DEEPFAKE DETECTION USING

DEEP LEARNING

A PROJECT REPORT

Submitted by

GURUNATHAN M

(2019202015)

submitted to the Faculty of

INFORMATION SCIENCE AND TECHNOLOGY

in partial fulfillment for the award of the degree

of

MASTER OF COMPUTER APPLICATIONS

DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY

COLLEGE OF ENGINEERING, GUINDY

ANNA UNIVERSITYCHENNAI

600 025

MAY 2022

ANNA UNIVERSITY

CHENNAI - 600 025

BONAFIDE CERTIFICATE

Certified that this project report titled **"DEEPFAKE DETECTION USING DEEP LEARNING"** is the bonafide work of **GURUNATHAN M(2019202015)** who carried out project work under my supervision. Certified further that to the best of my knowledge and belief, the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or an award was conferred on an earlier occasion on this or any other candidate.

PLACE:CHENNAI **DR.L.SAIRAMESH**

DATE : 02.05.22 **PROJECT GUIDE**

DEPARTMENT OF IST, CEG

ANNAUNIVERSITY 600025

Dr.S. SRIDHAR

HEAD OF THE DEPARTMENT

DEPARTMENT OF INFORMATION SCIENCE AND TECHNOLOGY

COLLEGE OF ENGINEERING, GUINDY

ANNA UNIVERSITY

CHENNAI 600025

TABLE OF CONTENTS

CHAPTER 4: IMPLEMENTATION

4.1 PREPROCESSING

REFERENCES

LIST OF FIGURES

LIST OF ABBREVIATION

CNN	Convolutional Neural Networks
LSTM	Long Short Term Memory
CSV	Comma Separated Value
ResNeXt-50	Residual Network

ABSTRACT

With the recent developments on the creation of deepfake videos using Generative Adversarial Network (GAN), which can produce realistic photos and videos, the reliability of digital images is becoming more challenging to identify. This research is an approach to develop a deep learning model which can efficiently distinguish between a deepfake and a real video. Research work on transfer learning of computer vision to use the previously build features of the neural network of image categorization and build a new model over it. Deep learning is continuously evolving a lot in both areas of generating and detecting deepfakes. A model developed for detection of deepfake designed with older dataset may expire in time, and a need for new detection technique will always be there.

This system uses a convolutional Neural network (CNN) to extract features at the frame level. These features are used to train a Long Short Term Memory (LSTM) which learns to classify if a video has been subject to manipulation or not and able to detect the temporal inconsistencies between frames introduced by the DF creation tools. Expected result against a large set of fake videos collected from standard data set Result of the research is auspicious with more than 90% accuracy and the area of evolvement and advancement.

CHAPTER 1

INTRODUCTION

1.1 DEEPLEARNING TECHNIQUES

Deep learning is the subset of machine learning composed of algorithms that permits the machine to train itself and perform a task. These algorithms uses multiple layers to progressively extract the high level feature from the raw input(images, audio, video). In deep learning, each level learns to transform its input data into a slightly more abstract and composite representation. Most modern deep learning models are based on convolutional neural networks (CNNs).

A Convolutional Neural Network (CNN) is an algorithm which can take in an input image, assign importance (learnable weights and biases) tovarious aspects/objects in the image and be able to differentiate one from the other. The pre-processing required in a CNN is much lower as compared to other classification algorithms. It is able to successfully capture the Temporal dependencies in an image through the application of relevant filters. The role of the CNN is to reduce the images into a form which is easier to process, without losing features which are critical for getting a good prediction.

1.3 MOTIVATION AND OBJECTIVES

DeepFakes involves videos, often obscene, in which a face can be swapped with someone else's using neural networks. DeepFakes are a general public concern. As soon as fake videos go viral, people believe them initially and keep sharing them with others. This makes the targeted person embarrassed. Thus it's important to develop methods to detect them.

The creation of a synthesized video needs information to be examined in order to reveal the dishonest, particularly facial expression variables. It is defined as an in-depth

video study to find minor imperfections such as boundary points, background incoherence, double eyebrows or irregular twitch of the eye. The impetus behind this research is to recognize these distorted media, which is technically demanding and which is rapidly evolving. Many engineering companies have come together to unite a great deal of dataset. Competitions and actively include data sets to counter deepfakes.

Deepfake videos are now so popular that multiple political parties utilize this tool to produce faked images of the leader of their opposing party to propagate hate against them. Fake political videos telling or doing things that have never happened is a threat to election campaigns. These images are the primary source of false media controversies and propagate misleading news. In order to expose the forgery in extremely detailed facial expression, such details to be investigated frame by frame in the production of a deepfake picture. The goal of this research is to create a deep learning model that is capable of recognizing deepfake images. The model will learn what features differentiate a real image from a deepfake.

1.4 SCOPE OF THE PROJECT

The scope of the project is to avoid false media controversies, propagate misleading news and fake political videos telling things that have never happened is a threat to election campaigns.

1.5 ORGANIZATION OF THE REPORT

The thesis is organized into 6 chapters, describing each part of the project with detailed illustration and system design diagrams. The chapter are as follows:

Chapter 1 : This module consists of Introduction, Problem statement, Motivation and Objectives etc.

Chapter 2 : This module consists of Literature survey details of the project alongside their detailed methodologies, advantages, disadvantages etc.

Chapter 3 : This module consist System design of the project with its preliminary design such as overall Architecture diagram and process flow diagram which tells about the modules integration in the project.

Chapter 4 : This module consists of Detailed system design or module description with their input and algorithmic steps involved in each module to derive the output as per the user requirement.

Chapter 5 : This module consists the details about the hardware and software requirement to the project and the experiments that has been performed along with their outcomes. The detailed result of the project is also portrayed in this chapter.

Chapter 6 : This module conclude the project report with all the results and implementation procedure that has been underwent during the project development. The future works and excellence of implemented project is detailed.

The above mentioned six modules are followed up with the Reference which deliberately explains and list all the reference documents used during the various phases of the project, which includes the journal papers, conference papers, white papers, articles and websites referred for tutorials.

CHAPTER 2

LITERATURE REVIEW

This Chapter explains about the literature survey made on the existing system, analyzing the problem statements and issues with the existing system and proposed objectives for the new system.

2.1 EYE BLINKING DETECTION

Exposing AI Created Fake Videos by Detecting Eye Blinking describes a new method to expose fake face videos generated with deep neural network models. The method is based on detection of eye blinking[1] in the videos, which is a physiological signal that is not well presented in the synthesized fake videos. The method is evaluated over benchmarks of eye-blinking detection datasets and shows Deepfake Video Detection using Neural Networks.

Drutarovsky et al. [2] analyzed the variance of the vertical motions of eye region which is detected by a Viola-Jones type algorithm. Then a flock of KLT trackers are used on the eye region. Each eye region is divided into 3x3 cells and an average motion in each cell is calculated. This system proposed a scalar quantity that measures the aspect ratio of the rectangular bounding box of an eye corresponding to the eye openness degree in each frame. They then trained an SVM of EARs within a short time window to classify final eye state. Kim et al. [3] studied CNN-based classifiers to detect eye open and close state.

2.2 FACE BASED VIDEO MANIPULATION

Multiple approaches that target face manipulations in video sequences have been proposed since the 1990s [4]. This demonstrated the first real-time expression transfer for faces and later proposed Face2Face, a real-time facial reenactment system, capable of altering facial movements in different types of video streams.

Several face image synthesis techniques using deep learning have also been explored as surveyed by Generative adversarial networks (GANs) are used for aging alterations to faces [5], or to alter face attributes such as skin color. Deep feature interpolation shows remarkable results in altering face attributes such as age, facial hair or mouth expressions. Most of these deep learning based image synthesis techniques suffer from low image resolution. Karras et al. [6] show high quality synthesis of faces, improving the image quality using Generative Adversarial Network(GAN).

2.3 DETECTION USING MESONET

This section aims to discuss and analyze various techniques that have been used for deepfake detection. Various attempts have been made to detect deepfakes which use deep learning at their core. These approaches work either on detecting faults in video or in separate frames of the video. Approaches which involve image analysis target various parameters like face warping artifacts, eye blinking [2]. In 2018, “MesoNet” [7] was developed which used Inception model to detect faults at mesoscopic level. Convolutional Neural Networks (CNN) have shown excellent feature extraction properties that can be used by a model to detect deepfake videos.

Various other approaches have used CNN along with other learning models like Recurrent Neural Network (RNN), Long Short-Term Memory Networks (LSTM) and Capsule Network[8] to further improve the accuracy by detecting temporal discrepancies and have shown good results on dataset containing videos generated by FaceSwap and Deepfacelab.

2.4 PROPOSED SYSTEM

This approach for detecting the DF will be great contribution in avoiding the problems of the DF over the world wide web. It will be a web-based platform for the user to upload the video and classify it as fake or real. This project can be scaled up from developing a web-based platform to a browser plugin for automatic DF detections. Even big application like WhatsApp, Facebook can integrate this project with their application for easy pre detection of DF before sending to another user. The important objective is to evaluate its performance and acceptability in terms of security, user-friendliness, accuracy and reliability.

CHAPTER 3

SYSTEM DESIGN

3.1 SYSTEM ARCHITECTURE

The proposed work of the system architecture is shown in Fig 3.1. The proposed system works on DeepFake detection. Celebrity dataset are labeled to the defined classes. Then they are loaded to the pre-defined model for training the dataset. In deepfake detection, ResNext50 is chosen to trained the dataset. Also use LSTM to enrich the accuracy gives the checkpoint file. At last the user video tested with the checkpoint file to detect fake or real.

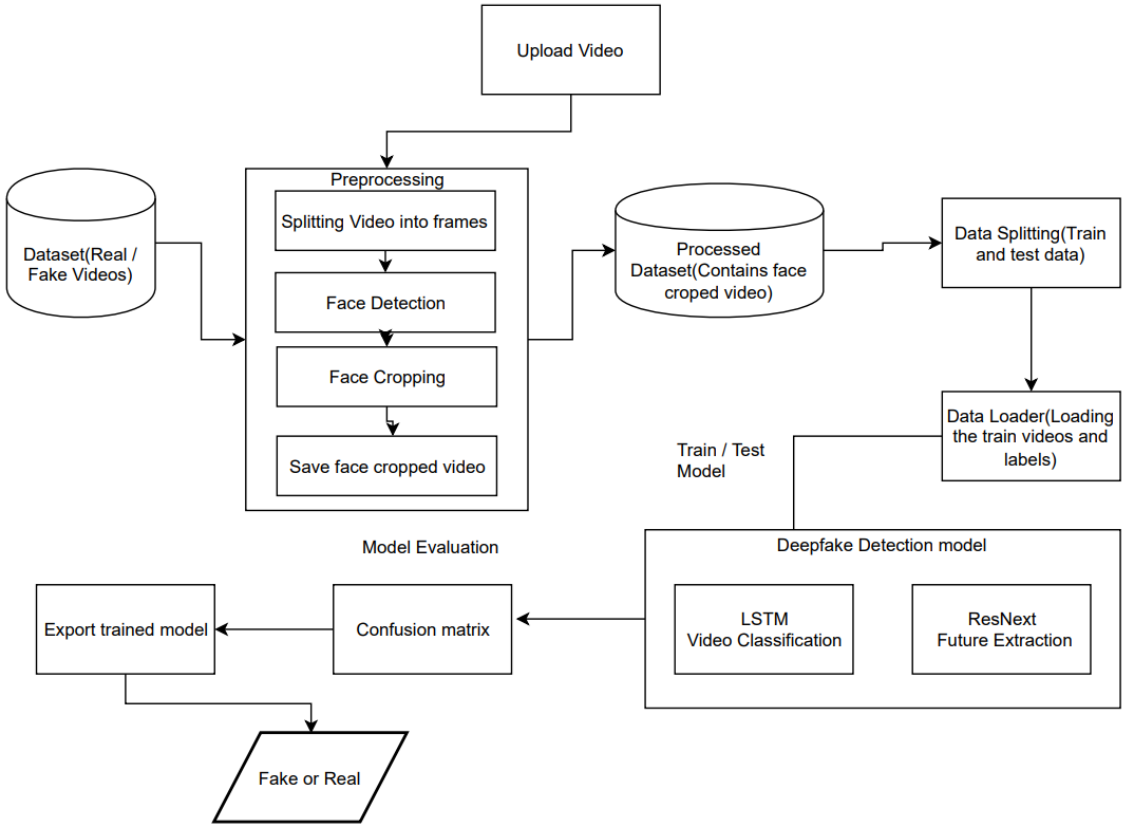

Fig 3.1 SYSTEM ARCHITECTURE

3.1.1 TRAINING DATASET

For DeepFake detection, Celebrity video datasets are collected from theGitHub. This contains 50% real and 50% fake videos.

3.1.2 DATASET PREPROCESSING

Dataset preprocessing includes the splitting the video into frames. Followed by the face detection and cropping the frame with detected face.

3.1.3 EXTRACT FEATURE

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.

3.1.4 LSTM

The output of the ResNext are fed into the Long Short Term Memory (LSTM) cell which is supposed to extract global temporal features entire sequences of data. 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.

3.1.5 FULLY CONNECTED LAYER

The outputs of the LSTM cell are classified by a fully-connected layer which contained two neurons that represent the two categories (fake and real). The final result is in the form of checkpoint file which helpsto detect the input video sequence as fake or real.

3.2 FLOWCHART DESIGN

This is the Flowchart design of the proposed system. The Flowchart design of the prediction workflow is shown in Fig.3.2.

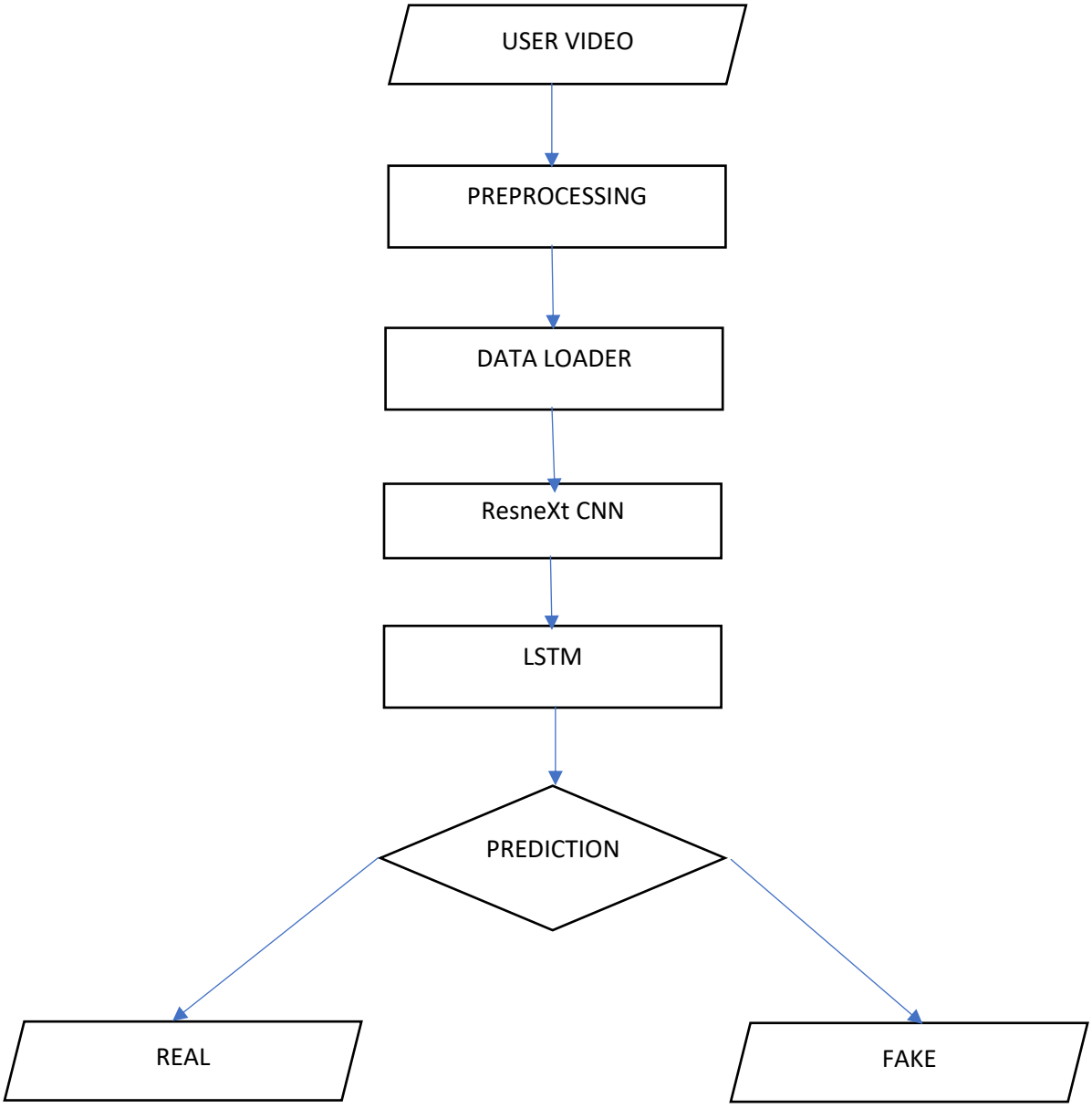

Fig:3.2 PREDICTION FLOW

3.3 ALGORITHM DESIGN

In the proposed system the detection of DeepFake are based on the pre-trained model. Here Faster ResNext model used for and detection.

3.3.1 ResNeXt-50

The ResNeXt architecture is an extension of the deep residual network which replaces the standard residual block with one that leverages a “split-transform-merge” strategy used in the Inception models. The ResNeXt architecture is shown in Fig.3.3

The ResNext refers to the number of branches or groups as the cardinality of the ResNeXt cell and performs a series of experiments to understand relative performance Gains between increasing the cardinality, depth, and width of the network. The experiments show that increasing cardinality is more effective at benefiting model performance than increasing the width or depth of the network. The experiments also suggest that “residual connections ae helpful for optimization”, whereas aggregated transformations are strong representations.

stage	output	**ResNeXt-50 (32×4d)**
conv1	112×112	7×7, 64, stride 2
conv2	56×56	3×3 max pool, stride 2 [1×1, 128 3×3, 128, C=32 1×1, 256] ×3
conv3	28×28	[1×1, 256 3×3, 256, C=32 1×1, 512] ×4
conv4	14×14	[1×1, 512 3×3, 512, C=32 1×1, 1024] ×6
conv5	7×7	[1×1, 1024 3×3, 1024, C=32 1×1, 2048] ×3
	1×1	global average pool 1000-d fc, softmax

Fig 3.3 ResNext50 Architecture

CHAPTER 4 – IMPLEMENTATION

4.1 PREPROCESSING

Dataset preprocessing includes the splitting the video 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. The preprocessing workflow is shown in Fig.4.1.

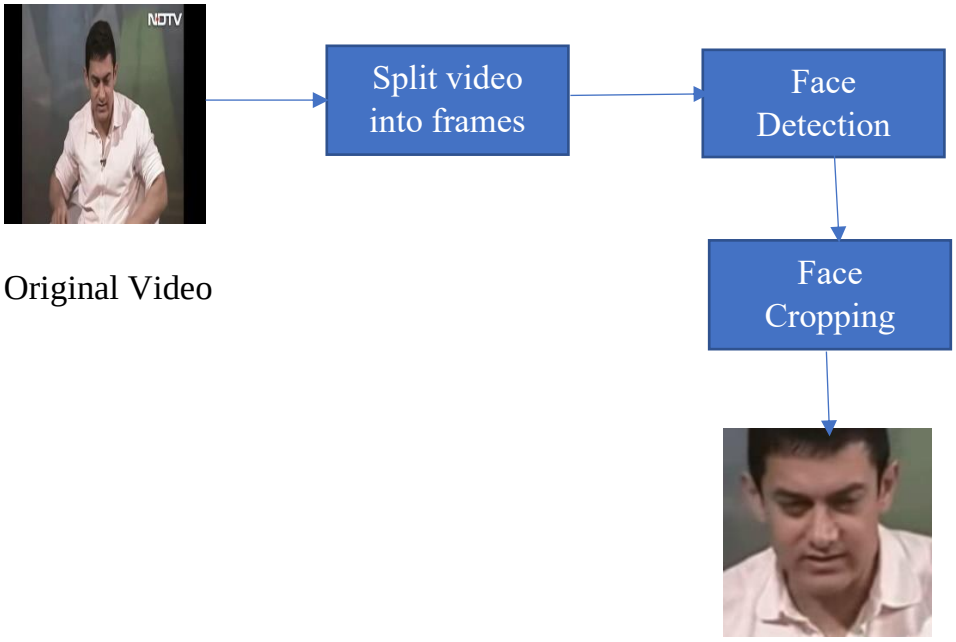

Fig 4.1 Preprocessing

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.

REFERENCES

[1] Yuezun Li, Ming-Ching Chang and Siwei Lyu “Exposing AI Created Fake Videos by Detecting Eye Blinking” in arxiv.

[2] T. Drutarovsky and A. Fogelton. Eye blink detection using variance of motion vectors. In ECCV, pages 436–448, 2014.

[3] K. W. Kim, H. G. Hong, G. P. Nam, and K. R. Park. A study of deep cnn-based classification of open and closed eyes using a visible light camera sensor. Sensors, 17(7):1534, 2017

[4] David Guera and Edward J.Delp. Deepfake Video Detection Using Recurrent Neural Network. Video and Image processing Laboratory, Purdue University, In Ieeexplore, 2018

[5] G. Antipov, M. Baccouche, and J.-L. Dugelay. Face aging with conditional generative adversarial networks. arXiv:1702.01983, Feb. 2017

[6] T. Karras, T. Aila, S. Laine, and J. Lehtinen. Progressive growing of gans for improved quality, stability, and variation. arXiv:1710.10196, Oct. 2017.

[7] DariusAfchar, Vincent Nozick, Junichi Yamagishi and Isao Echizen, “MesoNet: a Compact Facial Video Forgery Detection Network”, arXiv:1809.00888v1, 4 Sep 2018.

[8] Huy H. Nguyen, Junichi Yamagishi, and Isao Echizen, “Use of a capsule network to detect fake images and videos”, arXiv:1910.12467v2 [cs.CV] 29 Oct 2019.