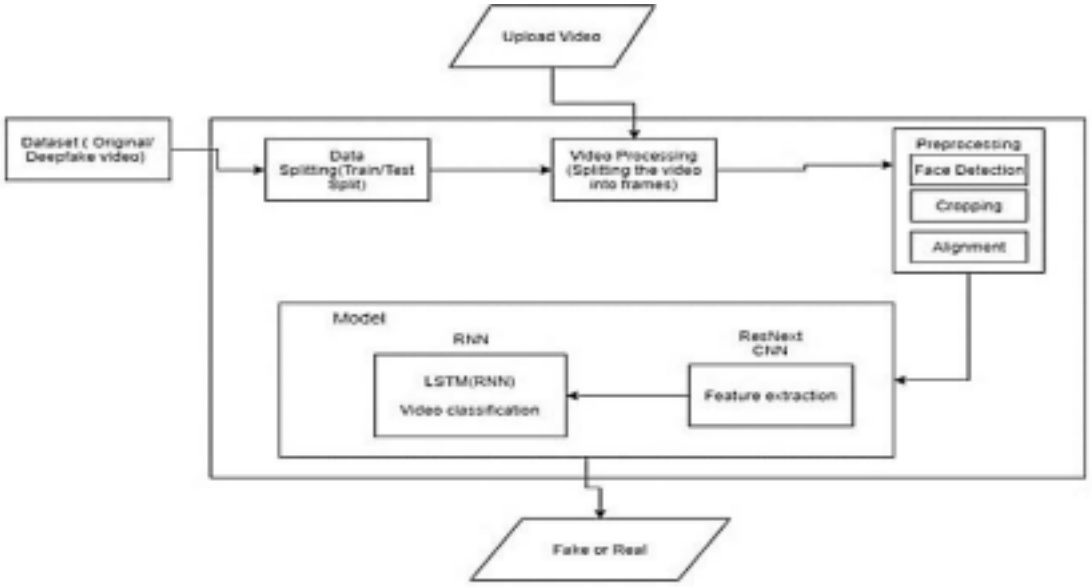
  • Laptop CPUs can often run the software, but will not be fast enough to train at reasonable speeds

- **A POWERFUL GPU**

  • Currently only Nvidia GPUs are supported. AMD graphics cards are not supported. This is not something that we have control over. It is a requirement of the Tensorflow library.

  • The GPU needs to support at least CUDA Compute Capability 3.0 or higher.

PROPOSED SYSTEM:-



Dataset: We are using a mixed dataset which consists of equal amount of photo from different dataset sources like YouTube, Deep fake detection challenge dataset. Our newly prepared dataset contains 50% of the original video and 50% of the manipulated deepfake videos. The dataset is split into 70% train and 30% test set.

Preprocessing: Dataset preprocessing includes the splitting the photo into frames. Followed by the face detection and cropping the frame with detected face. To maintain the uniformity in the number of frames the mean of the dataset video is calculated and the new processed face cropped dataset is created containing the frames equal to the mean. The frames that doesn't have faces in it are ignored during preprocessing. As processing the 10 second video at 30 frames per second i.e total 300 frames will require a lot of computational power. So for experimental purpose we are proposing to used only first 100 frames for training the model.

Model: The model consists of resnext50_32x4d followed by one LSTM layer. The Data Loader loads the preprocessed face cropped videos and split the videos into train and test set. Further the frames from the processed videos are passed to the model for training and testing in mini batches.

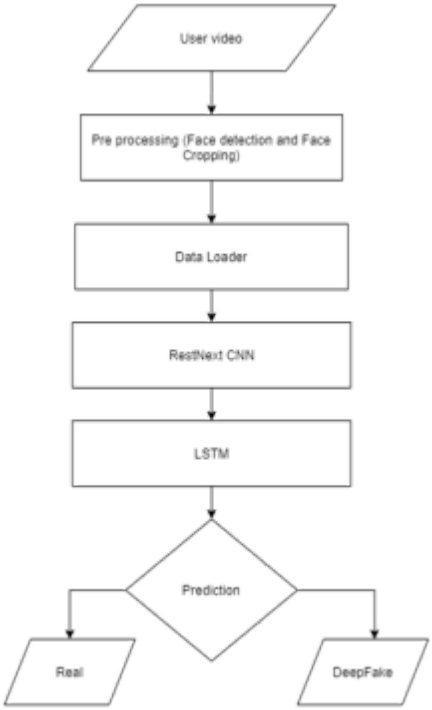
ResNext CNN for Feature Extraction: Instead of writing the rewriting the classifier, we are proposing to use the ResNext CNN classifier for extracting the features and accurately detecting the frame level features. Following, we will be fine-tuning the network by adding extra required layers and selecting 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.

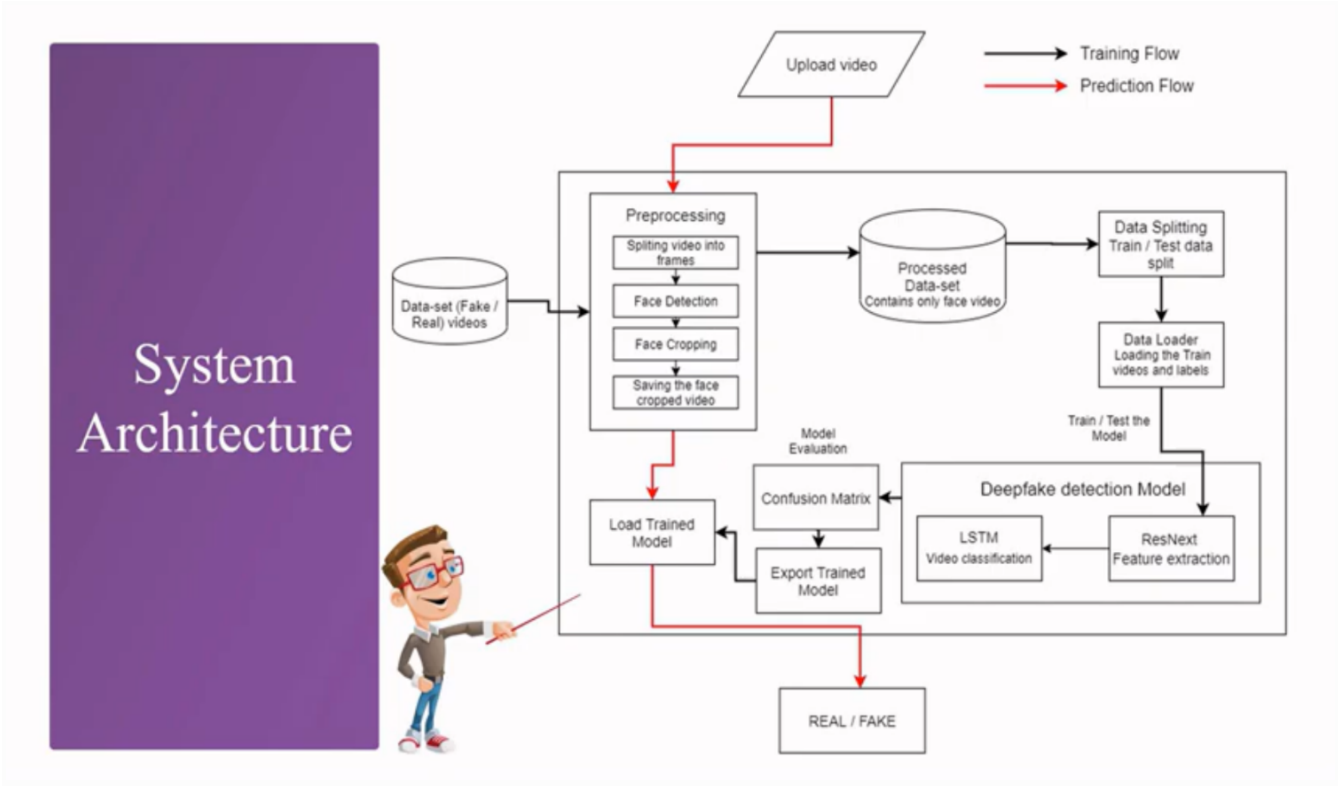
LSTM for Sequence Processing: Let us assume a sequence of ResNext CNN feature vectors of input frames as input and a 2-node neural network with the probabilities of the sequence being part of a deep fake video or an untampered video. The key challenge that we need to address is the de- sign of a model to recursively process a sequence in a meaningful manner. For this problem, we are proposing to the use of a 2048 LSTM unit with 0.4 chance of dropout, which is capable to do achieve our objective. LSTM is used to process the frames in a sequential manner so that the temporal analysis of the video 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.

Predict: A new video is passed to the trained model for prediction. A new video is also preprocessed to bring 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.

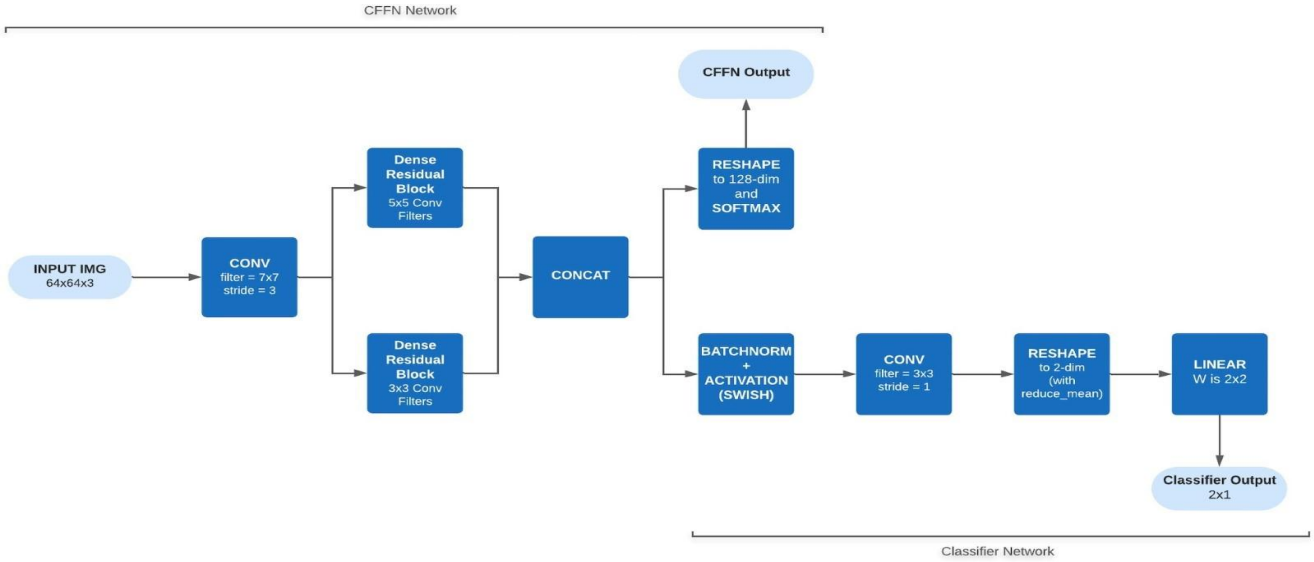
CHAPTER-3:- Functionality/Working of Project

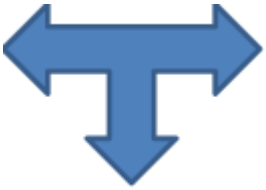




CHAPTER-4: Result & Conclusion

- 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

Deepfakes are hyper-realistic digitally manipulated videos of modern-day humans doing or pronouncing things they simply don’t. Mere visual verification is not enough to make a judgment on the veracity and also contemporary technologies to test if the footage has been altered aren't reliable. since the visible excellent modern day Deepfakes will soon come to be so flawless that it'll be difficult to make a judgment on veracity through mere visual verification. Digger’s solution for this is to apply state-of-the-art technology to broaden a toolkit that may locate the forgery in Deepfakes. So, it's very much essential to develop a system that could discover the forgery inside the Deepfake movies. Accordingly, destiny research can endorse a machine that may automatically detect the Deepfake motion pictures that use an audio-visible approach that detects the inconsistency that exists with lip movements and speech in audio. here, we also can practice several baseline methods which include simple principal component analysis (PCA) and linear discriminant evaluation

(LDA) techniques used for the extraction latest the characteristic vectors corresponding to the input video and the method primarily based on image high-quality metrics (IQM) and support vector machine (SVM) in CNN community may be used for the category modern day the video as real or faux based totally on the correlation between the characteristic vectors.

REFERENCES

[1] Raahat Devender Singh, Naveen Aggarwal," Video content authentication techniques: a comprehensive survey", Springer, Multimedia Systems, pp. 211- 240, 2018.

[2] David G'uera Edward J. Delp," Deepfake Video Detection Using Recurrent Neural Networks", Video and Image Processing Laboratory (VIPER), Purdue University,2018.

[3] Ekraam Sabir, Jiaxin Cheng, Ayush Jaiswal, Wael AbdAlmageed, Iacopo Masi, Prem Natarajan," Recurrent Convolutional Strategies for Face Manipulation Detection in Videos", In proceeding of the IEEE Xplore Final Publication, pp. 80-87, 2018.

[4] Xinyi Ding, Zohreh Razieiy, Eric C, Larson, Eli V, Olinick, Paul Krueger, Michael Hahsler," Swapped Face Detection using Deep Learning and Subjective Assessment", Research Gate, pp. 1-9, 2019.

[5] Peng Zhou, Xintong Han, Vlad I. Morariu Larry S. Davis," Two- Stream Neural Networks for Tampered Face Detection", IEEE Conference on Computer Vision and Pattern Recognition, 2019

[6] Chih-Chung Hsu, Yi-Xiu Zhuang, and Chia-Yen Lee," Deep Fake Image Detection based on Pairwise Learning", MDPI, AppliedScience,2020, doi:10.3390/app10010370.

[7] Fang Liu, Licheng Jiao, Fellow, IEEE, and Xu Tang, Member" Task-Oriented GAN for PolSAR Image Classification and Clustering", IEEE Transactions On Neural Networks and Learning Systems, Volume 30, Issue 9, 2019.

[8] B. Bayar and M. C. Stamm. A deep learning approach to universal image manipulation detection using a new convolutional layer. In Proceedings of the 4th ACM Workshop on Information Hiding and Multimedia Security, pages 5–10. ACM, 2016.

[9] F. Chollet. Xception: Deep learning with depthwise separable convolutions. arXiv preprint, pages

[10]D. Erhan, Y. Bengio, A. Courville, and P. Vincent. Visualizing higher-layer features of a deep network. University of Montreal, 1341(3):1, 2009.

[11] S. Fan, R. Wang, T.-T. Ng, C. Y.-C. Tan, J. S. Herberg, and B. L. Koenig. Human perception of visual realism for photo and computer-generated face images. ACM Transactions onApplied Perception (TAP), 11(2):7, 2014.

[12] H. Farid. A Survey Of Image Forgery Detection. IEEE Signal Processing Magazine, 26(2):26–25, 2009. 1 [9] P. Garrido, L. Valgaerts, O. Rehmsen, T. Thormahlen, P. Perez, and C. Theobalt. Automatic face reenactment. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 4217–4224, 2014.

[13] S. Ioffe and C. Szegedy. Batch normalization: Accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167, 2015.

[14] T. Julliand, V. Nozick, and H. Talbot. Image noise and digital image forensics. In Y.-Q. Shi, J. H. Kim, F. Pe´rez-Gonza´lez, and I. Echizen, editors, Digital-Forensics and Watermarking: 14th International Workshop (IWDW 2015), volume 9569, pages 3–17, Tokyo, Japan, October 2015.

[15] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980, 2014.

[16] A. Krizhevsky, I. Sutskever, and G. E. Hinton. Imagenet classification with deep convolutional neural networks. In Advances in neural information processing systems, pages 1097–1105, 2012