Facial Forgery Detection

Thesis Submitted By:
Haider Ali Khan
01-243172-043

Supervised By:
Dr. Samabia Tahsin

A dissertation submitted to the Department of Computer Science, Bahria University, Islamabad as a partial fulfillment of the requirements for the award of the degree of Masters in Computer Science

Session (2017-2019)

Bahria University
Discovering Knowledge

MS-13

Thesis Completion Certificate

Scholar's Name: Haider Ali Khan Registration No. 01-243172-043

Programme of Study: MSCS

Thesis Title: Facial Forgery Detection

It is to certify that the above student's thesis has been completed to my satisfaction and, to my belief, its standard is appropriate for submission for Evaluation. I have also conducted plagiarism test of this thesis using HEC prescribed software and found similarity index at _____ that is within the permissible limit set by the HEC for the MS/MPhil degree thesis.
I have also found the thesis in a format recognized by the BU for the MS/MPhil thesis.

Principal Supervisor's Signature: _____

Date: 02-07-2019 **Name:** Dr. Samabia Tehsin

Bahria University
Discovering Knowledge

MS-14A

Author's Declaration

I, Haider Ali Khan hereby state that my MS thesis titled "Facial Forgery Detection"

is my own work and has not been submitted previously by me for taking any degree from this university

Bahria University Islamabad

or anywhere else in the country/world.

At any time if my statement is found to be incorrect even after my Graduate the university has the right to withdraw/cancel my PhD degree.

Name of scholar: Haider Ali Khan
Date: 25-06-2019

Bahria University
Discovering Knowledge

MS-14B

Plagiarism Undertaking

I, solemnly declare that research work presented in the thesis titled "Facial Forgery Detection" is solely my research work with no significant contribution from any other person. Small contribution / help wherever taken has been duly acknowledged and that complete thesis has been written by me.

I understand the zero tolerance policy of the HEC and Bahria University towards plagiarism. Therefore I as an Author of the above titled thesis declare that no portion of my thesis has been plagiarized and any material used as reference is properly referred / cited.

I undertake that if I am found guilty of any formal plagiarism in the above titled thesis even after award of PhD degree, the university reserves the right to withdraw / revoke my PhD degree and that HEC and the University has the right to publish my name on the HEC / University website on which names of students are placed who submitted plagiarized thesis.

Student / Author's Sign: 

Name of the Student: Haider Ali Khan

Abstract

In the recent past, fields of computer vision and graphics have become more advanced and it is now easy to generate realistic fake videos and images. At present facial forgery techniques, such as Deepfake and Face2Face are very popular. Such forged videos and images can be used for fake news and to deceive biometric recognition systems. Detection of facial forgery in images and videos is a very complex and challenging task. For the last years, different methods have been employed to detect such forgeries. However such facial forgeries are still challenging to detect. In this paper, we proposed a facial forgery detection solution which automatically detects facial forgeries in videos and images. The proposed technique will classify fake and original videos and images. To classify the images we used Inception-ResNet, a deep neural network. The proposed technique is validated on publicly available Deepfake TIMIT dataset and reported very effective results.

Acknowledgments

All praises to the ALLAH Subahanhu-Wa-Taa-Ala, for giving me the knowledge and ability to achieve this milestone - MS thesis. Without his blessings completion of this research was not possible.
I am very thankful to my supervisor, Dr. Samabia Tehsin for guiding me to accomplish this goal.
I would like to thank my friends Mr. Shahbaz Hassan , Syed Abdul Basit and Mr. Aaqib Mehran for their encouragement and moral support throughout my MS work.
My acknowledgment would be incomplete without expressing my gratitude to my parents and family members, the prime source of my strength for always believing in me and encouraging me to accomplish this milestone - MS thesis.

HAIDER ALI KHAN
Bahria University Islamabad, Pakistan

2019

"Success is no accident. It is hard work, perseverance, learning, studying, sacrifice and most of all, love of what you are doing or learning to do."

Pele

Contents

1 Introduction **1**

1.1 Facial Forgery Methods 1

1.2 Problem Statement 4

1.3 Research Objectives 4

1.4 Research Contribution 4

1.5 Thesis Organization 4

2 Literature Review **5**

2.1 History of Facial forgery 5

2.2 History of Facial forgery detection 5

2.3 Datasets 8

2.3.1 DeepfakeTIMIT 9

2.3.2 FaceForensics 9

2.3.3 Deepfake Dataset 11

2.3.4 Other Datasets 11

3 Methods **13**

3.1 Dataset 13

3.2 Proposed Methodology 13

3.2.1 Data Preparation 14

3.2.2 Data Preprocessing 14

3.2.3 Overview of CNNs 15

3.3 Overview of InceptionResnet 18

3.3.1 Classification 21

3.4 Conclusion 21

4 Experiments and Results **23**

4.1 Experimental Protocol 23

4.2 Classification rate on Deepfake TIMIT dataset 24

4.3 Comparison 27

5 Conclusions & Perspectives **30**

5.1 Conclusion 30

List of Figures

1.1 An Example of Face2Face technique 2
1.2 Deepfake technique 3

2.1 An Example of Deepfake TIMIT dataset 9
2.2 An Example of Source-to-Target Reenactment Dataset 10
2.3 An Example of Source-to-Target Reenactment Dataset 10
2.4 An Example of Deepfake Dataset 11

3.1 An overview of proposed methodology 14
3.2 High-level general CNN architecture 16
3.3 3D representation of CNN input layer 16
3.4 A Convolutional layer of CNN along with input, output volumes. 17
3.5 Convolution of an image (32 x 32 x 3) with two (5 x 5 x 3) filters 17
3.6 An overview of inception net module 18
3.7 A residual block of Resnet 19
3.8 An overview of identity residual block 19
3.9 An overview of Conv residual block 20
3.10 Combination of residual block and inception net 20
3.11 Forged image classification using inceptionResNet-V2 21
3.12 Few detection results of Real and Fake images. 22

4.1 The training loss of model on Deepfake TIMIT dataset as a function of a number-of-epochs with learning rate 0.001 25
4.2 The accuracy of model on Deepfake TIMIT dataset as a function of a number of epochs with learning rate 0.001 25
4.3 The training loss of a model on Deepfake TIMIT dataset as a function of a number of epochs with learning rate 0.001 26
4.4 The accuracy of a model on Deepfake TIMIT dataset as a function of a number of epochs with learning rate 0.001 27

4.5 A comparison of the training loss of the model at learning rate 0.001 and 0.0001. 28
4.6 A comparison of the accuracy of the model at learning rate 0.001 and 0.0001. 28

List of Tables

2.1 Summary of techniques 8
2.2 Details of Source-to-Target Reenactment Dataset 10
2.3 Details of Self-Reenactment Dataset 11

4.1 Deepfake TIMIT dataset division for first experiment 23
4.2 Deepfake TIMIT dataset division for second experiment 24
4.3 Comparison of proposed solution and other solutions. 29

Chapter 1

Introduction

During the last decades, there is tremendous growth in the social network and the usage of technology gadgets such as cell phones and tablets has increased. Millions of digital images and pictures are uploaded on the internet on daily basis. Moreover, technology is growing day by day. Biometric systems have improved and the digital image of faces are used instead of finger or thumb impression. Such large use of digital images gives inflation to the methods to forge images, using different software, e.g Photoshop. Faces are most vulnerable part of an image as the face can be forged to hack biometric systems and making the fake news. Therefore digital face forensics research has emerged to detect fake images & videos.

Faceforensics is a popular area of digital image processing. Nowadays different applications are used to forge human faces to make fake videos and images. Even synthetic faces are used to generate videos in real time. Video face replacement[1], video rewrite[2], vDub[3], Generate adversarial networks (GANs)[4], Fader Networks[5] etc. are different methods used for human face forging. To detect such face forgeries is a very challenging task. The field of digital image forensics has become very popular in examining the authenticity of images.

1.1 Facial Forgery Methods

There are different falsification methods which are used to create forged images and videos. Face2Face[6] and Deepfake[7] are one of the famous techniques.

Face2Face This technique is proposed by Thies et al.[6]. It is another facial reenactment system, which is developed to alter the facial movements in videos.
Face2Face is the most advanced form of facial forgery. A monocular video is created by using a single camera e.g Videos uploaded on Youtube. Such videos are usually captured with a webcam. In this technique, such videos are reenacted and facial expressions of one video are imposed on the other video. The resultant video is also same as target video, i.e monocular video. This technique first reconstructs its facial model and for this purpose, it requires recorded video of the target person for a training. Then, at runtime, the program tracks expressions of both videos. In the final step, the source facial expressions are embedded on the target video frames respectively. An example of Face2Face is illustrated in Figure 1.1. Virtual reality also adopt the same technique for eye-tracking and reenactment [8].

Deepfake [7] replaces the face of one person with someone else in a video. It was used

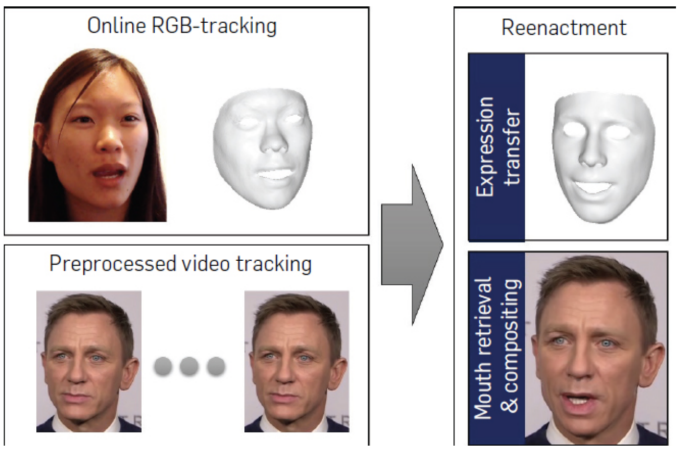

Figure 1.1: An Example of Face2Face technique

to generate face-swapped adult contents. FaceSwap[9] application is developed on the principle of Deepfake.
Two auto-encoders are used in the Deepfake architecture. The encoder reduces the dimension of data from the input layer by encoding. By this encoding process, the data of input layers are reduced to a number of variables. The same variables are used by the decoder for output, which is an approximation of the input. In the optimization step, the approximation generated by the input and the input is compared. Take the difference in this comparison and penalize it by using an L2 distance.
In Deepfake technique, images of resolution 64 x 64 x 3 = 12, 288 variables are fed to the original auto-encoder. After encoding the images, the encoder generates images of the same size as of input..

For the generation of Deepfake images, parallel faces of 2 persons X and Y are gathered. After that, an autoencoder Ex is trained to regenerate the faces of X from the dataset of X containing facial images and another autoencoder Ey is trained to regenerate the faces of Y from the dataset of Y containing facial images. The auto-encoders share the weights of their encoding part but their respective decoders are kept separated. After the optimization process, an image taken from dataset X is encoded by the shared encoder but it can be decoded with the Ey decoder. It is illustrated in Figure 1.2. The encoder encodes the

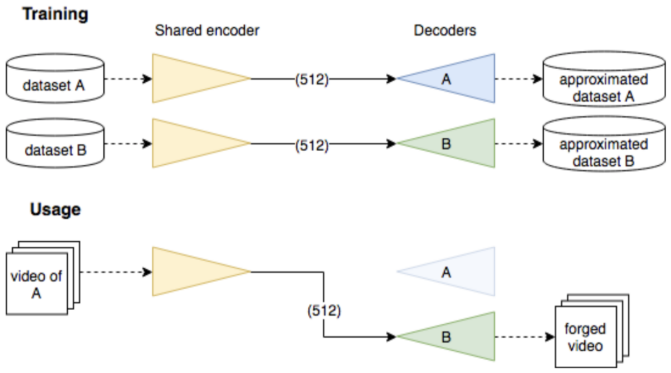

Figure 1.2: Deepfake technique

general information of illumination, facial position, facial expression and a decoder which is dedicated for individual person to regenerate constant shapes and other details of the face of the person.
Practically this technique is very popular because of its impressive results. The last step is the extraction and alignment of the target face from target video frames and the generation of another face with the same illumination and facial expressions. After the generation of the new face, it is merged back in the video.This technique has some flaws especially in case of occlusion the extraction and reintegration of faces can fail.

Nowadays Fake news has become a global risk and its danger is widely acknowledged. On social media, people watch more than 100 million videos on a daily basis, therefore, the chances of spreading fake news are increasing rapidly. Though different researchers have been putting their efforts for the detection of forged images, fake video detection still remains a challenging task.

In this research, our aim is to develop a comprehensive Facial forgery detection technique. The performance of the proposed technique will be evaluated on a standard benchmark.

1.2 Problem Statement

Facial forgery is very vulnerable to social media, news and sharing websites. Because of their large use, they have raised the techniques to make fake images and videos. Therefore there is always a need to detect such forgeries. Many studies have beed conducted for such facial forgery detection with different detection rates. As the techniques to forge the videos and images have become more advanced day by day, therefore detection of facial forgery needs more attention and improvement.
The proposed research aims to develop a robust technique for facial forgery detection with good accuracy. Deepfake TIMIT dataset[10] will be used for training, testing and validation.

1.3 Research Objectives

The key objectives of this research study are discussed in the following.

- To propose a robust technique for facial forgery detection in images/videos.
- To investigate the effectiveness of machine-learned features for fake images/video detection.
- To evaluate the proposed technique on publicly available benchmark datasets and also study the system generalization by merging multiple datasets.

1.4 Research Contribution

The key contributions of present study are discussed in following section.

- An automatic system to detect the forged images/videos.
- Classification of forged images/videos is carried out through machine learning classifier.
- The proposed technique is validated on Deepfake TIMIT dataset [10].

1.5 Thesis Organization

The thesis is organized as follow Chapter 2 provides an overview of Facial forgery and challenges associated with Facial forgery detection. Chapter 3 introduces pre-processing

steps and data preparation steps and experimental scenarios in our research. Chapter 4 presents a comprehensive discussion on CNN based classification results. Conclusion and further research directions are presented in the last chapter.

Chapter 2

Literature Review

2.1 History of Facial forgery

In previous years, a lot of work is done on face manipulation methods and these methods have improved to such extent that it is very difficult to detect forgery detection. Breglera et al.[2] presented Video Rewrite approach which automatically creates a forged video in which the movements of mouth are employed on the face of the person that the person did not speak in the original video. Garrido et al.[11] proposed a system in which the face of a target video is replaced with a face from the source video. Thies et al.[12] proposed a technique in which facial expressions of source video to target video are transferred in real-time. In this approach, the RGB-D sensor is used to get the source and target facial expressions in real time and recreated a 3D-model of both subjects. Recently, Suwajanakorn et al.[13] proposed a technique similar compositional methodology to Face2Face[6]. In this approach RNN is used to learn the mapping from raw audio features to mouth shapes. Lu et al.[14] provided an overview of several face image synthesis approaches using deep learning techniques. Deep Feature Interpolation[15] gave impressive results on facial transformations like "add smile", "make older/younger" and adding or removing mustache etc.

2.2 History of Facial forgery detection

ca Digital Image Forensics aims to ensure the originality of digital images and videos without any embedded security system. In past, the proposed methods were based on handcrafted features. H Farid [16] and HT Sancar et al.[17] provided surveys on these methods. B Bare et al.[18], L Bondi et al.[19] and JH Bappy et al.[20] proposed

CNN-Based solutions for image forgery detection. In videos, the major work is done on forgery dectection and it can be created with low efforts, like copy-move forgeries[21] and dropped or duplicated frames[22, 23].

Specifically for face forgeries detection, few methods have been proposed. DT Dang-Nguyen et al.[24] proposed a five-step method to differentiate natural human faces from computer-generated faces on the basis of facial expression. They performed the experiment on Boğaziçi-University-Head-Motion-Analysis-Project-Database (BUHMAP-DB)[25] and The Japanese-Female-Facial-Expression (JAFFE) Database[26]. The classification scores of the experiment were 96.67%.

Conotter, V et al.[27] proposed a forensic technique to distinguish real human faces from computer-generated faces by identifying fluctuations in the appearance of faces due to the heart beats. The drawback of this technique was that it was only applicable to the video of a person and might be modeler can artificially introduce a pulse to a CG character, or a forger can remove the pulse from a real human face.

In another study, Aparna Bharati et al.[28] proposed a supervised deep Boltzmann machine algorithm. In this technique, different parts of the face were used to learn discriminative features to classify whether face images are original or forged. They achieved accuracy over 87% on ND-IIITD Retouched-Faces database and 99% on Celebrity database.

In another study, R. Raghavendra et al.[29] proposed Deep CNN approach to detect whether an image is being morphed or not. The experiment was done on a morphed-image database, which was created from the morphed face database.[30]. This approach reported 8.23% Detection Equal Error Rate for digital images, 17.64% for HP Print-Scan and 12.47% for RICOH Print-Scan. Moreover, this approach reported 14.38% Attack Presentation Classification Error Rate (APCER) on digital images, 41.78% on HP Print-Scan and 28.76% on RICOH Print-Scan.

In another study, Peng Zhou et al.[31] proposed a two-stream network for face forgery detection. In this framework, a CNN was trained to classify whether an image is tempered or not. In the second phase which is patch triplet stream Steganalysis feature extractor is used to extract features and SVM is trained on the learned features for classification. The experiment was carried out on a newly created dataset and the dataset was created by using an iOS app called SwapMe [32] and an open-source face

swap application called FaceSwap [9]. The accuracy of face-level ROC for the proposed two-stream network was 0.927.

In another study Rozita Teymourzadeh et al.[33] designed an advanced Computer-Aided tool for image authentication and classification. In this approach, the originality of the image was authenticated by a neural network. The system was tested on two databases containing 1-10 training images each. The system reported $\pm 2\%$ error rate for image forgery detection.

In another study Shahroz Tariq et al.[34] proposed a NN based classifier for the detection of the fake faces created by Computer as well as Humans. The proposed method was able to detect Human created fake faces as well as GANs-created fake faces. The system was tested on CelebA Dataset[35] and Progressive Growing GANs Dataset (PGGAN)[36]. The system reported 94% AUROC score to detect GANs-created images and 74.9% AUROC score to detect human-created fake images.

In another study, Cuicui Guo et al.[37] proposed a Support-Vector-Machine (SVM) technique to detect facial expression reenacted forgery (FERF). The experiment was conducted on newly created FERF dataset and the system reported 70.11% accuracy.

In another study, Darius Afchar et al.[7] proposed MesoNet technique for facial-forgery-detection in videos. They introduced two methods: Meso-4 and MesoInception-4, both are based on CNN. The experiments were carried out on FaceForensics dataset [38] and Deepfake dataset [7]. The experiments demonstrated 98% detection rate for Deepfake and 95% detection rate for Face2Face.

P. Korshunov et al.[10] presented Deepfake TIMIT dataset. The dataset was evaluated on IQM and SVM [43], [44], and 91.03% detection rate was reported. An overview of all techniques for facial forgery detection is presented in Table 2.1

Table 2.1: Summary of techniques

Paper	Technique	Database	Evaluation result	
DT Dang-Nguyen et al.[24]	five-step method based on CNN	Boğaziçi-University-Head-Motion-Analysis-Project-Database (BUHMAP-DB)[25] and The Japanese Female Facial Expression (JAFFE) Database[26]	96.67% classification performance	Differentiate natural human faces from computer-generated faces on the basis of facial expression.
Aparna Bharati et al.[28]	Supervised deep Boltzmann machine algorithm	Private	upto 99% classification accuracy, but the database size was very small	Classify whether face images are original or retouched. e.g for driving license or models appearance
R. Raghavendra et al.[29]	Deep CNN approach	Newly constructed morphed image database from morphed face database[30]	8.23% Detection Equal Error Rate for digital images	Detect whether an image is being morphed or not. For face biometric systems
Peng Zhou et al.[31]	CNN, Steganalysis feature extractor & SVM	Private	0.927 classification for face level ROC	Classify whether an image is tempered or not, to identify the faceSwap
Rozita Teymourzadeh et al.[33]	Computer-Aided tool based on Neural Network	Private	$\pm$2% error rate for image forgery detection	Authentication of images for image recognition system.
Shahroz Tariq et al.[34]	Neural Network based classifier	CelebA Dataset[35] & Progressive Growing GANs Dataset (PGGAN)[36]	94% AUROC score to detect GANs-created images and 74.9% AUROC score to detect human-created fake images	detect fake human faces created by both machines and humans
Cuicui Guo et al.[37]	SVM	Private	70.11% classification accuracy	Detect facial expression reenacted forgery (FERF)
Darius Afchar et al.[7]	Meso-4 and MesoInception-4, both are based on CNN	FaceForensics[38] and Deepfake[7]	98% detection rate for Deepfake and 95% detection rate for Face2Face.	Detection of fake videos
P. Korshunov et al.[10]	IQM+SVM	Deepfake TIMIT[10]	91.03% detection rate.	Detection of fake videos

2.3 Datasets

Availability of a large of datasets is the fundamental requirement in the development and evaluation of facial forgery detection systems. Following datasets are created for this purpose.

2.3.1 DeepfakeTIMIT

The DeepfakeTIMIT[10] dataset contains videos in which faces are swapped by using open source GAN-based technique.
The dataset is created by selecting 16 similar looking pairs of people from publicly available VidTIMIT database. Two different models are trained to create fake videos. A lower quality model which created the fake videos with 64 x 64 resolution and a higher quality model which created the fake videos with 128 x 128 resolution. Deepfake TIMIT contains 620 fake videos. Figure 2.1 presents an example of Deepfake TIMIT dataset. We will use Deepfake TIMIT dataset for testing, training and validation of our proposed methodology.

Figure 2.1: An Example of Deepfake TIMIT dataset

2.3.2 FaceForensics

Faceforensics[38] dataset is generated by using Face2face technique. The data for this dataset was collected from youtube and the resolution of all the youtube videos was greater than 480p.
There are two variations of this dataset. Source-to-Target Reenactment Dataset and Self-Reenactment Dataset.
Face2Face[6] reenactment approach is applied to two randomly chosen videos to build Source-to-Target Reenactment Dataset. Mouth retrieval approach is used to select the mouth interiors from a mouth database based on the target expression. In the preprocessing step the mouth database is created from the tracked videos, which contains the video of target video. An example of Source-to-Target Reenactment Dataset is presented in Figure 2.2
The details of the Source-to-Target Reenactment Dataset is shown in Table 2.2.

Figure 2.2: An Example of Source-to-Target Reenactment Dataset

Table 2.2: Details of Source-to-Target Reenactment Dataset

Dataset	Videos	Images
Training	704	364,256
Validation	150	76,309
Testing	150	78,562

Self-Reenactment Dataset is created by using Face2Face[6] technique. In self-reenactment technique, similar video is used as the source and the target video. To obtain video pairs self-reenactment technique is applied to a video. These pairs contain ground truth and re-rendered facial images. These pairs are well suited to train generative approaches for Face Forensics. Figure 2.3 presents an example of Self-Reenactment Dataset.

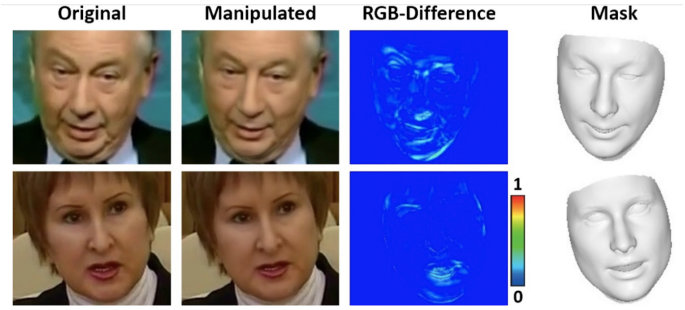

Figure 2.3: An Example of Source-to-Target Reenactment Dataset

The details of the Self-Reenactment Dataset is given in Table 2.3.

Table 2.3: Details of Self-Reenactment Dataset

Dataset	Videos	Images
Training	704	368,135
Validation	150	75,526
Testing	150	77,745

2.3.3 Deepfake Dataset

Deepfake dataset is generated by using Deepfake technique [7]. 175 forged videos were collected from different platforms. The length of these videos was up to three minutes and the resolution of each video was 854 x 480 pixels. H.264 codec was used to compress the videos with different compression levels. An example of Deepfake dataset is presented in Figure 2.4.

Figure 2.4: An Example of Deepfake Dataset

2.3.4 Other Datasets

There are also other datasets, on which different researchers conducted their experiments for facial forgery detection. Like CelebA Dataset[35], Progressive Growing GANs Dataset (PGGAN)[36], Head-Motion-Analysis-Project-Database (BUHMAP-DB)[25], Japanese-Female-Facial-Expression (JAFFE) Database[26], ND-IIITD Retouched-Faces database [28], Celebrity database [28] and Morphed face database[30]. The problem with these datasets is that some of these datasets are not publicly available and some datasets are very smaller in size. We have chosen FaceForensics dataset [38] to conduct our experiment.

The FaceForensics dataset [38] is specifically designed for facial forgery detection and it has many videos and images for training, validation, and testing.

Chapter 3

Methods

This chapter presents the details on Facial forgery detection using InceptionResNet model, which is a combination of InceptionNet and ResNet. We first present the details of Deepfake TIMIT dataset [10], an overview of basic CNN and the architecture of Inception-ResNet [39]. We then discuss each phase of the proposed methodology in detail i.e data preparation, data preprocessing and classification.

3.1 Dataset

As explained earlier, a number of datasets have been employed for facial forgery detection. In this study, we used a publicly available standard benchmark Deepfake TIMIT [10] dataset to detect the forgery from images. The dataset contains two types of videos i.e low-quality videos having 64 x 64 resolution and high-quality videos having 128 x 128 resolution. We performed different preparation and preprocessing steps over dataset in order to use in our proposed methodology. The detailed description of the proposed methodology is explained in the subsequent section.

3.2 Proposed Methodology

We extracted different frames from videos as a first step. The height and width of the images are normalized in order to feed the InceptionResNet model of CNN. It is a combination of ResNet and Inception methods. In Inception model, different sizes of feature maps can be used and pass through the convolutions of the existing layer and let the model choose the best one. Inception model can give a very better performance at a very low cost. In ResNet, the core idea is to introduce an identity shortcut connection to skip some layers. ResNet architecture maintains the gradient. It accelerates the training of Inception networks. In

inception networks, inception blocks reduces the computational complexity by introducing bottle neck layers. Therefore we used a combination of ResNet and Inception for feature extractions. We also added Dense-layer at the bottom layer of architecture to classify the extracted features. We used 26,315 images of real and fake videos as a training set and cross-validation is performed using 10, 693 images. we evaluated the performance of the model using 1500 images of real and fake videos respectively, where we realized very promising results. An overview of the proposed methodology is presented in Figure 3.1. In subsequent sections, we explained details of each step carried out in this study.

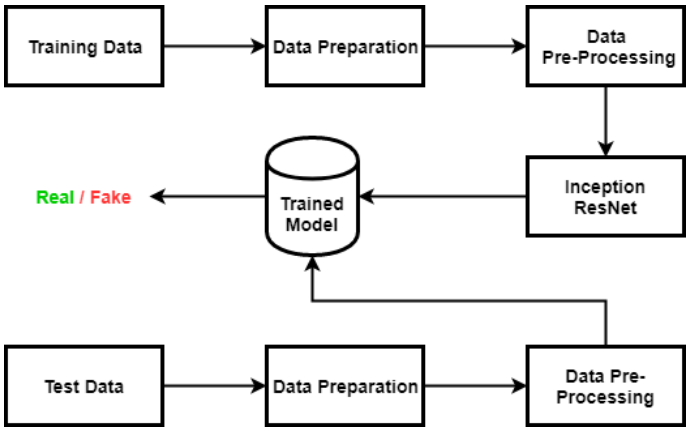

Figure 3.1: An overview of proposed methodology

3.2.1 Data Preparation

As mentioned above, we employed InceptionResnet model of CNN for feature extraction as well as for classification. The InceptionResnet followed different convolutional layers to extract the feature map from input images. The deepfake dataset contains two different types of videos therefore, we extracted image frames from videos in order to accomplish this study. We extracted total 38,508 images from 620 videos including real and fake where we extracted 50 to 90 image frames per video. We used 26,315 images for our training purpose, 10,693 images for validation and 1500 images for our testing purpose.

3.2.2 Data Preprocessing

Data preprocessing is a technique in which data is transformed into useful and readable formate. Most of the time there are many issues in the data and there is a need to preprocess the data before using it. Data preprocessing varies according to the research domain. e.g in Natural language processing: preprocessing apply to resolve case sensitivity, stemming and lemmatization issues. In the field of image processing: image resizing, noise removal,

segmentation, and morphology are some data preprocessing tasks. As we explained in the Data preparation section, videos are converted into image frames. The resolution of each image frame for high-quality videos is 128 x 128 and 64 x 64 for low-quality videos and all the frame images have different height and width. In data preprocessing these images are resized and the height and width of all the images are set to 299 x 299 so that these images can be used to give input to the classification model. After normalization of height and width of all the images, these images are fed to the InceptionResnet model and the convolutional layer of the model extract feature maps from the images. Though inceptionResnet is a type of CNN architecture, the detail of CNN is explained in the subsequent section.

3.2.3 Overview of CNNs

A Convolutional Neural Network is most commonly used deep learning architecture for image classification problem. CNN is so far been most popularly used for analyzing images. It performed well in object recognition and it is very popular among the image classification networks. As compared to other traditional approaches CNN based architectures provided the best results. CNN based architectures are widely used in Face recognition, Optical character recognition, and object detection methods.
CNN also performed good at analyzing sounds and also have been used in natural language translation/generation [40] and sentiment analysis [41].
Image recognition and classification is one of the main ability of CNN based architectures. A CNN takes images as an input, assign weights and biases to different aspects in the image and differentiates image from other images.
Traditional neural networks cannot perform well on images. Whereas CNN(s) take an image as an input where neurons are attached to the region of the image. In CNN(s) neurons are arranged in a three-dimensional structure i.e height, width, and depth. CNN(s) takes input data and transform it through all connected layers into scores given by the output layer. There are different variants of CNN architecture and all are based on the number and pattern of layers, as demonstrated in the Figure 3.2.
CNN architecture contains an Input layer, convolutional layer, pooling layer, Relu layer, and FC layer. Each layer is discussed in the following section.

Input layer CNN takes input via input layer in the form of an image. Raw input data of the image is stored and load by the CNN via input layer for the processing in the network. Width, height, and number of channels are specified by the input data. A 3D representation of input layer is presented in Figure 3.3.

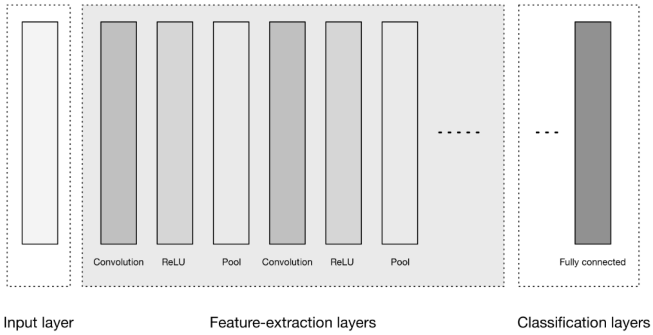

Figure 3.2: High-level general CNN architecture

Convolutional Layer Convolutional layers are the core building blocks of CNN architecture. In convolutional layer, different filters are used for feature extraction. As Figure 3.4 illustrates, convolutional layer takes input and transform the input data by using a patch of locally connecting neurons from the previous layer.

Convolutional layer extract features from the previous Layer and maps it into a feature

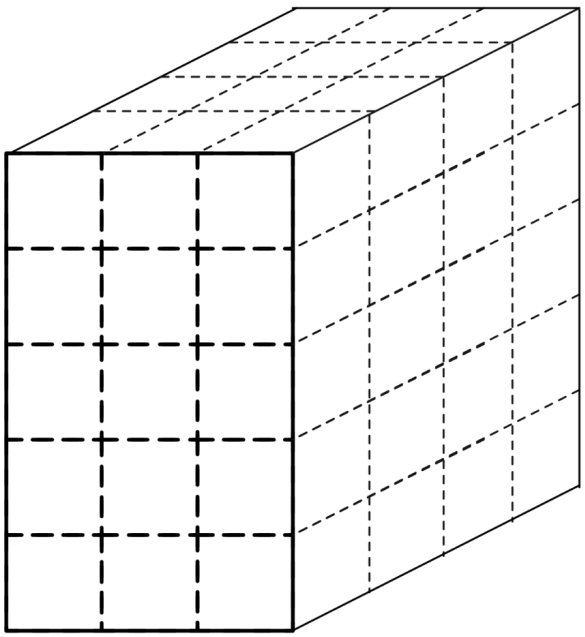

Figure 3.3: 3D representation of CNN input layer

map. In CNN architecture the initial convolutional layers extract low-level features, such as lines and curves, etc. The next layers use these features to extract problem specific features, such as eye, nose, etc. in a human image.

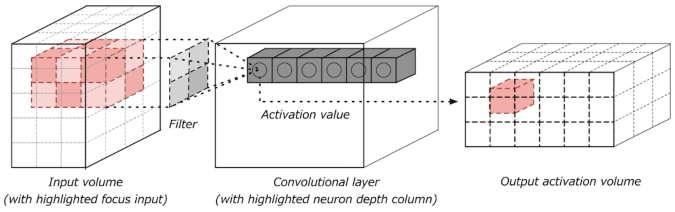

Figure 3.4: A Convolutional layer of CNN along with input, output volumes.

CNN uses a different type of filters to detect patterns in the images. These filters are used for different purposes such as image blurring, image sharpening, and edge detection, etc. A separate feature map is generated by each filter. For the computation of output, all of the feature maps are combined against the depth of CNN. In Figure 3.5 32 X 32 X 3 image is convolved with two 5 X 5 X 3 filters which produce two activation maps. Depth of the filter remains the same as input whereas the number and size of the filter

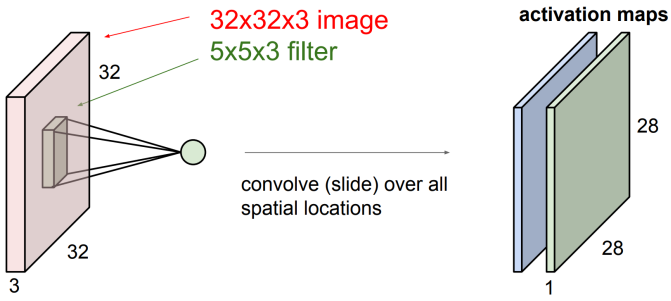

Figure 3.5: Convolution of an image (32 x 32 x 3) with two (5 x 5 x 3) filters

may vary. Three parameters involved in the size of output volume, i.e padding, depth, and stride. Strides represent the number of pixels while moving the filter in the input volume to calculate the next cell. In padding extra rows and columns are added at the border of an input image so that convolution can be performed at border pixels. It is used to adjust the size of input according to our requirements. Depth shows filter number in a layer.

Relu layer It is a nonlinear activation function which is applied on a feature map to introduce non-linearity in a convolutional neural network. The main reason for using ReLU is, it avoids the vanishing gradient problem and its performance is better than the other activation functions, such as tang or sigmoid. The mathematical representation of ReLU function is: r(x) = max(0,x).

Pooling layer The pooling layer normalizes the size of the activation-map. It is

periodically used after multiple convolutional layers. Pooling layer normalizes the parameters/ size of the activation map and this downsampling Prevents the model from over-fitting. There are different kinds of pooling, such as Max-pooling, Avg-pooling, and Sum-pooling. Max pooling is the most commonly used pooling operation. Max pooling uses a 2x2 filter for pooling operation.

Fully connected layer In a fully connected layer, each neuron is fully connected to every neuron in the previous layer creating a mesh topology. Pooling layer contains high Level features. It computes the class scores on the basis of these high-level features and classifies the data into different classes. The number of neurons in the Fully connected layer is the same in number as of the unique classes in the problem.
Though we used inceptionResNet as the proposed classification model, therefore the detailed description of inceptionResNet is explained in the following section.

3.3 Overview of InceptionResnet

The proposed architecture for our facial forgery detection system is InceptionResnet. It is a combination of ResNet and Inception methods. In Inception model, different sizes of feature maps can be used and pass through the convolutions of the existing layer and let the model choose the best one. Inception model can give a very better performance at a very low cost. An inception net module is given in Figure 3.6

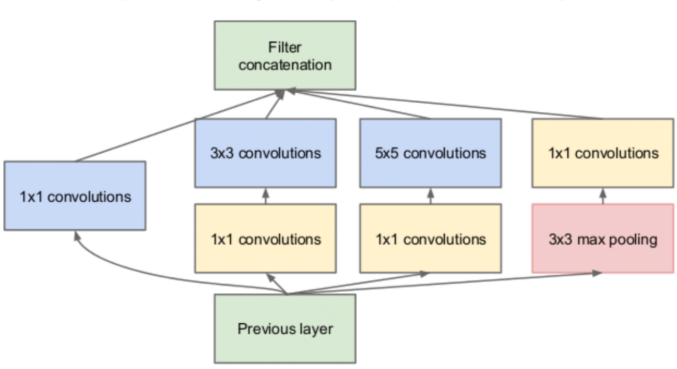

Figure 3.6: An overview of inception net module

In ResNet, the core idea is to introduce an identity shortcut connection to skip some layers as shown in Figure 3.7. There are two types of residual blocks: Identity block, in which input is directly passed to the output as shown in Figure 3.8. Conv block, in

which one convolutional layer and one batch Norm is added to the shortcut as shown in the Figure 3.9.

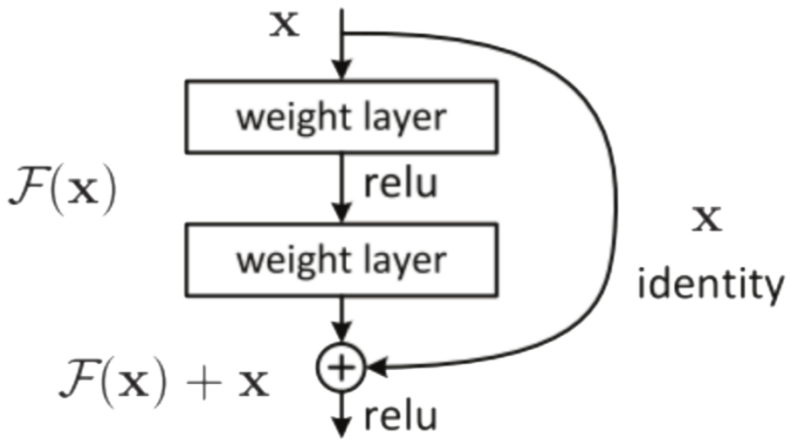

Figure 3.7: A residual block of Resnet

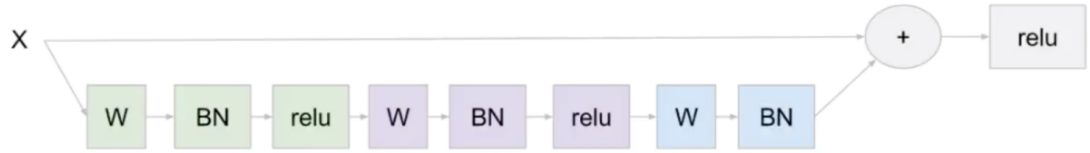

Figure 3.8: An overview of identity residual block

The idea is to bypass all the intermediate layers and just pass input directly to the output. ResNet architecture maintains the gradient. It accelerates the training of Inception networks. The architecture of Inception net and Residual connection are combined to form InceptionResNet architecture as described in Figure 3.10.

As there are multiple blocks in inceptionResNet, therefore each residual connection combined different sizes of convolutional filters. Inception net is a very deep neural network, therefore it has a high training time. By using this identity function, InceptionResNet reduces training time and the problem of degradation. The architecture of inceptionResNet in this study is given in Figure 3.11.

We adopted this network as a pre-trained architecture where InceptionResnet was trained on ImageNet dataset. we used previously learned weights and fine-tuned the model

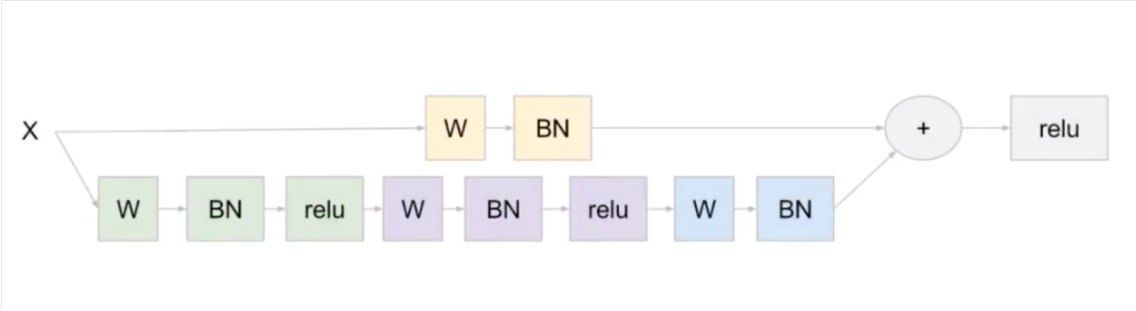

Figure 3.9: An overview of Conv residual block

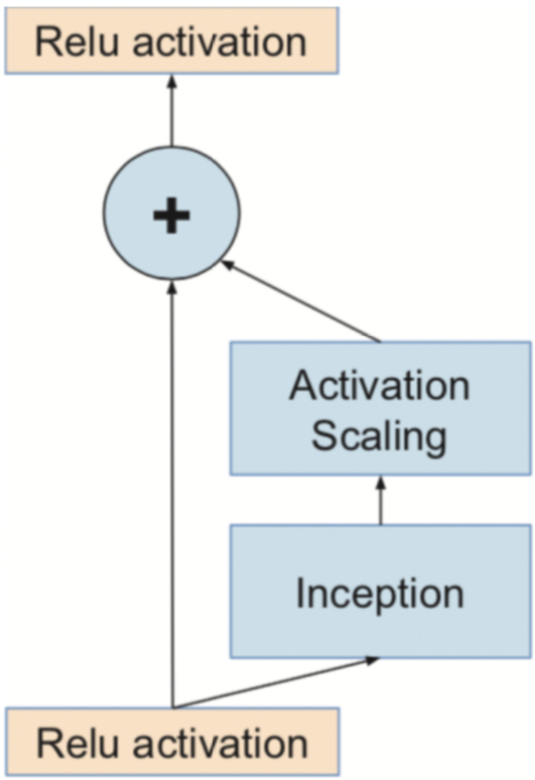

Figure 3.10: Combination of residual block and inception net

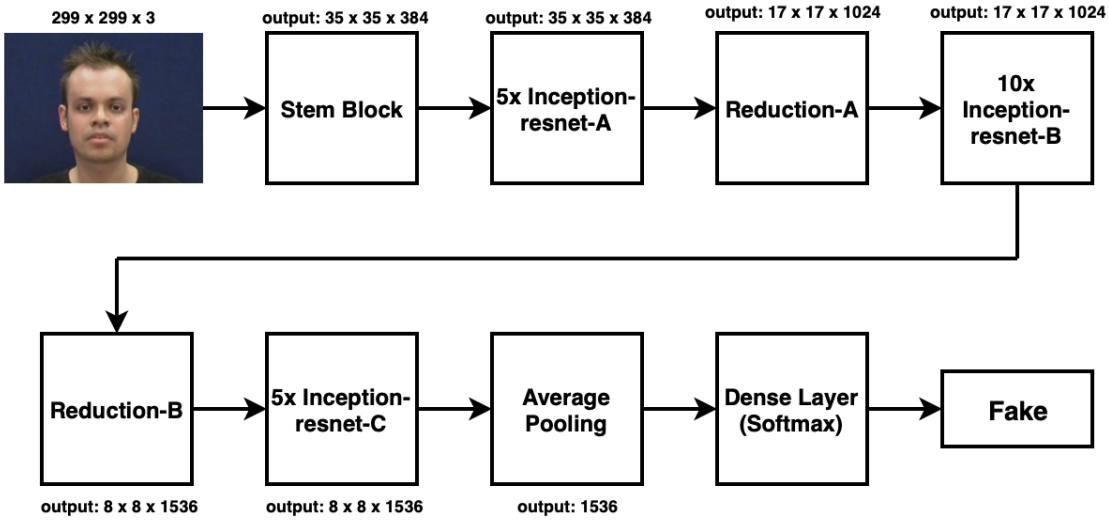

Figure 3.11: Forged image classification using inceptionResNet-V2

on deepfake dataset. We also embed a dense layer according to input image dimension for classification purpose. the detail of the classification step is illustrated in the given section.

3.3.1 Classification

Classification is a problem of supervised learning in which labels are assigned to different classes according to their categories. In classification, the classifier discriminates one class from other classes on the bases of feature values. In this study, we classified extracted features using Dense layer. For this purpose, a dense layer is added to the end of inceptionResNet architecture in order to discriminate the real and forged images. A dense layer is just a common layer and it receives input from the neurons of the previous-layer with own weight-matrix, bias-value, and activation function.

3.4 Conclusion

After the complete description of each phase of the proposed technique, we present summarised steps of our proposed architecture. In data preparation step we converted the videos in image frames. The input images are normalized and resized to 299 x 299 as a preprocessing step. After preprocessing, the images are fed to the inceptionResNet for feature extraction and these feature maps were used for classification where we employed a dense layer. Some examples of detection results of proposed system are illustrated in Figure 3.12.

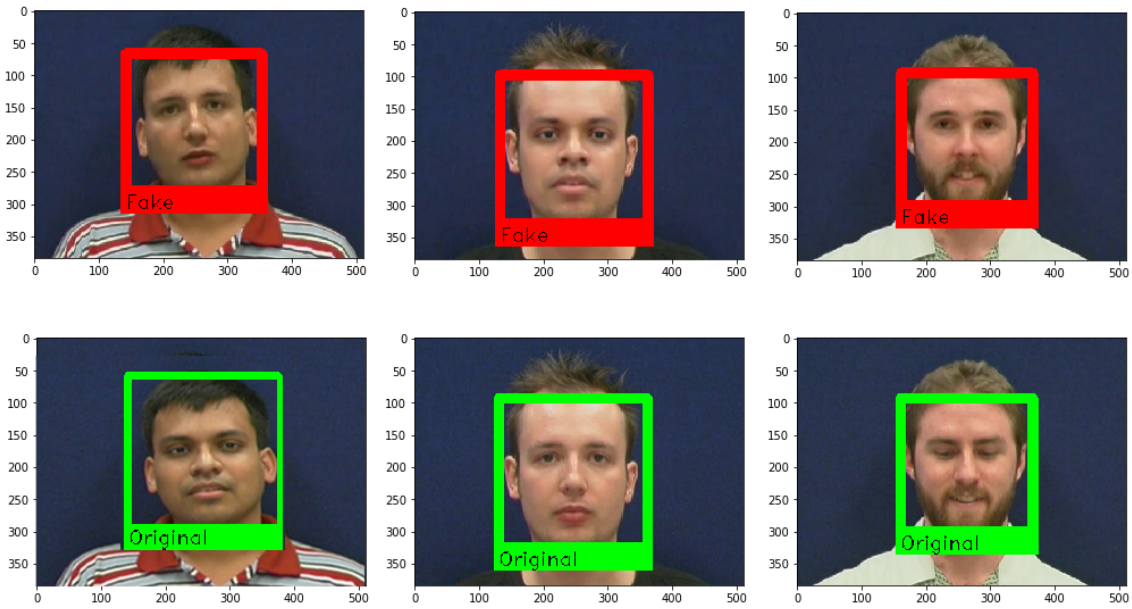

Figure 3.12: Few detection results of Real and Fake images.

Chapter 4

Experiments and Results

This chapter contains the experimental details of the experiment which we carried out to check and evaluate our proposed facial forgery detection system. We used Deepfake TIMIT dataset [10] to conduct our study. First of all, we present the experimental protocols. Later we present the classification rate which we computed via different experiments.

4.1 Experimental Protocol

To evaluate the efficiency of the classification system we employed Deepfake TIMIT dataset [10] which is publicly available dataset. As explained earlier in chapter 2, Deepfake TIMIT [10] contains 620 fake videos. This dataset contains 38,508 images. In the first experiment, we used 26,315 images for our training purpose, 10,693 images for validation and 1500 images for our testing purpose. The division of Deepfake TIMIT dataset for our experiment is illustrated in Table 4.1.

Table 4.1: Deepfake TIMIT dataset division for first experiment

Dataset	**Images**
Training	26,315
Validation	10,693
Testing	1500

In the second experiment, we used a small portion of the dataset. Only 19 frames per video were selected for this experiment. We employed 4,750 images for our training purpose, 2,053 images for validation and 1500 images for our testing purpose. The division of Deep fake TIMIT dataset for the above experimental setup is illustrated in Table 4.2.

Table 4.2: Deepfake TIMIT dataset division for second experiment

Dataset	Images
Training	4,750
Validation	2,053
Testing	1500

The details of the model training and the evaluation of the classification rate on Deepfake TIMIT dataset [10] is given in the next section.

4.2 Classification rate on Deepfake TIMIT dataset

Initially, we evaluated our detection system on the whole Deepfake TIMIT dataset [10]. As we explained earlier that in the first experiment we used 26,315 images for our training purpose and 10,693 images for validation purpose. We used the transfer learning approach and adopted InceptionResnet which was already trained on a very large imageNet dataset [42]. We converted all the videos of Deepfake TIMIT dataset into frames and resize the input images to 299 x 299 resolution. The resultant images are fed to the InceptionResnet model for the extraction of feature maps. As mentioned earlier in the methodology section, we employed a dense layer at the end of the inceptionResNet model. The extracted feature maps were used by the dense layer for classification. We trained our model on Google Colab with 12GB Ram and 48GB Storage capacity for 12 hours. it has Tesla k80 GPU with 14 GB graphics memory. The learning rate is set to 0.001 for training and the batch size is set to 5. The training of the model took about 8 to 10 hrs for only 2 epochs as the number of training images was very large. Therefore we use the concept of checkpoints, i.e saving epochs along with model training. We trained the model for 10 epochs. We also validated the model on 10,693 validation images. The training-loss of the model is shown in Figure 4.1

It is clear from the Figure 4.1, that as the number of epochs increases, the loss decreases. After 7th epoch the loss suddenly increased, therefore we stopped training our model and finalize our model at a 7th epoch which has minimum loss 0.0175. Also, the accuracy of the model as a function of a number of epochs is shown in Figure 4.2. It shows that the accuracy increases as we increase the number of epochs and the accuracy is maximum at 7th epoch i.e 99.5%.

In the second experiment, we used a subpart of Deepfake TIMIT dataset. We took only 19 frame images per video. We used 4,750 images for our training purpose and 2,053

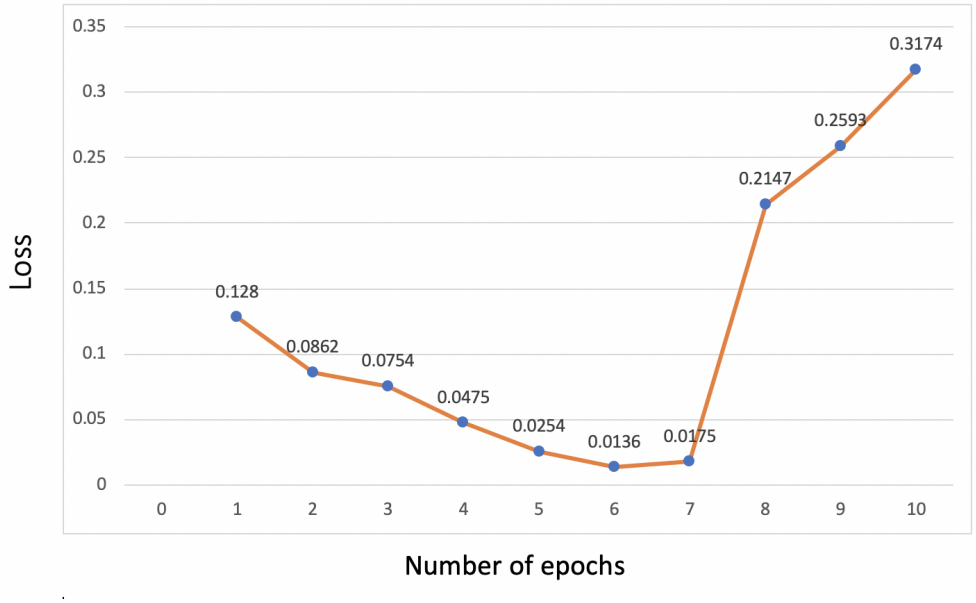

Figure 4.1: The training loss of model on Deepfake TIMIT dataset as a function of a number-of-epochs with learning rate 0.001

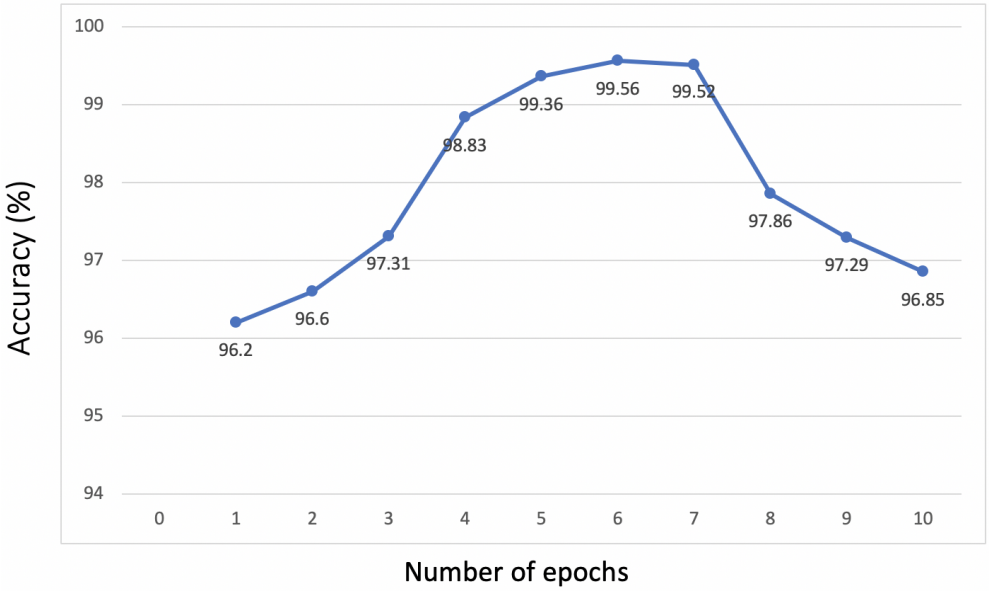

Figure 4.2: The accuracy of model on Deepfake TIMIT dataset as a function of a number of epochs with learning rate 0.001

images for validation purpose while conducting the experiment. The learning rate was set to 0.001 and the model was trained for 10 epochs. The batch size is set to 19 for this experiment.

The training loss of the model as a function of a number of epochs in this experiment is illustrated in Figure 4.3. It shows that the number of loss decreases with the increase of epochs. The loss decrease stopped at 8th epoch and began increasing, therefore we stopped our training at this point.

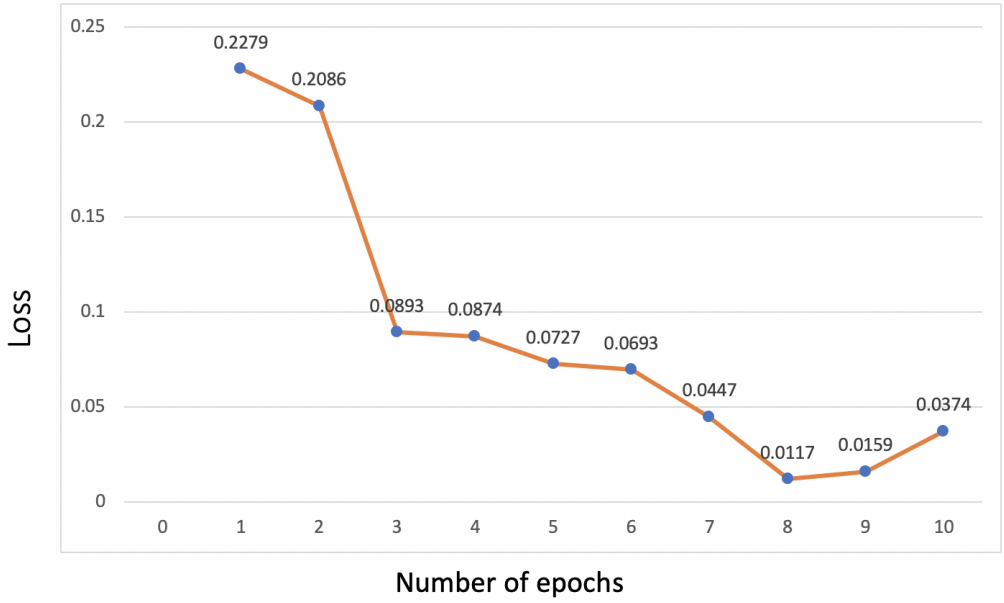

Figure 4.3: The training loss of a model on Deepfake TIMIT dataset as a function of a number of epochs with learning rate 0.001

The accuracy of the experiment is illustrated in Figure 4.4. It is clear from Figure 4.4 that accuracy is increasing with the number of epochs. The accuracy stopped increasing at 8th epoch, therefore we stopped training the model and finalized our model at 8th epoch.

As the loss of the model starts increasing after certain epoch, therefore we conducted another experiment. In this experiment, we used the same subpart of Deepfake TIMIT dataset as we used in experiment 2. We took only 19 frame images per video. We used 4,750 images for our training purpose and 2,053 images for validation purpose while conducting the experiment. This time, the learning rate was set to 0.0001 and the model was trained for 10 epochs. The batch size is set to 19 for this experiment. It is clear from the Figure 4.3, that at learning rate 0.0001, training loss becomes stable, instead of decreasing. The accuracy also stopped decreasing and become stable. Therefore we stopped training our model at this point. The comparison of model accuracy at learning rate 0.001 and 0.0001 is illustrated in Figure 4.3.

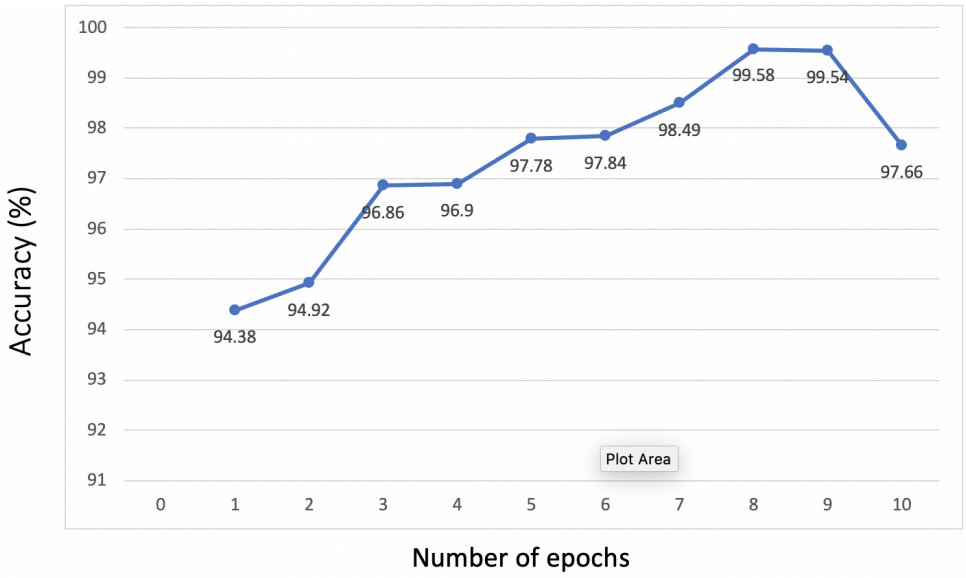

Figure 4.4: The accuracy of a model on Deepfake TIMIT dataset as a function of a number of epochs with learning rate 0.001

The training loss of the model as a function of a number of epochs in this experiment is illustrated in Figure 4.3. It shows that the number of loss decreases with the increase of epochs. The loss decrease stopped at 8th epoch and began increasing, therefore we stopped our training at this point.

We computed classification rate on Deepfake TIMIT dataset using InceptionRenet which is a combination of Inception and Resnet architectures. We conducted two experiments with different parameters to highlight the efficiency of the classification system. We achieved 99.52% and 99.58% of classification accuracy.

4.3 Comparison

As explained earlier in chapter 2, different studies have been done for the detection of facial forgery. Cuicui Guo et al.[37] proposed a Support Vector Machine (SVM) technique to detect facial expression reenacted forgery (FERF) and reported 70.11% accuracy. After that Darius Afchar et al. [7] proposed MesoNet technique for facial forgery detection in videos. They introduced two methods: Meso-4 and MesoInception-4 and reported 98% for detection rate tor Deepfake dataset [7]. P. Korshunov et al.[10] presented Deepfake TIMIT dataset. The dataset was evaluated on IQM and SVM [43], [44], and 91.03% detection rate

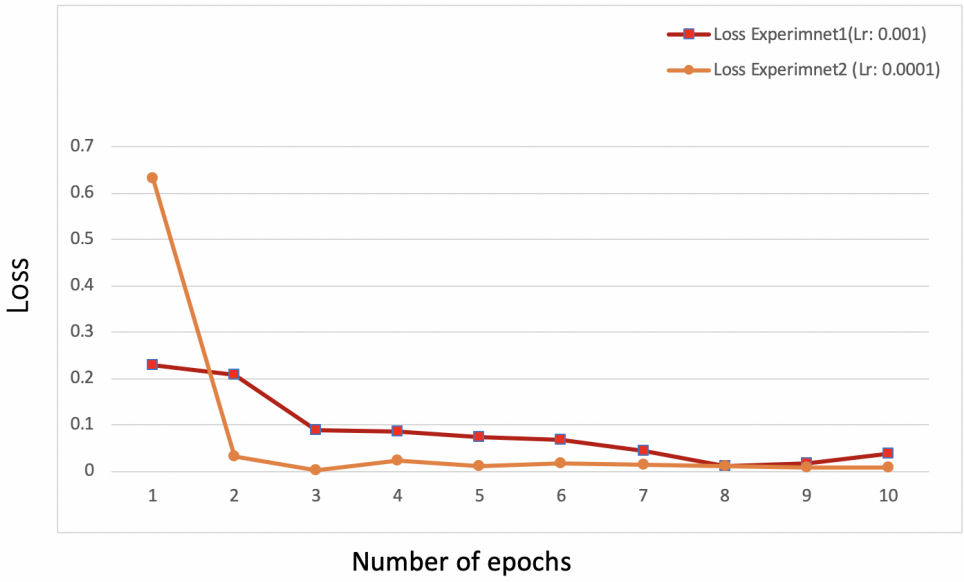

Figure 4.5: A comparison of the training loss of the model at learning rate 0.001 and 0.0001.

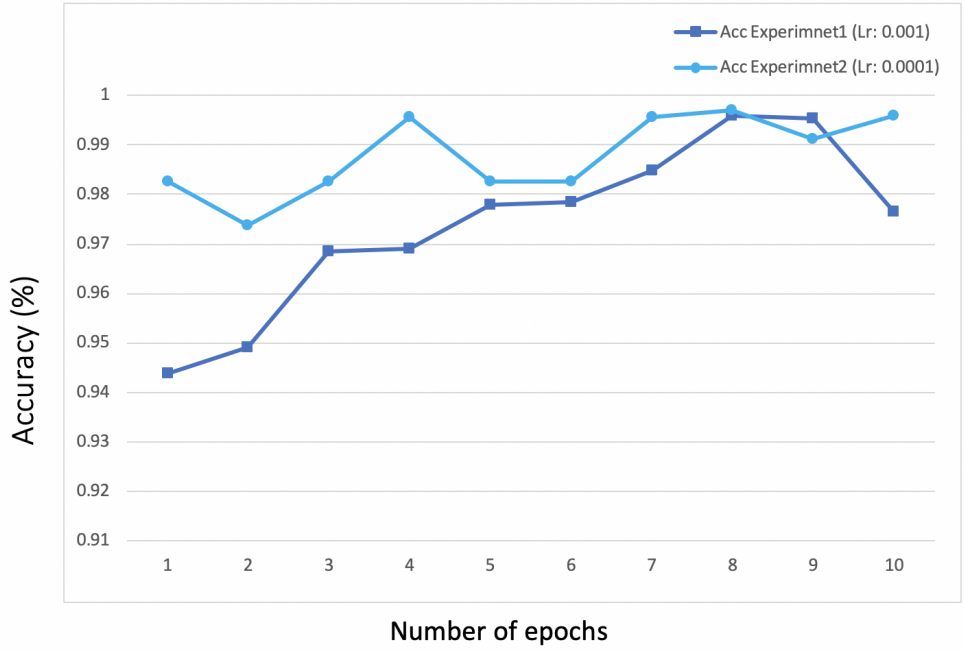

Figure 4.6: A comparison of the accuracy of the model at learning rate 0.001 and 0.0001.

was reported. We compared our proposed solution with [43], [44]. Our proposed solution reported 99.5% detection rate for publicly available Deepfake TIMIT dataset [10]. An overview of the comparison is shown in Table 4.3

Table 4.3: Comparison of proposed solution and other solutions.

Source	**Technique**	**Database**	**Evaluation result**
P. Korshunov et al.[10]	IQM+SVM	Deepfake TIMIT[10]	91.03% detection rate.
Proposed solution	InceptionResNet	Deepfake TIMIT Dataset [10]	99.5% classification accuracy

Chapter 5

Conclusions & Perspectives

Facial forgery detection is the most difficult problem. Many types of research have been done on forgery detection and this research has been refining day by day. But as the methodologies of creating forged videos are also maturing, therefore, its detection has become more difficult. This study investigated the problem of facial forgery detection. The remarks on the results of the study and future work direction are presented in the next section.

5.1 Conclusion

This study presented an effective technique for the detection of facial forgery in images. The images were fed to the learning algorithm to learn the forgeries in the images. We used InceptionResnet which is a combination of inceptionNet and Resnet. Validation of the proposed detection system is carried out on Deepfake TIMIT dataset [10]. We used the transfer learning approach while training our model and achieved the best results. We achieved 99.5% of classification accuracy score.

Bibliography

[1] K. Dale, K. Sunkavalli, M. K. Johnson, D. Vlasic, W. Matusik, and H. Pfister, "Video face replacement," *ACM Transactions on Graphics (TOG)*, vol. 30, no. 6, p. 130, 2011. Cited on p. 1.

[2] C. Bregler, M. Covell, and M. Slaney, "Video rewrite: Driving visual speech with audio," in *Proceedings of the 24th annual conference on Computer graphics and interactive techniques*, pp. 353–360, ACM Press/Addison-Wesley Publishing Co., 1997. Cited on pp. 1 and 5.

[3] P. Garrido, L. Valgaerts, H. Sarmadi, I. Steiner, K. Varanasi, P. Perez, and C. Theobalt, "Vdub: Modifying face video of actors for plausible visual alignment to a dubbed audio track," in *Computer Graphics Forum*, vol. 34, pp. 193–204, Wiley Online Library, 2015. Cited on p. 1.

[4] G. Antipov, M. Baccouche, and J.-L. Dugelay, "Face aging with conditional generative adversarial networks," in *Image Processing (ICIP), 2017 IEEE International Conference on*, pp. 2089–2093, IEEE, 2017. Cited on p. 1.

[5] G. Lample, N. Zeghidour, N. Usunier, A. Bordes, L. Denoyer, *et al.*, "Fader networks: Manipulating images by sliding attributes," in *Advances in Neural Information Processing Systems*, pp. 5967–5976, 2017. Cited on p. 1.

[6] J. Thies, M. Zollhofer, M. Stamminger, C. Theobalt, and M. Nießner, "Face2face: Real-time face capture and reenactment of rgb videos," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2387–2395, 2016. Cited on pp. 1, 2, 5, 9, and 10.

[7] D. Afchar, V. Nozick, J. Yamagishi, and I. Echizen, "Mesonet: a compact facial video forgery detection network," *arXiv preprint arXiv:1809.00888*, 2018. Cited on pp. 1, 2, 7, 8, 11, and 27.

[8] J. hies, M. Zollhöfer, M. Stamminger, C. Theobalt, and M. Nießner, "Facevr: Real-time gaze-aware facial reenactment in virtual reality," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 25:1–25:15, 2018. Cited on p. 2.

[9] "Faceswap. https://github.com/ marekkowalski/faceswap/.," Cited on pp. 2 and 7.

[10] P. Korshunov and S. Marcel, "Deepfakes: a new threat to face recognition? assessment and detection," *arXiv preprint arXiv:1812.08685*, 2018. Cited on pp. 4, 7, 8, 9, 13, 23, 24, 27, 29, and 30.

[11] P. Garrido, L. Valgaerts, O. Rehmsen, T. Thormaehlen, P. Perez, and C. Theobalt, "2014 ieee conference on computer vision and pattern recognition (cvpr)(2014)," Cited on p. 5.

[12] J. Thies, M. Zollhöfer, M. Nießner, L. Valgaerts, M. Stamminger, and C. Theobalt, "Real-time expression transfer for facial reenactment.," *ACM Trans. Graph.*, vol. 34, no. 6, pp. 183–1, 2015. Cited on p. 5.

[13] S. Suwajanakorn, S. M. Seitz, and I. Kemelmacher-Shlizerman, "Synthesizing obama: learning lip sync from audio," *ACM Transactions on Graphics (TOG)*, vol. 36, no. 4, p. 95, 2017. Cited on p. 5.

[14] Z. Lu, Z. Li, J. Cao, R. He, and Z. Sun, "Recent progress of face image synthesis," *arXiv preprint arXiv:1706.04717*, 2017. Cited on p. 5.

[15] P. Upchurch, J. R. Gardner, G. Pleiss, R. Pless, N. Snavely, K. Bala, and K. Q. Weinberger, "Deep feature interpolation for image content changes.," in *CVPR*, pp. 6090–6099, 2017. Cited on p. 5.

[16] H. Farid, *Photo forensics*. MIT Press, 2016. Cited on p. 5.

[17] H. T. Sencar and N. Memon, "Digital image forensics," *Counter-Forensics: Attacking Image Forensics*, pp. 327–366, 2013. Cited on p. 5.

[18] 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*, pp. 5–10, ACM, 2016. Cited on p. 5.

[19] L. Bondi, S. Lameri, D. Guera, P. Bestagini, E. J. Delp, S. Tubaro, *et al.*, "Tampering detection and localization through clustering of camera-based cnn features.," in *CVPR Workshops*, pp. 1855–1864, 2017. Cited on p. 5.

[20] J. H. Bappy, A. K. Roy-Chowdhury, J. Bunk, L. Nataraj, and B. Manjunath, "Exploiting spatial structure for localizing manipulated image regions," in *Proceedings of the IEEE International Conference on Computer Vision*, pp. 4970–4979, 2017. Cited on p. 5.

[21] L. D'Amiano, D. Cozzolino, G. Poggi, and L. Verdoliva, "A patchmatch-based dense-field algorithm for video copy-move detection and localization," *arXiv preprint arXiv:1703.04636*, 2017. Cited on p. 6.

[22] A. Gironi, M. Fontani, T. Bianchi, A. Piva, and M. Barni, "A video forensic technique for detecting frame deletion and insertion.," in *ICASSP*, pp. 6226–6230, 2014. Cited on p. 6.

[23] C. L. E. S. A. Basharat and A. Hoogs, "A c3d-based convolutional neural network for frame dropping detection in a single video shot," 2017. Cited on p. 6.

[24] D.-T. Dang-Nguyen, G. Boato, and F. G. De Natale, "Identify computer generated characters by analysing facial expressions variation," in *Information Forensics and Security (WIFS), 2012 IEEE International Workshop on*, pp. 252–257, IEEE, 2012. Cited on pp. 6 and 8.

[25] O. Aran, I. Ari, A. Guvensan, H. Haberdar, Z. Kurt, I. Turkmen, A. Uyar, and L. Akarun, "A database of non-manual signs in turkish sign language," in *Signal Processing and Communications Applications, 2007. SIU 2007. IEEE 15th*, pp. 1–4, IEEE, 2007. Cited on pp. 6, 8, and 11.

[26] M. Lyons, S. Akamatsu, M. Kamachi, and J. Gyoba, "Coding facial expressions with gabor wavelets," in *Automatic Face and Gesture Recognition, 1998. Proceedings. Third IEEE International Conference on*, pp. 200–205, IEEE, 1998. Cited on pp. 6, 8, and 11.

[27] V. Conotter, E. Bodnari, G. Boato, and H. Farid, "Physiologically-based detection of computer generated faces in video," in *Image Processing (ICIP), 2014 IEEE International Conference on*, pp. 248–252, IEEE, 2014. Cited on p. 6.

[28] A. Bharati, R. Singh, M. Vatsa, and K. W. Bowyer, "Detecting facial retouching using supervised deep learning," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 9, pp. 1903–1913, 2016. Cited on pp. 6, 8, and 11.

[29] R. Raghavendra, K. B. Raja, S. Venkatesh, and C. Busch, "Transferable deep-cnn features for detecting digital and print-scanned morphed face images," in *Computer Vision and Pattern Recognition Workshops (CVPRW), 2017 IEEE Conference on*, pp. 1822–1830, IEEE, 2017. Cited on pp. 6 and 8.

[30] U. Scherhag, R. Raghavendra, K. Raja, M. Gomez-Barrero, C. Rathgeb, and C. Busch, "On the vulnerability of face recognition systems towards morphed face attacks," in *Biometrics and Forensics (IWBF), 2017 5th International Workshop on*, pp. 1–6, IEEE, 2017. Cited on pp. 6, 8, and 11.

[31] P. Zhou, X. Han, V. I. Morariu, and L. S. Davis, "Two-stream neural networks for tampered face detection," in *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). IEEE*, pp. 1831–1839, 2017. Cited on pp. 6 and 8.

[32] "Swapme. https://itunes.apple.com/us/app/ swapme-by-faciometrics/.," Cited on p. 6.

[33] R. Teymourzadeh, Y. Samir, and M. Othman, "Design an advance computer-aided tool for image authentication and classification," *arXiv preprint arXiv:1808.02085*, 2018. Cited on pp. 7 and 8.

[34] S. Tariq, S. Lee, H. Kim, Y. Shin, and S. S. Woo, "Detecting both machine and human created fake face images in the wild," in *Proceedings of the 2nd International Workshop on Multimedia Privacy and Security*, pp. 81–87, ACM, 2018. Cited on pp. 7 and 8.

[35] Z. Liu, P. Luo, X. Wang, and X. Tang, "Deep learning face attributes in the wild," *2015 IEEE International Conference on Computer Vision (ICCV)*, Dec 2015. Cited on pp. 7, 8, and 11.

[36] T. Karras, T. Aila, S. Laine, and J. Lehtinen, "Progressive growing of gans for improved quality, stability, and variation," 2017. Cited on pp. 7, 8, and 11.

[37] C. Guo, G. Luo, and Y. Zhu, "A detection method for facial expression reenacted forgery in videos," in *Tenth International Conference on Digital Image Processing (ICDIP 2018)*, vol. 10806, p. 108061J, International Society for Optics and Photonics, 2018. Cited on pp. 7, 8, and 27.

[38] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "Faceforensics: A large-scale video dataset for forgery detection in human faces," *arXiv preprint arXiv:1803.09179*, 2018. Cited on pp. 7, 8, 9, 11, and 12.

[39] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the inception architecture for computer vision," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2818–2826, 2016. Cited on p. 13.

[40] J. Gehring, M. Auli, D. Grangier, and Y. N. Dauphin, "A convolutional encoder model for neural machine translation," *arXiv preprint arXiv:1611.02344*, 2016. Cited on p. 15.

[41] C. Dos Santos and M. Gatti, "Deep convolutional neural networks for sentiment analysis of short texts," in *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pp. 69–78, 2014. Cited on p. 15.

[42] O. Russakovsky, J. Deng, H. Su, J. Krause, S. Satheesh, S. Ma, Z. Huang, A. Karpathy, A. Khosla, M. Bernstein, *et al.*, "Imagenet large scale visual recognition challenge," *International journal of computer vision*, vol. 115, no. 3, pp. 211–252, 2015. Cited on p. 24.

[43] J. Galbally and S. Marcel, "Face anti-spoofing based on general image quality assessment," in *2014 22nd International Conference on Pattern Recognition*, pp. 1173–1178, IEEE, 2014. Cited on pp. 7, 27, and 29.

[44] D. Wen, H. Han, and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 746–761, 2015. Cited on pp. 7, 27, and 29.

ORIGINALITY REPORT

17%	10%	13%	9%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	export.arxiv.org Internet Source	3%
2	"Computer Vision – ECCV 2016 Workshops", Springer Science and Business Media LLC, 2016 Publication	1%
3	Submitted to Higher Education Commission Pakistan Student Paper	1%
4	www.cs.dartmouth.edu Internet Source	1%
5	individual.utoronto.ca Internet Source	1%
6	Submitted to University of Melbourne Student Paper	<1%
7	Shahroz Tariq, Sangyup Lee, Hoyoung Kim, Youjin Shin, Simon S. Woo. "Detecting Both Machine and Human Created Fake Face Images In the Wild", Proceedings of the 2nd	<1%