**PROJECT SYNOPSIS**

**ON**

**Deepfake Video Detection using Neural Network**

**Submitted to**

**Rashtrasant Tukadoji Maharaj Nagpur University,**

**NAGPUR**

In partial fulfillment of the requirement of

**M.Sc. Semester - IV (Computer Science) Examination**

*Submitted by*

**Pranay V. Shahare**

**&**

**Tejaswini V. Hirudkar**

**Under the guidance of**

**Dr. S. R. Gedam**

**Asst. / Professor & Head**

**(Department of Computer Science)**



**DEPARTMENT OF COMPUTER SCIENCE**

**Shri Shivaji Education Society Amravati’s**

**SCIENCE COLLEGE**

**Congress Nagar, Nagpur-12.**

**2021-2022**

**Table Of Contents**

|  |  |  |
| --- | --- | --- |
| **Sr. No.** | **Topic** | **Page No.** |
| **1.** | **Introduction** | 1 |
| **2.** | **Objectives** | 2 |
| **3.** | **Literature Survey** | 2 |
| **4.** | **Methodology** | 3 |
| **5.** | **Plane of Research** | 6 |
| **6.** | **Implications** | 6 |
| **8.** | **References** | 7 |

**Introduction:-**

The term “Deepfake” is a combination of two words “Deep Learning” and “Fake”. Deepfake are a type of algorithm based on deep learning and Artificial Intelligence which can harness fake videos and images. Deepfakes can be created by traditional visual effects or computer-graphics approaches. Creation of deep fake videos are a simple task, but when it comes to detection, it’s a major challenge.

The first deepfake video emerged in 2017, which was created by Chris Ume using visual and AI effects. The idea behind deepfake videos is to make fake videos look realistic with the help of AI but this technology has mostly received negative criticism due to the dangerous possibilities it offers. The back bone of deep fake are deep neural networks trained on facial images to automatically map the expression from a source to a target. In this work, we will effectively distinguish fake videos from real videos. We will focus only on facial manipulations in the video. We won’t be discussing the audio manipulations in the deep fake video. We emphasize that deepfake should not be confused with adversarial machine learning which is to fool the machine learning algorithms.

Most Deep Fakes on the Internet include pornographic images of men, usually by female celebrities such as their often used without their consent. Extraordinary pornography is being released surfing the Internet in 2017, especially Reddit. Deep Fake is also used to misrepresent famous politicians. In separate videos, Argentine President Mauricio Macri's face has been replaced by Adolf Hitler's face, and Angela Merkel's face has been replaced by Donald Trump's. The first known attempt to make a face-to-face exchange was seenin the photograph of Abraham Lincoln. The lithography superimposes his head with the body of John Calhoun. The engravings of his head on other bodies appeared quite often after his assassination. [1]

In particular, a new vein of AI-based fake video generation methods known as DeepFake has attracted a lot of attention recently. It takes as input a video of a specific individual (’target’), and outputs another video with the target’s faces replaced with those of another individual (’source’). The backbone of DeepFake are deep neural networks trained on face images to automatically map the facial expressions of the source to the target. With proper post-processing, the resulting videos can achieve a high level of realism. [2]

The democratization of creating realistic digital humans has positive implications, there is also positive use of deep fakes such as their applications in visual effects, digital avatars, snapchat filters, creating voices of those who have lost theirs or updating episodes of movies without reshooting them. [3]

Generative Adversarial Network (GAN) creates the DF. Generative adversarial network takes as input a video and an image of a specific individual (‘target’), and outputs another video with the target’s faces replaced with those of another individual (‘source’). The backbone of DeepFakes are deep adversarial neural networks trained on face images and target videos to automatically map the faces and facial expressions of the source to the target. With proper post processing, the resulting videos can have a high level of real like effect. The generative adversarial network splits the video into frames and replaces the input image in every frame. Further it reconstructs the video. This process is usually done by using autoencoders.

**Objectives:-**

1. To create deep fake video using available applications (Zao, Deep Art, Wombo).
2. To detect a deep fake video using neural network.

**Literature Survey:-**

Abdul Jamsheed V., Janet B., “Generative adversarial networks progressed to the point where it is very difficult to distinguish the difference between what is real or fake. In recent times, different face manipulating tools are used to generate credible face swaps in videos that leave a very little trace of manipulation, which is commonly referred as AI based Deep Fake videos”.[5]

T. T. Nguyen and D. T. Nguyen, “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.”[3]

Pavel Korshunov and Sebastien Marcel, “It is becoming increasingly easy to automatically replace a face of one person in a video with the face of another person by using a pre-trained generative adversarial network (GAN). The first publicly available set of Deepfake videos generated from videos of Texas Instruments and Massachusetts Institute of Technology(VidTIMIT) database.”[6]

Arti karandikar, Vedita Deshpande, Sanjana Singh, Sayali Nagbhidkar “With the advent of new technological enhancements in artificial intelligence, new sophisticated AI techniques are used to create fake videos. Such videos can pose a great threat to the society in various social and political ways and can be used for malicious purposes. These fake videos are called deepfakes. Deepfakes refer to manipulated videos, or other digital representations produced by sophisticated artificial intelligence, that yield fabricated images and sounds that appear to be real. A deep-learning system can produce a persuasive counterfeit by studying photographs and videos of a target person from multiple angles, and then mimicking its behaviour and speech patterns.”[4]

Akhil Sunil Kumar, Amruta Khavase, Himesh Rajendran “Creating the DeepFakes using Artificially intelligent tools are simple tasks. But, when it comes to detection of these DF, it is a major challenge. Because training the algorithm to spot the DeepFake is not simple. We have taken a step forward in detecting the DeepFakes using Convolutional Neural Network and Recurrent neural Network. System uses a convolutional Neural network to extract features at the frame level. These features are used to train a recurrent neural network which learns to classify if a video is manipulated or not and able to detect the temporary inconsistencies between frames introduced by the DeepFake creation tools. Expected result against a large set of fake videos collected from standard data sets.” [7]

**Methdology:-**

As deep fakes are spreading faster than expected and creating very serious issues; it’s necessary to have automatic tools and technologies to detect fake content. Social media channels are one of the widely used platforms where such contents are dumped without any verification. We are focusing mainly on the detection of deep fake videos from real ones. The model focuses on detecting all types of deep fakes such as Replacement Deep Fake, Interpersonal Deep Fakes and Retrenchment Deep Fakes.

**Flow of methodology to be applied:-**

Deep Fake Creation

Architecture

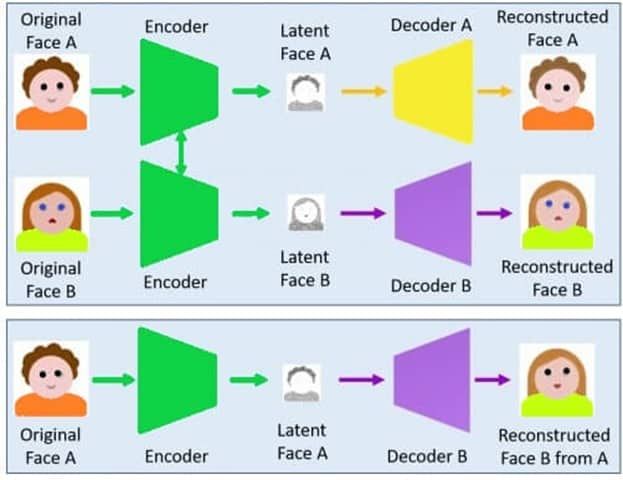
Dataset

Pre-processing

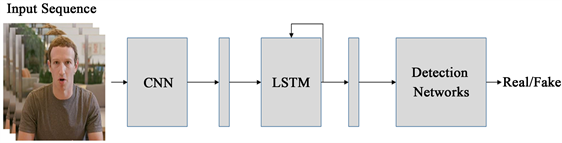
CNN

LSTM

1. **Deep Fake creation:-** The creation of a DeepFake video starts with an input video of a specific individual (’target’), and generates another video with the target’s faces replaced with that of another individual (’source’), based on a GAN model trained to translate between the faces of the target and the source.



1. **Architecture:-** Proposed model will effectively detect manipulated videos from real videos. Our system uses convolutional neural networks (CNN) to extract frame level features. These Features are then trained using a Long Short Term Memory (LSTM) network that classifies real and fake videos separately. Architecture of the system is given in figure 3. This novel deep fake detection model has 3 sections, first the data pre-processing stage, then feature extraction using CNN and finally LSTM network is used to classify if the video is fake or not.



1. **Dataset:-** We built a custom dataset by combining different data from multiple sources. Majority of the videos are taken from Deep Fake Detection challenge (DFDC) repository which is an open source data contributed by tech giants like amazon, Facebook, Microsoft etc. 3000 plus videos are collected from DFDC and around 2000 plus videos from Face Forensics++ where manipulations of the video are done by latest video manipulation software like deepfake, Face2Face, Face Swap and Neural Textures. A good amount of videos are contributed by Caleb-DF repository which contains manipulated videos of celebrities with different age, ethnic group and gender. These videos are later split into train and test data with ratio of 70:30.

1. **Pre-Processing:-** Since our work only concentrates on the face manipulations on the video, we need proper pre-processing of data to avoid unnecessary computation. Video is split into frames and then faces in the videos are recognized using facial recognition model. Frames which doesn’t have faces are removed in the pre-processing stage. To maintain uniformity, the first 300 frames are cropped from every video. Detecting faces from video is a difficult task. The system demands every feature of the face to properly diagnose if any manipulations have been done to the video. We use special haarcascade facial features to capture the faces from the video. Combining the haarcascade feature vector and openCV facial recognition module we can detect faces. This slightly modified function can extract face, profile face, eyes and smiles. For each object we are using different shapes and colours to distinguish between the features as shown in Figure 4. Green rectangle for frontal face, red circle for eye, red rectangle for smile and blue rectangle for profile face are used for this purpose.

1. **Convolutional Neural Network:-** Convolutional neural networks are widely used for image recognition and classification problems. Pre-processing required for CNNs are very less when compared to other deep learning models and it can extract more features from a video or image. So the deep learning models also require a technology which can extract all the useful features from the videos before feeding them into a classification algorithm. Convolution operation preserves spatial relationship between the pixels while extracting the useful features from the video. Pre-trained ResNeXt-50 CNN serves this purpose. We will fine tune the model by using hyper parameters and adding extra layers the process is repeated until the model converges to global optimum.

1. **LSTM (Long Short Term Memory) for sequence processing:-** Unlike recent works, we use the LSTM version of recurrent neural networks for classification. Features which are extracted using convolutional neural networks are fed into LSTM network for classifying a video as fake or real. LSTM’s are very much capable of learning long term dependencies. Chain like structure of LSTM decides what information we are going to pass in each stage and helps to keep the dependencies for long term. This way the model learns the manipulations made in the videos throughout the duration and identifies if the video has been subjected to any type of manipulation. Key challenge in LSTM that we need to address is the design flow to process the sequence of frames in a meaningful manner. LSTMs are very helpful in locating the spatial and temporal difference in the videos.

**Plan of Research:-**

* Literature Survey.
* Study of related work on deep fake video creation and detection techniques.
* Study of methodology to solve the problem of deep fake video.
* Perform the experiments to detect the deep fake videos using neural network technique.
* Conclude the result of deep fake video detection using neural network.

**Implications:-**

We can represent a neural network based solution for detecting deep fake videos. This model is when trained using neural network can give quiet impressive results compared to other deep learning models. Creation of deep fake videos is a simple task but when it comes to the detection, it’s a major challenge. The advancement in creation of AI based deep fake videos has made the older detection system less accurate. But in this work, we can describe a new method which use deep learning based methodology to effectively detect manipulated or fake videos from real videos. We can be successful in obtaining a competitive result of maximum number of percent accuracy using a simple architecture. As the technology behind the deep fake keeps evolving, we will continuing improve the detection method.

**References:-**

1. Anuj Badale, Chaitanya Darekar, Lionel Castelino, Joanne Gomes, “Deepfake Video Detection Using Neural Netwok”, International Journal of Engineering Research and Technology, ISSN: 2278-0181, NTASU 2020, (April 2021)
2. Yuezn Li, Siwei Lyu, “ Exposing Deepfake Videos By Detecting Face Warping Artifacts”, Institute of Electrical and Electronics Engineers, arXiv: 1811.00656v3 [cs.CV] (May 2019)
3. T.T. Nguyen, Q. V. H. nguyen, C. M. Nguyen, D. T. Nguyen, Saied Nahavandi, “ Deep Learning for Deepfakes creation and Detection”, Institute of Electrical and Electronics Engineers, arXiv: 1909.11573v3 [cs.CV] (April 2021)
4. Aarti Karandikar, Vedita Deshpande, Sanjana Singh, Sayali Nagbhidkar, Saurabh Agrawal, “ Deepfake Video Detection Using Convolutional Neural Network” , International Journal of Advanced trends in computer Science and Engineering, ISSN 2278-3091, Volume 9 No. 2 (April 2020)
5. Abdul Jamshhed V. Janet B. , “ Deep Fake Video detection Using Recurrent Neural Network” , International Journal of Scientific Research in Computer Science and Engineering, E-ISSN: 2320-7639, Volume 9, Issue:2 (April 2021)
6. Pavel Korshunov, Sebastien Marcel, “Deepfakes A New Threat to face recognition? Assessment And Detection” , arXiv:1812.08685v1 [cs.CV] (December 2020)
7. Akhil Sunil Kumar, Amruta Khavase, Himesh Rajendran, “ Deepfake Video Detection Using Neural Nnetwork”, International Journnal of Innovative research in technology, ISSN; 2349-6002, Volume 7, Issue 12, (May 2021)
8. Ruben t. Ruben Vera-Rodriguez, Julian Fierrez, Aythami Morales, Javier Ortega-Garcia, “ DeepFake and Beyond: A Survey of Face Manipulation and Fake Detection” , arXiv: 2001.00179v3 [cs.CV] (June 2020)
9. Deressa Wodajo, Solomon Atnafu, “ Deepfake Video detection Using Convolutional Vision Transformer”, Institute of Electrical and Electronics Engineering, arXiv: 2102.11126v3 [cs.CV] (March 2021)
10. Brian dalhansky, russ Howes, Ben Pflaum, Nicole Baram, Cristian Canton Ferrar, “ The Deepfake zdetection Challenge Preview Dataset” , Institute of Electrical and Electronics Engineering, arXiv: 1910.08854v2 [cs.CV] (23 October 2019)