Towards Real-Time Deep Fake Forensic System: An Attention Guided Approach

by

Maj Rai Sabir Hussain
00000325423
Supervisor
Asst Prof Dr Saddaf Rubab

A thesis submitted to the faculty of CSE Department, Military College of Signals, National University of Sciences and Technology, Rawalpindi in partial fulfilment of the requirements for the degree of MS in Software Engineering

Sept 2023

THESIS ACCEPTANCE CERTIFICATE

Certified that final copy of MS thesis written by Mr Rai Sabir Hussain, Registration No. 0000325423, of Military College of Siqnals has been vetted by undersigned, found complete in all respect as per NUST Statutes/Regulations/ MS Policy, is free of plagiarism, errors and mistakes and is accepted as partial, fulfillment for award of MS degree. It is further certified that necessary amendments as pointed out by GEC members and local evaluators of the scholar have also been incorporated in the said thesis.

Signature:

Name of Supervisor: Asst Prof Dr Saddaf Rubab

Date.

Brig

Head of Dept of CSE
Mil College of Sigs (NUST)

Signature (HOD):

Date:

Signature (Dean/Principal):

30|9|23

Dean, MCS (N
(Asif Masood,

asood,

Brig

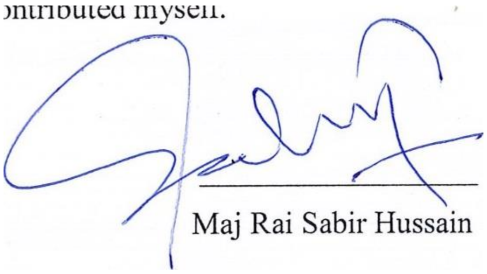
(NUST) Date. Phd)

Declaration

I, Rai Sabir Hussain, declare that this thesis titled "Towards Real-Time Deep Fake Forensic System: An Attention Guided Approach" and the work presented in it are my own and has been generated by me as a result of my own original research.

I confirm that :-

1. This work was done wholly or mainly while in candidature for a Master of Science degree at NUST.
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at NUST or any other institution, this has been clearly stated.
3. Where I have consulted the published work of others, this is always clearly attributed
4. Where I have quoted from the work of others, the source is always given. With the exception of 'Such quotations, this thesis is entirely my own work.
5. 1 have acknowledged all mam sources of help.
6. Where the thesis is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.



ontributed myself.

Maj Rai Sabir Hussain

7.

Dedication

“In the name of Allah, the most Beneficent, the most Merciful"

I dedicate this thesis to Prof Dr Ali Muhammad Ch, who has always remained a source of inspiration and motivation for me in undertaking all endeavors of my life towards academic enlightenment and progress.

Acknowledgments

All praises to Allah for the strengths and His blessing in completing this thesis.

I would like to convey my gratitude to my supervisor, Asst Prof Dr Saddaf Rubab, PhD, for her supervision and constant support. Her invaluable help of constructive comments and suggestions throughout the experimental and thesis works are major contributions to the success of this research. Also, I would thank my committee members; Lt Col Imran, Asst Prof Yawwar Abas Bangash for their support and knowledge regarding this topic. I would also like to thank MCS for providing me with opportunities to work and in my endeavor to contribute towards cyber security in these challenging and information flooded times.

Last, but not the least, I am highly thankful to my family for bearing me out with time and mental peace towards completion of my work.

Abstract

DF is a relatively newer domain of vision research which emerged in recent years. The actual term "DF" was introduced in late 2017 by a Reddit user, where the authors claimed to develop a learning-based algorithm to incorporate celebrity faces into the pornography videos [3]. Considering his attempt as a success, many malicious users tried to appropriate similar principles to conceive social chaos [4,5]. Later, free to access smartphone applications like ZAO and FaceApp also played a prominent role in popularizing such DF tools. These applications encourage mass users to appropriate such DF applications without any prior experience. Most notably, all of these DF tools used social media as a platform for spreading their conspiracies and misinformation within a short span [1]. The growing interest in DF techniques shocked the research community as well as security concerns. As soon it begins to get popularity, several tech giants like Facebook, Google, and Apple took immediate actions to counter DF generators. Simultaneously, well-known security defenses like the Defense Advanced Research Project Agency (DARPA) and the National Institute of Standards and Technology (NIST) commenced developing DF forensic systems along with arranging DF detection competition like Media Forensics Challenge MFC2018) and the DF Detection Challenge (DFDC). All these attempts encouraged this study to investigate further into the DF domain in such a manner that a substantial push can be achieved into the forensic system(s). This research aims at developing a framework to systematically address the issue for prevention and detection of such DFs. A mechanism will be studied/ developed to prevent use of critical image/video data from use by forgers that use AI algorithms to generate DFs.

Table of Contents

Chapter **Page**

List of Figures

Chapter 1

Introduction

1.1 Overview

Intro of high quality video/ image capturing devices and cams into smart phones and other hand held gadgetries have intrigued users to continuously upload captured data onto social media platforms like Facebook, Instagram, Whatsapp etc etc. However, with the fast pace recent developments in AI and AI based application / platforms, shared user data (Videos and images) may be manipulated using various AI based algorithms (specifically GANs) to create forged videos/ images, also known as DFs (DFs). Criminal use of these forged DFs can put every one prone to devastating threats to privacy, reputation and sometimes with financial implications as well. Example scenario could be creating and releasing fake pornographic videos, or creating and releasing a fake statement by higher hierarchy of MNCs having severe financial impacts in market. This threats get more horrific with the fact that a large no of freely available tools and apps (e.g. ZAO FaceApp etc) enables normal users without much of technical know-how of AI or smart systems, to easily manipulate data to create DFs using their everyday smart devices like smart Phones or tabs. Apropos, it necessitates that robust and effective measures should be undertaken to mitigate the risk of DFs and to protect security and privacy of users and their data [9].

With increased recognition of the threats to data privacy and user security involving DFs, the need to develop DF prevention and Detection techniques have seen a surge worldwide both at private and official level. This is evident from the fact that large no of detection methods/ techniques has been developed in recent years, most of which are AI based using deep learning algorithms. Secondly, several large-scale data sets of DFs videos have been made available for researchers and two open challenges related to DF detection have recently been conducted, the DARPA MFC 2018 Synthetic Data Detection Challenge (https://www.nist.gov/itl/iad/mig/media-forensics-challenge-2019-0) and the 2020 DF Detection Challenge (https:// DFdetectionchallenge.ai) conducted with sponsors from technical giants including Microsoft, Facebook, and Amazon. Despite this serious involvement and progress, still a number of challenges stand tall for DF detection in wild data. Most important of these challenges is the fast pace in improvement and development of Deep Neural Networks (DNNs), that are widely used for DF creation [5].

Therefore, it is now pertinent to consider DF prevention with more importance and concentration as the threat is evolving at a pace where detection will slowly become near impossible.

1.2 Motivation and Problem Statement

Capturing and sharing of millions of images and videos round the clock by users on openly accessible social media platforms cannot be decreased or stopped and the only way forward is the endeavor to protect the privacy and integrity of created content. Hence, there is a need for development of a frame work/ systematic approach for effective prevention of user data. After prevention, the utmost requirement is of a mechanism which enables user to detect forged data (DFs) with some level of authenticity.

1.3 Objectives

The main objective of this research is to develop a framework to guide the forensics of DFs. The detailed objectives are:-

1.3.1 Develop DF prevention model for image/ video data.

1.3.2 Develop DF detection model for prevented data (data prevented through prevention model).

1.3.3 Propose a practical way forward for DF detection model for non- DFL data.

1.4 Thesis Contribution

To the best of our knowledge the mechanism proposed in this paper has not been used for handling prevention of DF video and imges.

The main contributions of this work are as follows.

- We propose a DF prevention model/ concept to enable users and organizations to prevent forgery of their data.
- Next, we have proposed a technique to identify the protected data for being DF or otherwise.
- Unlike existing work, our work is more leaned towards normal users that can use simple tools to protect their data against forgery.

1.5 Thesis Organization

The thesis is structured as follows: -

- Chapter 2 contains the literature review and work done in the field so far.
- Chapter 3 contains the proposed work
- Chapter 4 contains the testing of the concept, acquired results and detailed discussion
- Chapter 5 marks the end of the document. The conclusion and future work areas are revealed in this chapter.

Chapter 2

Literature Review

2.1 Threat Assessment

Deep Fakes have emerged as an existential threat at international security horizon with potentials to effect the personal privacy and security of individuals as well as pose a threat to create much larger and serious security concerns between countries, MNCs and other business and financial organization. Brookings Institution has identified that Deep fakes have a potential to Distorting state processes and eroding trust in institutions; weakening journalism; exacerbating social divisions; undermining public safety; and inflicting hard-to-repair damage on the reputation of prominent individuals, including elected officials and candidates for offices. A study by *Forbes* indicates that the consequences of Deep Fake technologies could be catastrophic both at individual as well as collective level. They have depicted a scenario where using Deep Fakes, footages can be created in which soldiers are shown committing humanitarian crimes against civilian population or some powerful ruler of a country may be shown approving a nuclear strike against any country leading to devastating consequences for countries/ humanity. Henceforth, we conclude that the threat posed by Deep Fakes and evolution of AI and its applications is quite real and significant in its potential of causing high magnitude devastations. Therefore, it is imperative to work on mitigation measures against this risk and a systematic and all-inclusive approach be devised for a safe way forward.

2.2 What are Deep Fakes?

The term Deep Fake (to be refered as DF hereon in this paper) appeared in 2017 by a Reddit user (whose nick was also DFs) who carried out image forgery through AI algorithms and created forged pornographic content with swapped faces. Theron, the term got attention of world wide researchers and users, and all the forged videos/images with impersonated characters had been termed as DFs.

Technology involved in Deep Fake creation is capable of seamlessly stitching anyone from real world, into a video or image in which they never actually participated. We can see from the example of actor Paul Walker, who even after his death was shown in Fast & Furious movie [7].

Though initially the DF content appeared as comic or humorous videos, that were mostly created through face swapping of popular personalities, however the criminal or negative uses of the technology like revengeful pornographic DFs, fake news, hoax or financial frauds have raised the concerns of public as well as intelligence and law enforcement personals [2].

2.3 Classification of Deep Fakes

With evolution of AI and related technologies, the quality and techniques for DF have also evolved and three major categories of DF videos have been seen emerging. First is *Head Puppetry,* which means creating a forged video of whole head and upper shoulders of a target user by synthesizing it with the same of a source user. This way the target person's head and upper shoulder region appears to be showing same actions and behaviors as the source user. Second category is *Face Swapping* in which target face is swapped with that of the source and is synthesized using AI so that it keeps the same facial expressions as that of the source. Third category is *Lip Syncing* in which only the lip area of a target is forged and manipulated to create wrong/false speaking gestures or words. This way a video of a person can be created, him saying such words that he actually never said [5].

A comprehensive survey published by ACM Computing Surveys in Jan 2020 has also discussed DF in detail and the researchers have divided the DFs videos into four broad categories based upon visible human characters i.e *Reenactment, Replacement, Editing, and Synthesis.* In Reenactment DF, the source image is used to create synthesized target image having expression, mouth, gaze, pose or body, derived from source image character. Expression reenactment is used to create expressions in target image derived through source image e.g the reenactment of historical figures for educational/ entertainment purposes. Mouth reenactment is used to create dubbed videos with realistic lip/ mouth expressions. Common example of its use are the realistic videos of characters in multiple languages whereas the source video speech has been recorded only in one language. Gaze reenactment is where the eye gaze position and eyelid movements are corrected/ synthesized in a target image using a source image. This is used to correct eye positions during important photographic events or to show a permanent eye contact with camera for a person appearing on a screen whereas actually the eye contact wasn't uniform or constant. In Pose reenactment the head position of the target image is derived from the source image. This technique has commonly been used for carrying out face frontalisation of camera footages where the target image or person is not directly facing into camera direction. In body reenactment the complete body pose of the target image is synthesized/ derived using a source image. This technique is commonly used in media industry to superimpose body poses into a scene where the target character is not actually present.

In Replacement DF the source images are either transferred into target images or they are swapped with them. In image transfer, the desired portion of source images are transferred to target images like in fashion industry different outfits are tried on same or different models without having them to wear the clothes. In swap replacement, commonly the face of the target image is swapped by source image e.g in entertainment memes the faces of actors are replaced with famous personalities. In Editing or Synthesis DF, the appearance characteristics of a target image are altered removed or added. Examples are the apps that are commonly used for image enhancements or editing like changing clothes, hair color, eye color, age, ethnicity or even beauty attributes [7].

To conclude, we define Deep Fake videos or images as such videos and images that contain characters that were not actually doing or saying whatever they appear to be doing or saying.

2.4 Deep Fake Creation

From a technical perspective, the term DF is a combination of terms deep learning and fake images/videos, involving use of AI algorithms and neural networks to deeply train a system to create manipulated images/ videos with people saying or doing things they actually never said or did. DF creation largely rely on neural networks that use large sets of data, mostly comprising source images, to train the system to learn to mimic a person's facial expressions, body movements, poses or speech expressions. Most commonly, Generative Adversarial Neural Networks (GANs) are used for the purpose, which are a pair of artificial neural networks that work together in a competitive framework to create DFs. Both these NNs are trained on the same data set comprising large no of sample of images, videos or voice of the target character, and then one of the network creates a new sample and the other network tries to classify it as real or created. This way both these networks drive each other in improving the results and resultantly create a near real sample of data. Tremendous work is in progress where scientists and researchers are trying to reduce the input data set for creation of more accurate results. This might end up at creation of near realistic DF using a single selfie of a user [1]. Whatever is the algorithm or type and technique of neural networks, it's the AI that's playing a pivotal role in creation of accurate and near real DF. From an end user perspective, although this technology has been existing in entertainment industry for quite some time, where the movie maker would join and stich anybody in any scene of the video. For example, Paul Walker a film actor, who even after his death was included in a movie named Fast and Furious 7. But to do this a lots of efforts in terms of setting up of studio and environment with huge loads of equipment and people was required. However, the evolution of technology in recent decade and availability of machine learning software have enabled the users to create DF in quite simple and fewer steps than ever before [7].

Recent survey conducted by Deeptracelab indicate at least 75 deep learning working models that have been developed to generate body and face reenactment for creating DFs and at least 20 DF creation online communities and forums with upto 100,000 active members are present making it a substantial community and means available [10].

2.5 Deep Fake Detection

Keeping in view the emerging and ever increasing threats posed by Deep Fakes, there is a significant increase in researcher's interest towards development of detection methods and techniques. According to a recent survey almost 55 detection models have been identified that have been proposed by different researches, some of which having an accuracy rate of upto 85 %, however, most of these detection models are data specific and require the system to be pre-trained on data sets for particular users. This makes their viability a bit lesser for normal users with no prior data sets and no experience of AI algorithms [2].

Though, the progress of DF detection is moving on, however, the evolution of DF creation algorithms is so rapid that it indicates that in future normal detection models won't be of any use for large scale DF detection. Availability of data sets by various security organizations has enabled the researchers to produce academic models and studies but any significant progress towards a real time forensic system and model is still awaited [11].

Most of the DF detection models irrespective of their techniques and methods suffer from two basic problems i.e firstly the post processing of videos to remove any artifacts and use of filters makes DFs near realistic and secondly the availability of data sets for researchers is still not formalized and errors and problems exists. Detection based on artifacts, GAN finger prints, biological indicators such as eye blinking pattern or skin wrinkling etc, emotional signatures such as smile pattern, all have achieved some progress towards detection however accuracy and wild data set implementation is still foggy [12].

2.6 Deep Fake Prevention

Since Deep Fake creation has become a very real threat to privacy and security with the rapid evolution of generative adversarial networks (GAN) in image regeneration and synthesis, therefore, needs serious mitigation steps and measures. However, the existing studies, models and techniques being developed for detecting these Deep Fakes are still at nascent stage. Presently the detection methods are struggling to match the speed of evolution and development in NNs and other techniques to evade detection. The existing effectiveness of detection methods is still around classification of some particular data set specific images or videos. However, Deep fake prevention methods can provide a mechanism to track, identify and classify any no of images, if appropriately configured to retain provenance. With this method the fake videos and images can be tracked and identified and further spread on SMNs can be prevented [9].

Passive defending against Deep Fakes have accrued little results and speed of advancement in GANs is likely to keep passive detection much behind active defense of privacy and security has to be undertaken to prevent the creation or at least propagation of false/ forged data. This could best be done by developing some image and video signature or copyright mechanism for embedding some information into image/ video data itself that can be later used for classifying any malicious images/ videos [12].

2.7 Conclusions

Through the study of various surveys, it may be concluded that more efforts are being put up into DF creation techniques and speed of evolution in artificial intelligence and deep learning algorithms might make the job of security personals more and more difficult. Table 1 shows the summary of various DF models identified and here we can see that the no of DF creation models is nearing 80 with those having least data dependencies is around 50, almost 62% of total DF creation working models can create many to many DF. This means that data of any user can be used to create DF videos through synthesizing the user with any other random user by training deep learning neural networks.

Although the study shows a considerable no of DF detection models being used however none of them is data set independent. They are based on neural networks that need to be trained for a specific user data set and after requisite training can only detect the possibility of forged data for that specific user.

Least worked on field is the DF prevention where only a few working models have been developed to protect user data from forged use in DFs.

Purpose	Type	Models	Dataset Specific	Data	Accy Range %	Avg Accuracy
DF Creation	One to One	7	Yes	Video/Image		
	Many to One	17	Yes	Video/Image		
	Many to Many	**52**	**No**	**Video/Image**		
DF Detection	Face Synthesis	9	Yes	“	70-100	93.33
	Identity Swap	17	Yes	“	53-100	77.5
	Attribute Manipulation	10	Yes	“	74-99	92
	Expression Swap	8	Yes	“	75-99	87
DF Prevention	Data Provenance	**3**	-	“		
	AI	**2**	-	“		

Table 1 **Summary of DF Models**

Chapter 3

Proposed Work

3.1 Introduction

This chapter focuses on the explanations of developed system model that will enable the user data to be prevented against forged use for creation of Deep Fakes. Concept of **Deep Fake Lock (DF*l*)** will be introduced, that is a mechanism to be used for locking/ preventing user data from Deep Fake creation. Various system and other diagrams will also be presented to describe the proposed system in detail.

Details of modules developed along with their description shall also be explained in this chapter. Discussion regarding software engineering concepts and implementation will also be explained in this chapter.

3.2 System Description

The proposed system comprises of processes and modules that will enable users to secure their data (Images and videos) for prevention of forged use for creating DFs. In case the data is forged and used for creating DFs, same can be verified through a no of steps using the system. Fig 3.1 shows the common use case scenario of the proposed system.

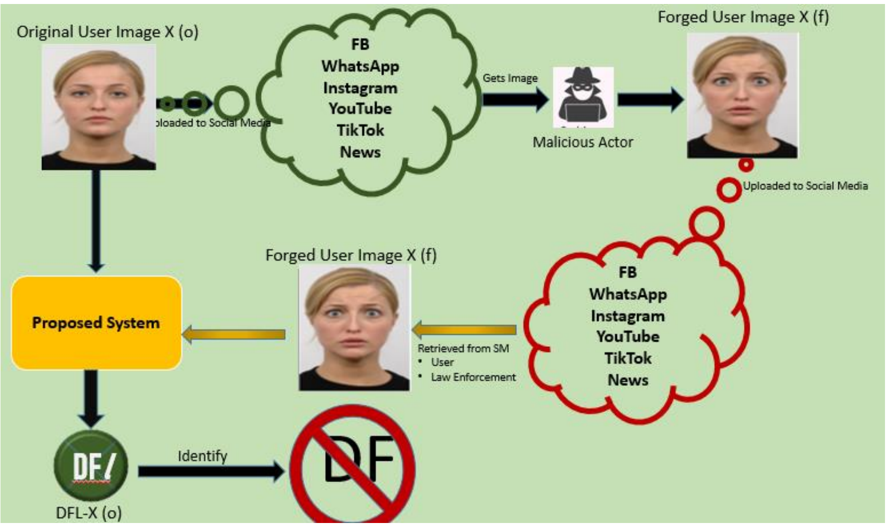


Figure 3.1 General Description and use case scenario of the Proposed System

The system introduces a concept called Deep Fake Locks for prevention of data forgery and a mechanism to identify such forgeries and classify the suspected images/ videos as being Deep Fake or not. Any original user image X (o) if deemed important to be protected through is system is provided by the user. The system creates a unique signature called Deep Fake Lock (DF*l*) for the image X(O). this **DF*l*- X(o)** is archived and secured both with user and within the data itself. If at any point of time during social media propagation, same image is used by a person or organization for creation of forged image X(f), and propagated, it can be retrieved, checked and can be classified by the system. Details of the system are explained in proceeding paragraphs.

3.3 Deep Fake Lock (DFL)

DF*l* is defined as a mechanism that can be used to protect the user data against forgery and unwanted use for creating Deep Fakes. It involves generation of unique signatures for a particular data item i.e. images (both in case of videos and pic) and then securing the generated signature both on the data and off the data. Generated signatures can further be secured through encryption algorithms. DF*l* can be secured On the data by embedding into the meta data of original images or into original images itself using various data hiding techniques. While DF*l* can be secured off the data into a secure file in user's device or in a central online repository.

The concept of DF*l* can be implemented using a no of data encryption and security algorithms, however we have used the **HASH 256** encryption method to generate our RAW signature for the data to be protected. Thus a unique key is generated for every piece of user data i.e user selected images or randomly selected images from user selected videos.

3.4 DFL Creation

DF*l* is created using pixel data of original image, through a series of steps involving mathematical calculations, encryption and data provenance as shown in fig 3.1 and explained in sections below.

3.4.1 User Data Acquisition

User selects the data to be locked or prevented for being used for DF creation. In case of the images, they are directly selected by the user, whereas in case of videos, user selects the video and system selects random images from that video for calculation of signatures. The data about which images are selected is also noted by system along with signatures and same images are selected again while in the identification and classification process at later stages.

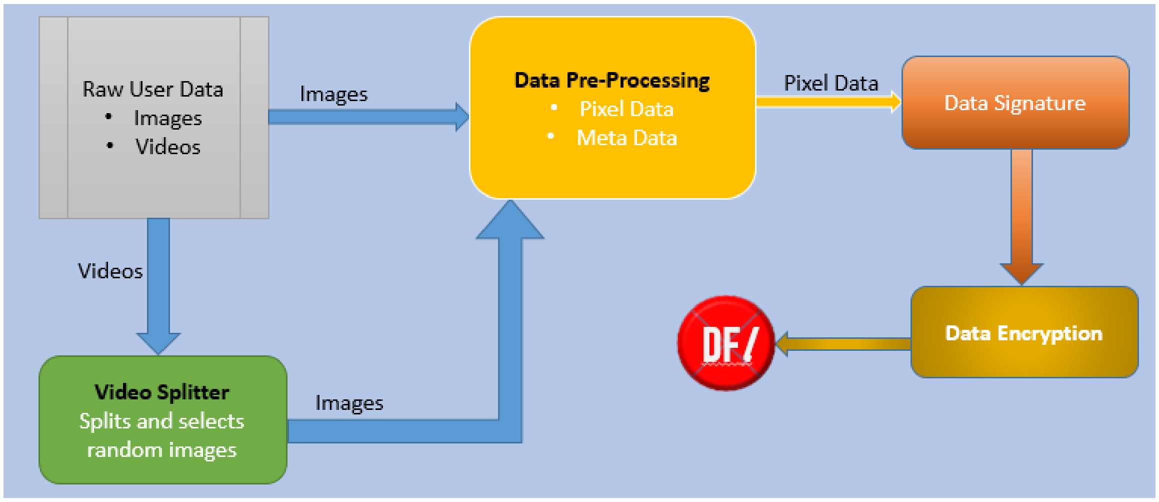


Figure 3.4.1 DF*l* Creation

3.4.2 Data Pre-Processing

After the selection of user data, the images are pre-processed and splitted into meta data and pixel data. Meta data is the raw information about the image including GPS location in case of modern cameras, camera model, lens information, and other details, whereas pixel data contains the raw pixels of image. Pixel data is used for calculation of unique signatures for any particular image.

3.4.3 Signature Calculation

After pre-processing of data, the splitted pixel data is fed into the signature generating process which uses HASH-256 algorithm to calculate a unique hash key for a particular image. It is worthwhile to note that this key is unique for every image with even slightest of changes and thus have a strong capability to detect any image forgery.

3.4.4 Signature Encryption

After the calculation of HASH key for a particular image, the generated key is again encrypted using **AES** encryption. This step further enhances the security of the key from being broken and provides a prevention mechanism for maintaining its integrity. After completion of this step, a DF*l* is created which is a unique signature generated for every image that is required to be protected against forged use or DF creation.

3.5 DF*l* Securing Process

After generating DF*l*, next important step is to secure it both along the data and off the data so as it can be used later for identification process. For our system, we have used two-way data securing mechanism to ensure that the unique data signature is available to users. These are explained in following sections.

3.5.1 On-Data Signature Embedding

For On-Data signature embedding, we have used meta data of the original image which is already available after pre-processing step and hence the unique signature is embedded into LSBs of meta data and same is retrieved during identification process.

3.5.2Off-Data Signature Embedding

DF*l* is archived and secured off the data through two techniques, firstly it is saved in user defined device or storage media, and secondly it is registered with a central certification server online. This will enable to identify DF by end users themselves using locally archived signatures for less harmful cases, and it will also enable establishment of a central data base for forensics and law enforcement agencies to identify DFs that may have consequences related to security and integrity of original data.

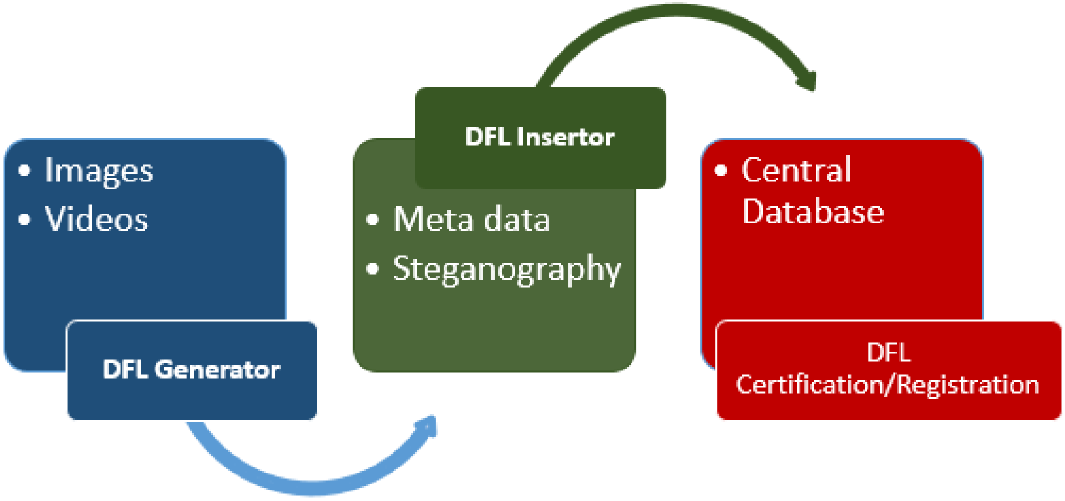


Fig 3.5 DF*l* Securing Process

3.6 DF*l* Authentication

While executing the authentication and validation process, the system again follows series of steps as shown in fig 3.3. Whenever a user or a law enforcement agency suspects that some image or video has been used to create a DF, with similar environment as the user was in original image/ video, same signature is again created for suspected data set (images from suspected DF videos/pics) and is compared with secured on the data and off the data signatures of original image or video. As an example if a user thinks that his portrait picture (that has already been prevented using DF*l*) has been used by some malicious actor to create forged image or video, showing him doing or saying something he never did in his original image/ video, the user can always use the archived signatures to identify and verify.

From security point of view same example can be set up for any president of important country who originally recorded a video or an address to the nation, and same data has been forged to make DF showing him saying things that he never said in original video. So if the original video was protected through DF*l*, the video can be used to re-create same signatures using same techniques and algorithms and will be compared with the archived signatures. Depending upon the results, the new forged video can be classified as not being original or DF. The process involves following steps.

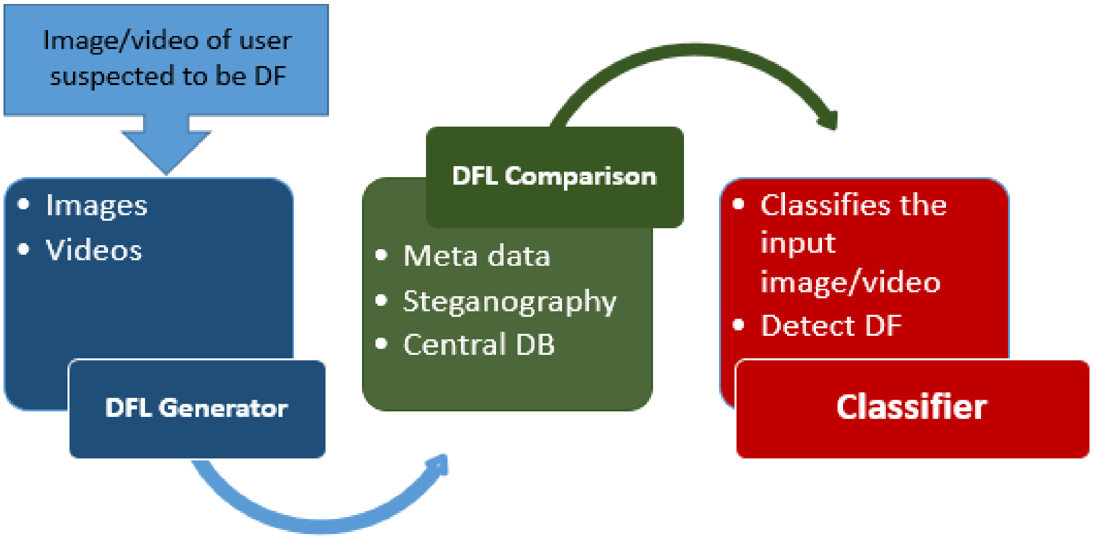


Fig 3.6 DF*l* Authentication

3.6.1 Suspected Data Acquisition

Suspected forged videos can be provided to the system both by users as well as law enforcement agencies, for which same procedure is used again. In case of image, it is directly passed on to the generator module, while in case of a video, same images are selected as that were during creation process. For the time being, we have used manual selection of images, however in future work and advanced version of system, the images will be selected on similarity basis with the original video. As an example, if a user thinks that his 10 seconds video has been used to create a 10 seconds DF, the original index no's of randomly selected images will be available in meta data or archived data. So same indexed images will be selected from the forged video and their signature re-calculation will be carried out and compared with original images.

3.6.2 Signature Re-Generation

After selection of images, same module of signature generation as explained in fig 3.2 , will be used to calculate the data for suspected images. After this process, we will have two signatures of user images calculated through exactly same algorithms.

3.6.3 DF*l* Comparison

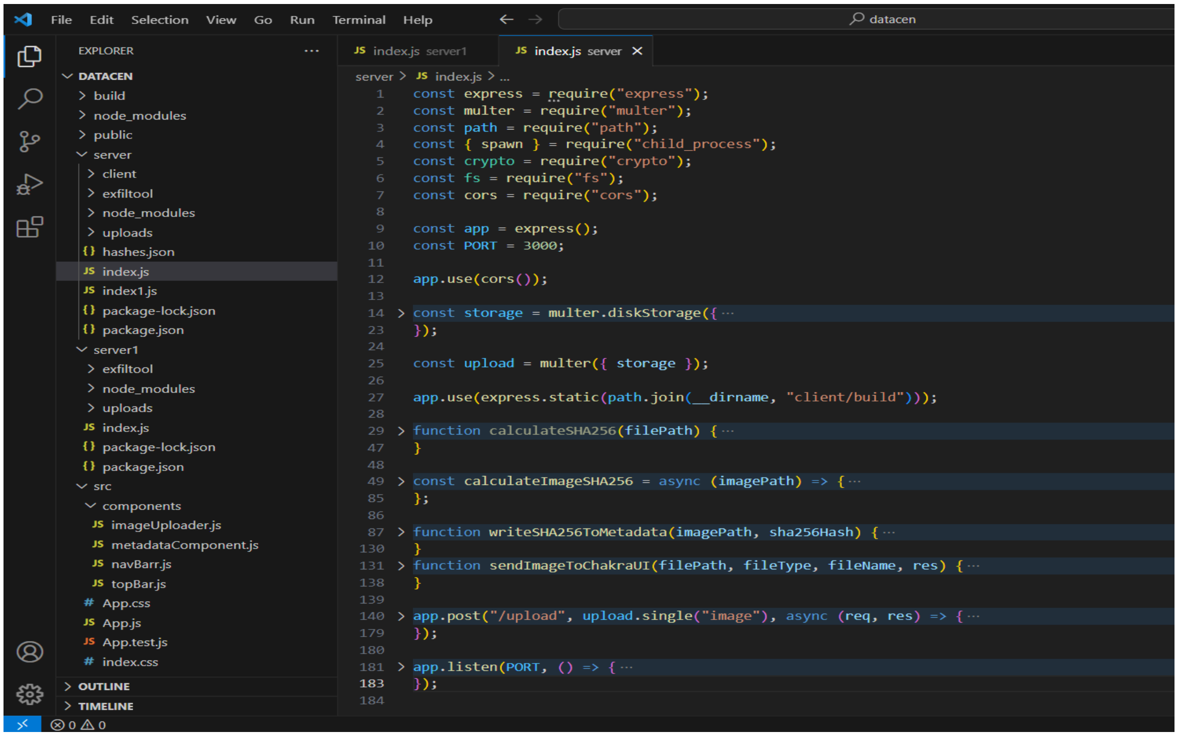
This process carries out comparison of the newly created signatures with those created and archived for original images. Archived signatures can be retrieved from the meta data or user files or from the central data base. These signatures being encrypted in original form, are first de-crypted using same algorithm i.e **ABCD** in this case.

3.6.4 Classification/ Authentication

This process validates the suspected images into either being DF or being original. This provides users and law enforcement with at least the basic foundation stone to fight DF and take a step towards its prevention.

3.7 Code Walkthrough

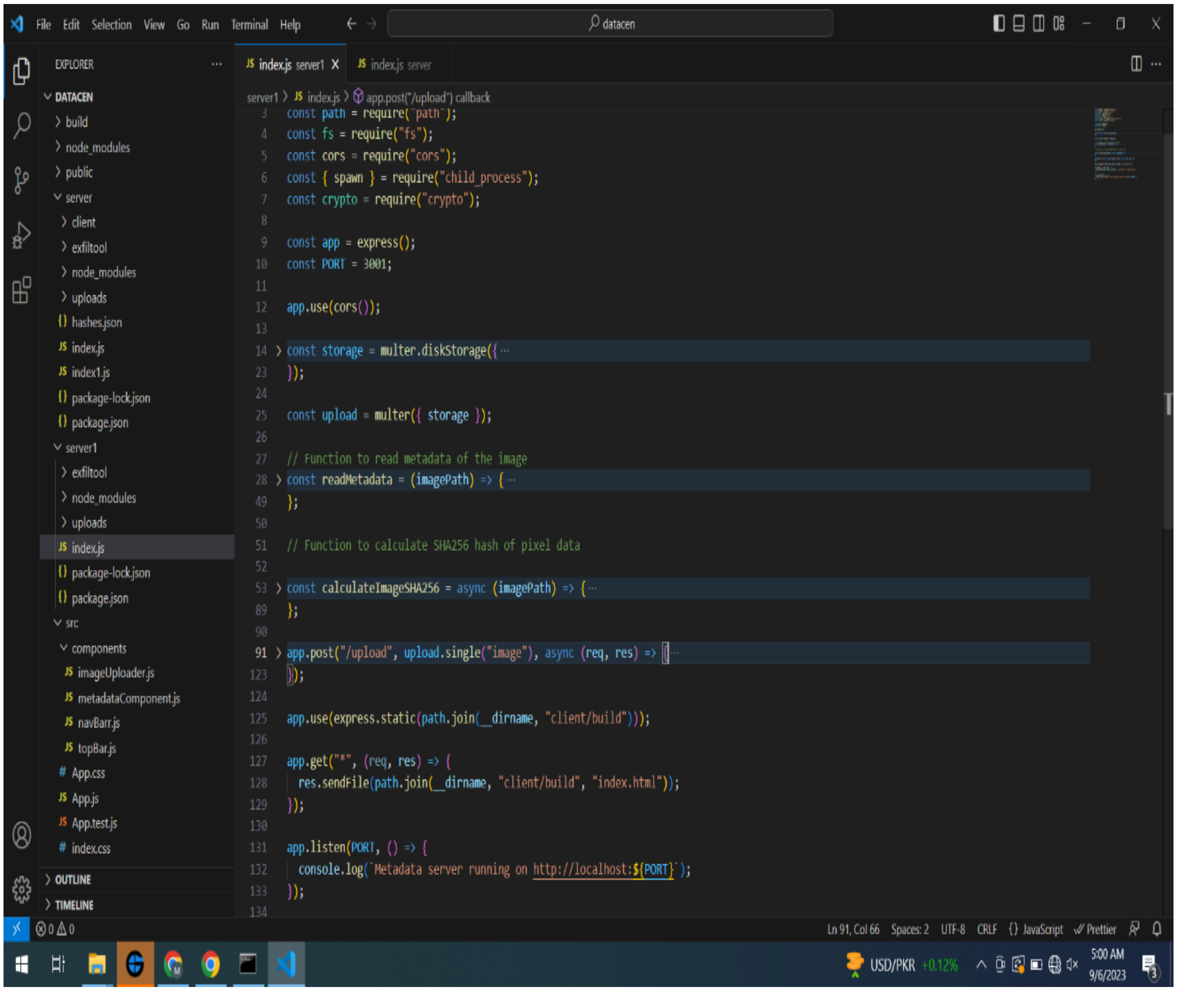
In this section we are going to provide a simple code walk through of the developed modules and application. Figure 3.5 shows the overall functionality of the system. Any image that's to be protected through DF*l* is provided to the system which splits the image into meta data and pixel data. Using pixel data of the image, the function calculates its hash and passes on to the next function that writes it into its meta data.



```
const express = require("express");
const multer = require("multer");
const path = require("path");
const { spawn } = require("child_process");
const crypto = require("crypto");
const fs = require("fs");
const cors = require("cors");

const app = express();
const PORT = 3000;

app.use(cors());

const storage = multer.diskStorage({ ⋯
});

const upload = multer({ storage });

app.use(express.static(path.join(__dirname, "client/build")));

function calculateSHA256(filePath) { ⋯
}

const calculateImageSHA256 = async (imagePath) => { ⋯
};

function writeSHA256ToMetadata(imagePath, sha256Hash) { ⋯
}
function sendImageToChakraUI(filePath, fileType, fileName, res) { ⋯
}

app.post("/upload", upload.single("image"), async (req, res) => { ⋯
});

app.listen(PORT, () => { ⋯
});
```

Fig 3.7 Overall Code Function Description

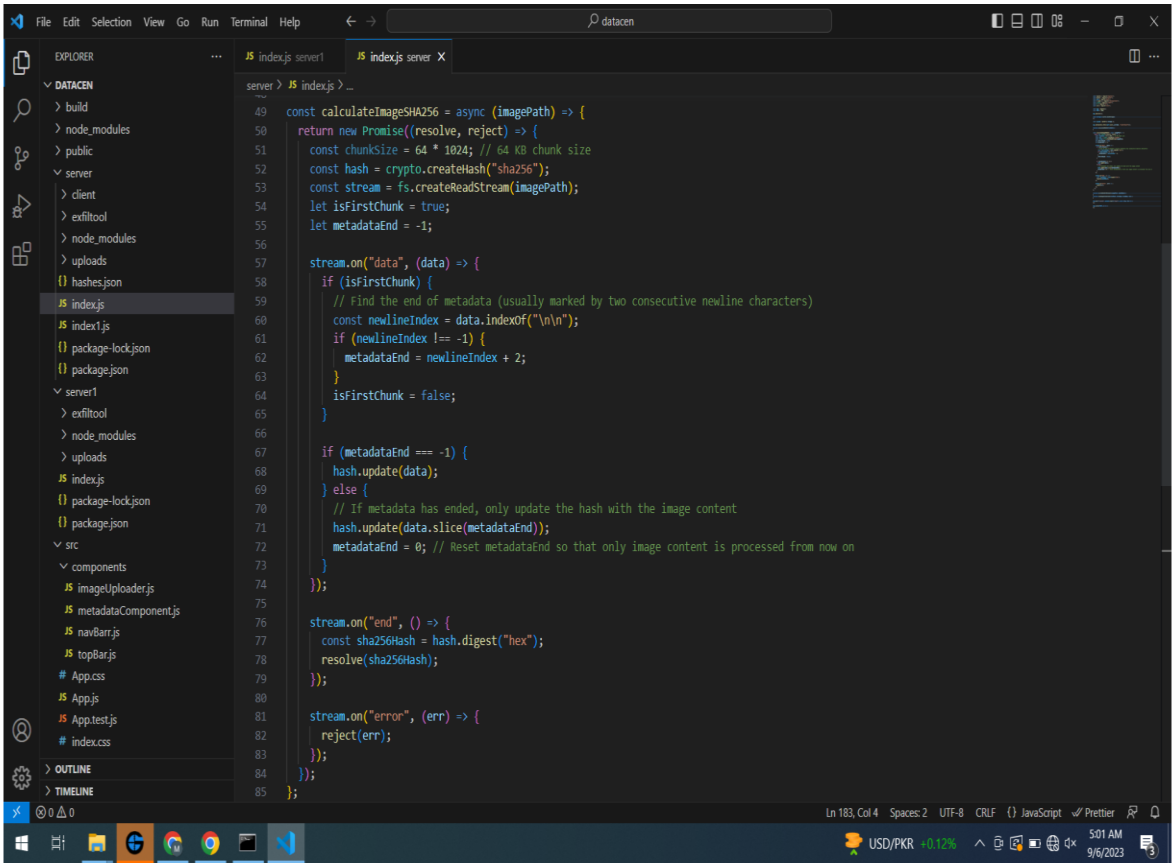
Figure 3.6 shows the detailed working of the function Calc SHA256, through which the desired image is uploaded, its meta data and image data is splitted and SHA256 is calculated for the pixel data. Same data is also displayed in UI where user can also see the calculated hash of the image. This has been incorporated for simulation purpose so as user can also see the DF*I* signatures for manual comparison.



```
const path = require("path");
const fs = require("fs");
const cors = require("cors");
const { spawn } = require("child_process");
const crypto = require("crypto");

const app = express();
const PORT = 3001;

app.use(cors());

const storage = multer.diskStorage({…
});

const upload = multer({ storage });

// Function to read metadata of the image
const readMetadata = (imagePath) => {…
};

// Function to calculate SHA256 hash of pixel data

const calculateImageSHA256 = async (imagePath) => {…
};

app.post("/upload", upload.single("image"), async (req, res) => {…
});

app.use(express.static(path.join(__dirname, "client/build")));

app.get("*", (req, res) => {
  res.sendFile(path.join(__dirname, "client/build", "index.html"));
});

app.listen(PORT, () => {
  console.log(`Metadata server running on http://localhost:${PORT}`);
});
```

Fig 3.7 Function to Generate DFL

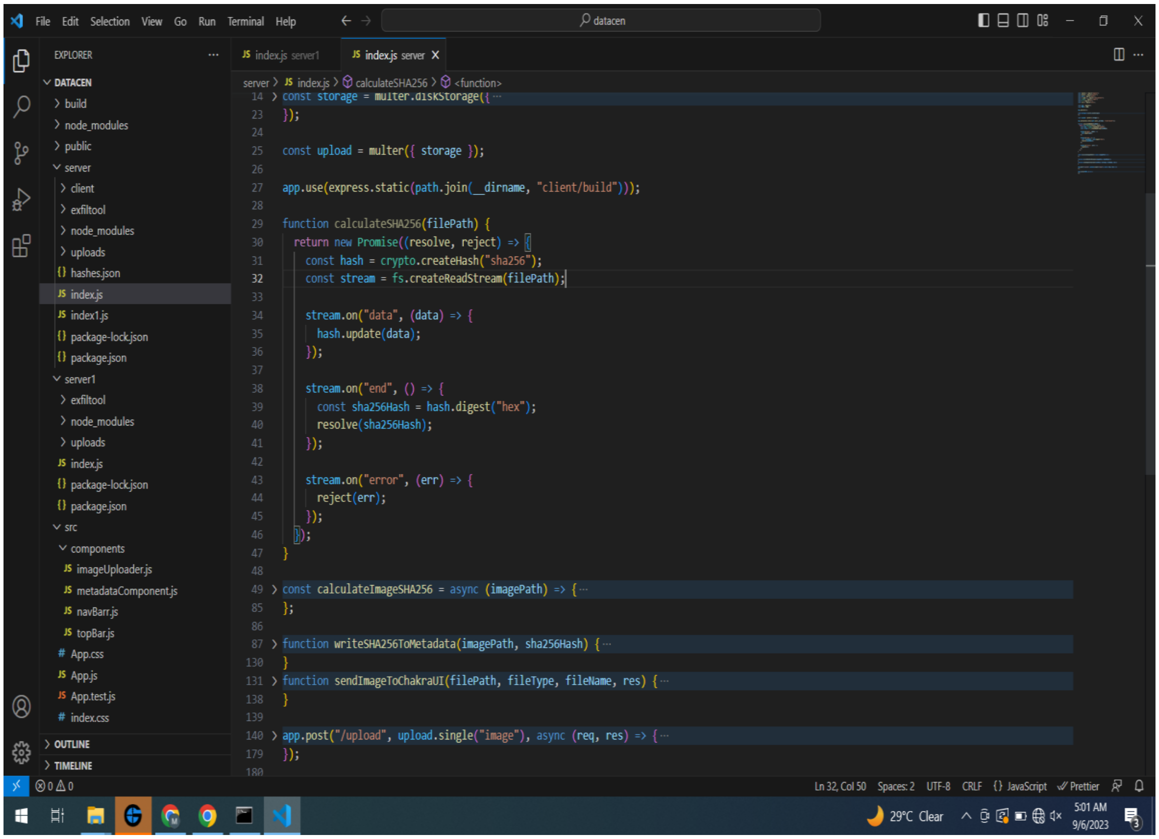
Figure 3.7 shows the function that handles the meta data of the input image. It reads the meta data, and saves it, and when sha-256 is calculated, it embeds the calculated signature into the meta data again for future reference and validation process. For the time being the signature is simply being embedded at the end of the meta data. However, in advanced version of the system in future, same data will be embedded into meta data at bit and byte level so as same cannot be altered or deleted during the social media laundering process where apps automatically delete the unnecessary portion of meta data or sometimes embed totally new meta data. From DF creation point of view, most of the creation tools delete the older meta data all together and embed new entries, making forensic process more difficult. We will discuss the is detail in our discussion section of the paper.



```
const calculateImageSHA256 = async (imagePath) => {
  return new Promise((resolve, reject) => {
    const chunkSize = 64 * 1024; // 64 KB chunk size
    const hash = crypto.createHash("sha256");
    const stream = fs.createReadStream(imagePath);
    let isFirstChunk = true;
    let metadataEnd = -1;

    stream.on("data", (data) => {
      if (isFirstChunk) {
        // Find the end of metadata (usually marked by two consecutive newline characters)
        const newlineIndex = data.indexOf("\n\n");
        if (newlineIndex !== -1) {
          metadataEnd = newlineIndex + 2;
        }
        isFirstChunk = false;
      }

      if (metadataEnd === -1) {
        hash.update(data);
      } else {
        // If metadata has ended, only update the hash with the image content
        hash.update(data.slice(metadataEnd));
        metadataEnd = 0; // Reset metadataEnd so that only image content is processed from now on
      }
    });

    stream.on("end", () => {
      const sha256Hash = hash.digest("hex");
      resolve(sha256Hash);
    });

    stream.on("error", (err) => {
      reject(err);
    });
  });
};
```

Fig 3.7 Function to Split Image Data

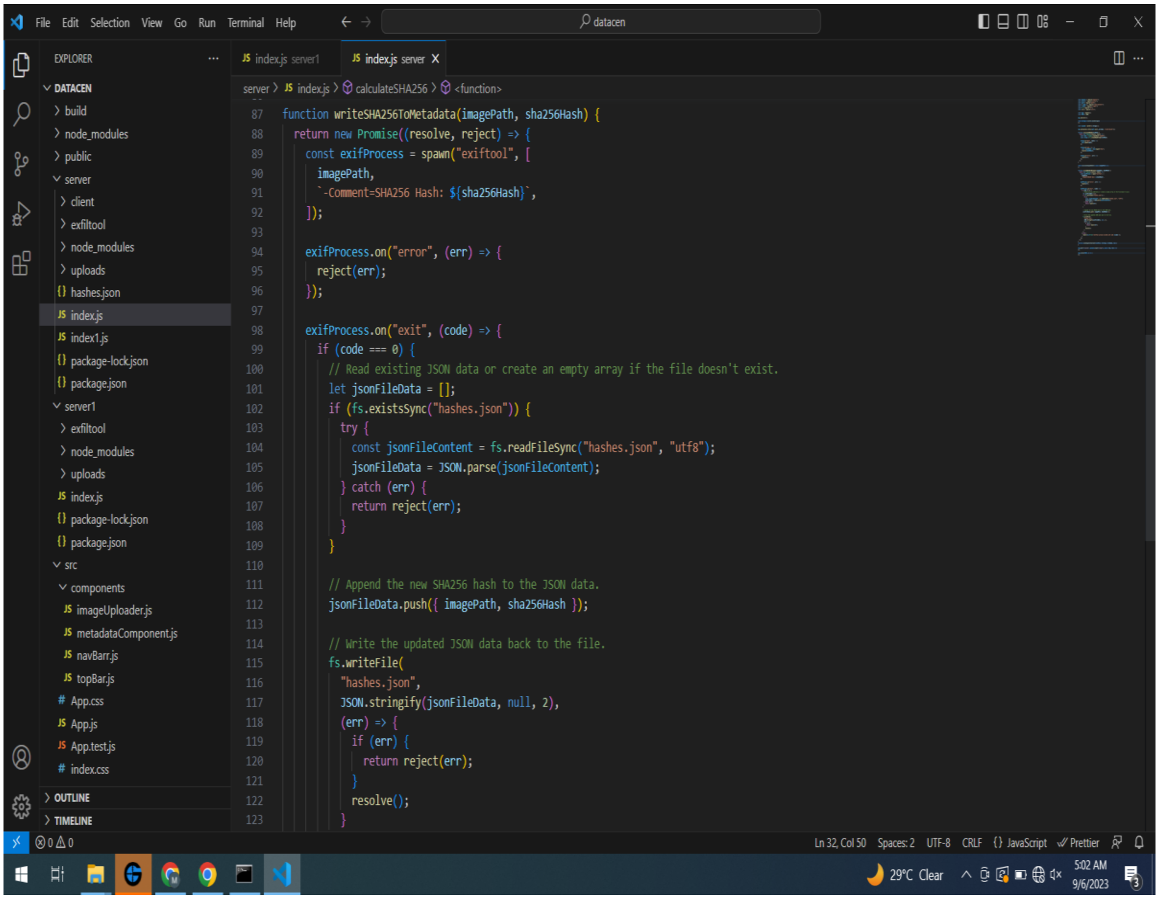
Figure 3.8 shows the function that actually parses the image pixel by pixel and calculates the unique signature using SHA256 algorithm. The unique signatures are calculated using the pixel data so as during forgery, any change on original image pixels can be identified for further classification of the image to be original or not.



```
const storage = multer.diskStorage({ ···
});

const upload = multer({ storage });

app.use(express.static(path.join(__dirname, "client/build")));

function calculateSHA256(filePath) {
  return new Promise((resolve, reject) => {
    const hash = crypto.createHash("sha256");
    const stream = fs.createReadStream(filePath);

    stream.on("data", (data) => {
      hash.update(data);
    });

    stream.on("end", () => {
      const sha256Hash = hash.digest("hex");
      resolve(sha256Hash);
    });

    stream.on("error", (err) => {
      reject(err);
    });
  });
}

const calculateImageSHA256 = async (imagePath) => { ···
};

function writeSHA256ToMetadata(imagePath, sha256Hash) { ···
}
function sendImageToChakraUI(filePath, fileType, fileName, res) { ···
}

app.post("/upload", upload.single("image"), async (req, res) => { ···
});
```

Fig 37 Pixel Data Processing

Figure 3.9 shows the code that handles the signature archiving in DB file that is envisaged to be residing in user selected local storage device. These entries can be used to classify any new image or video that is suspected to be a forged version of the user data that has already been protected through DF*I*. Advanced version of this app will also have an encryption mechanism for securing even the user device data. The app accesses the DB file, creates a new array if now older file exists, appends the calculated signature to the file against indexed entries and updates the file on user device.



```
function writeSHA256ToMetadata(imagePath, sha256Hash) {
  return new Promise((resolve, reject) => {
    const exifProcess = spawn("exiftool", [
      imagePath,
      `-Comment=SHA256 Hash: ${sha256Hash}`,
    ]);

    exifProcess.on("error", (err) => {
      reject(err);
    });

    exifProcess.on("exit", (code) => {
      if (code === 0) {
        // Read existing JSON data or create an empty array if the file doesn't exist.
        let jsonFileData = [];
        if (fs.existsSync("hashes.json")) {
          try {
            const jsonFileContent = fs.readFileSync("hashes.json", "utf8");
            jsonFileData = JSON.parse(jsonFileContent);
          } catch (err) {
            return reject(err);
          }
        }

        // Append the new SHA256 hash to the JSON data.
        jsonFileData.push({ imagePath, sha256Hash });

        // Write the updated JSON data back to the file.
        fs.writeFile(
          "hashes.json",
          JSON.stringify(jsonFileData, null, 2),
          (err) => {
            if (err) {
              return reject(err);
            }
            resolve();
          }
```

Fig 3.9 Pixel Data Processing

Chapter 4

Model Testing, Results and Discussion

4.1 Test Scenario

To test the frame work and concept model, few important personalities have been selected to validate the functionalities. In this basic version of the system video data (which is an array of images) has also been depicted through selection of single images. In advanced versions however, the selection of images from video data will be automated using AI based algorithm that will randomly select images from videos to be protected and calculate the signatures and embed it individually with respective images or collectively with the video itself.

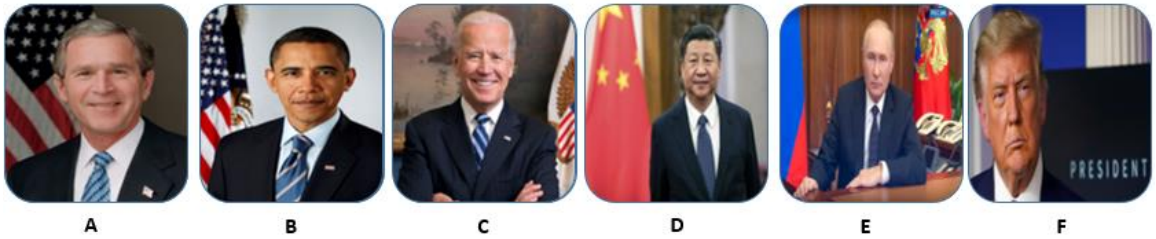


Fig 4.1 Test User Data - Original

Figure 4.1 shows the images that have been selected from different videos of the users whose data is to be protected as user A, B C D E F. These images will first be used to generate their unique signatures i.e DF*l*, and will be secured three ways, in meta data, in user defined device and on an online repository. After protecting the data through proposed concept, same will be used to create forged data through DF creating apps. FACEAPP has been used as test app for creating forged or deep fake images of videos.

4.2 Test Execution

Using the FACEAPP app, the test images have been altered to show the users in actions, they were actually not, the envisaged threat scenario of our study. Different change or forged depictions have been done for different users. Table 4.1 shows the details of alterations that have been done in the data.

Ser	User	Alteration Done	Platform Used	
1	A	Face gesture Altered	FACEAPP	
2	B	New face gesture	“	
3	C	Face tone changed	“	
4	D	New face gesture	“	
5	E	New face gesture	“	
6	F	New face gesture	“	

Table 4.1 Test Alterations

in case of user A, the original picture already contained the user with a smiling face, and same has been slightly altered with a deepened smile, that may go un noticed while watching this through a video. These changes may have different implications when comes to international diplomacy and sign language.

In case of user B, the user is serious and worried in his original image however the gesture has been changed to a relaxed and smiling face. User C can be seen with a high tone skin and clean looking textures however the altered image contains slightly changed skin tone with addition of freckles on the face which is a common phenomenon in the ethnicity of the user. User D also had a normal face gesture on a normal working day however we have altered the gesture to a very happy and relaxing mode. User E is also shown with a normal face gesture in original image however we have changed the gesture to a happy mode again. In case of user F, the original image shows user in a slightly anguished and disturbed state, whereas again using the DF app, the user is shown smiling and in a relaxed mode.

Fig 4.2 shows the processed data for same users with the altered or forged facial gestures. This gesture alteration of images will be same as in the case of forged videos where original images in sequence will be altered to show the user saying and doing things he never actually said or did.

After creation of forged images, the same images are then again provided to the system which carries out 3 way authentication process. It calculates the signature of the suspected image using same algorithm as it used during DF*l* generation process. After generating the data for new image the system than looks up for the archived DF*l* embedded in the meta data of the image and in case of social media laundering, the system checks for the signatures in user device DB and online repository. Both the DF*l*s are then compared to classify the suspected image.

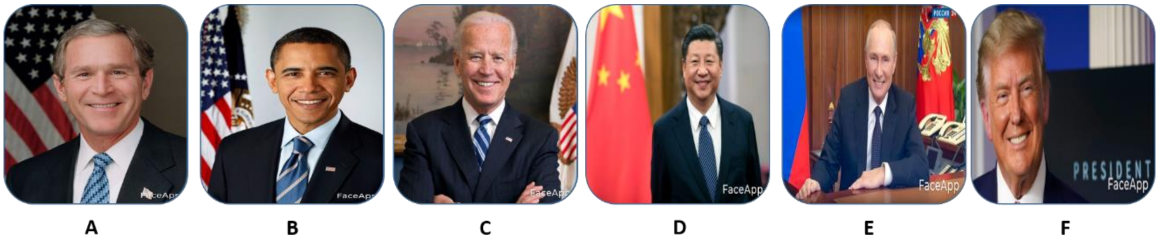


Fig 4.2 Test User Data - Forged

4.3 Test Results

Using the above mentioned test scenarios and data, the proposed concept is evaluated. Each user image was forged to create DFs and were classified based on archived DF*l*. In most of the cases, the meta data of the image was also altered through the DF creation app and hence required to refer to the DB where respective DF*l* was archived in user device and online repository. Table 4.2 shows the detailed test results.

Ser	**User**	**Alteration Done**	**Validation**	**Y/N**
1	A	Face gesture Altered	Meta Data	Y
2	B	New face gesture	Local DB	Y
3	C	Face tone changed	“	Y
4	D	New face gesture	“	Y
5	E	New face gesture	Online DB	Y
6	F	New face gesture	“	Y

Table 4.2 Test Results

Same forged images were also uploaded on SM sites such as Facebook and WhatsApp which automatically carry out social media laundering, mostly changing and reducing the meta data. When these images were retrieved after propagation of about a week, the meta data of the images was altered thus the classification process could not be validated through DF*l* embedded in meta data and the same was retrieved from the local DB/ online repository. Fig 4.3 shows the overall result for the sources of authentication after the data was subjected to SM propagation.

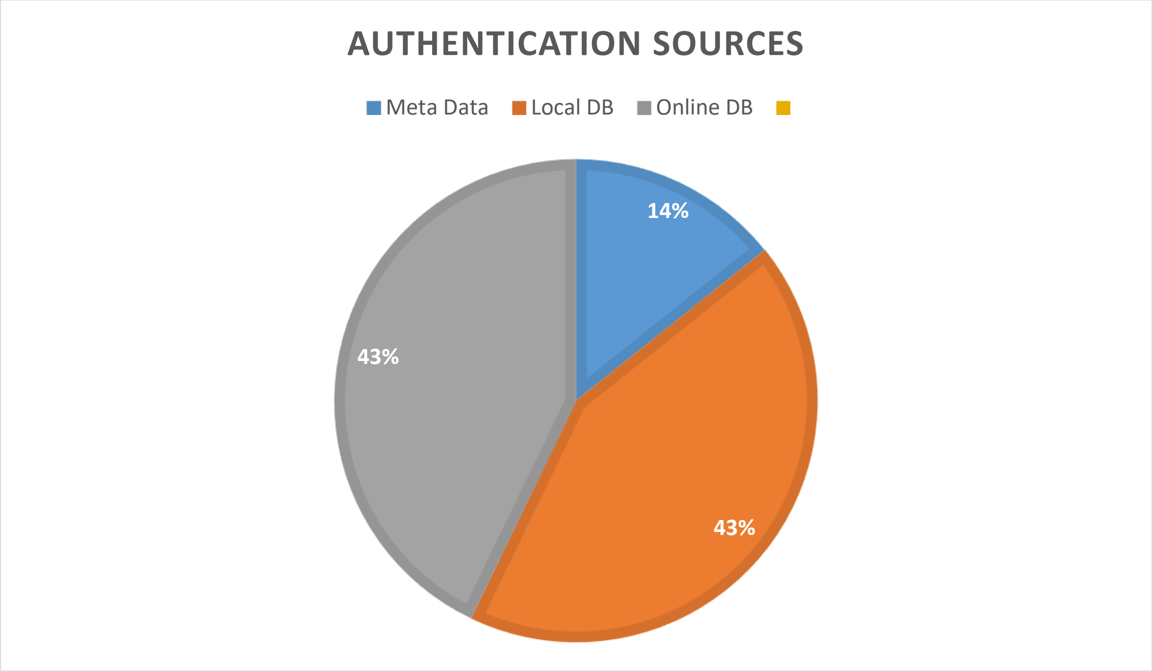


Fig 4.3 Authentication Sources

Moreover, the test results for the test data remained satisfactory for normal use cases, however the false positives increased with the change in size and resolution of same images without any forgery. Similarly, if any noise is added to the original image while transmitting or propagation on social media, again the false positives are high.

Ser	Change	Classified as DF	False Positive	%age
1	Resolution	Y	Y	100
2	Filters	Y	Y	100
3	Noise (AI Generated)	Y	Y	100
4	Platform Noise	Y	Y	100
5	Back ground change	Y	Y	100

4.4 Discussion

Although the proposed concept provides a basic stepping stone towards prevention and detection of DFs however the implementation is still through basic algorithms and techniques. The speed of evolution in AI and its applications requires that a robust implementation of the proposed concept may be undertaken to address the issue with more accuracy and effectiveness.

4.4.1 Data (Image) Selection

The selection of still images can be done manually however the selection of images from videos needs to be automated for both selecting the amount of data and images to be selected. This sequence of images with their index Nos will also be saved within DF*l* for later referral and selection of same images again to validate the suspected video.AI based image similarity algorithms be used in advanced version to quickly identify that which original video has been used to create any suspected DF. Same data will be used to then identify the archived DF*l* of original data for classification and validation process.

4.4.2 DF*l* Generation

We have used the SHA 256 Hash function to calculate the unique signature for images and videos. Though the robustness of the algorithm is still trusted by wide users and researchers however AI based technique may be developed to create DF*l* with more robust and unbreakable.

4.4.3 DF*l* Archiving within Data

Online embedding of DF*l* is currently being carried out into meta data of the images and videos. However, test result shows that most of the social media platforms carryout laundering of the meta data and reduce it or sometimes totally removes it and puts in new meta data. Therefore, some robust stenographic technique may be used in advanced version so as to cater for meta data laundering and availability of the DF*l* within the image data itself. This will enable users to use the suspected image itself for identification and classification. Felix Juefei-Xu et al proposed an AI based image provenance model that embeds the signatures within the image data, using GAN based implementation [9]. Our work combined with this can create a more robust and readily available data repository for classification of suspected DFs.

4.4.4 DF*l* Online Repository/ Certification Authority

We have implemented a user based online repository for archiving the DF*l*s however to address the issue systematically a central authentication or certification authority may be implemented providing users with a mechanism to protect the data and also provides a mechanism to all social media platforms to identify and mitigate propagation of DF videos and images. This might require some legislative steps to be taken by international cyber security organizations that will compel the platforms to authenticate the user data with this certification authority, whenever any new image or video is propagated. This can provide an automated mechanism for identifying and removing any DFs that are created using already uploaded data of users. At user level also, this can be implemented through subscriptions or certificate procurement like is being done through https in web based applications. Use of block chain technologies for creating such certification vaults can further increase the effectiveness of the concept and provide defensive mechanisms against hacking or forgery attempts.

4.4.5 Addressing the False Positives

Since the test results shows that for certain use case scenarios, like image resolution changes or resizing etc creates 100 % false positives hence the advanced version of the system might consider creating DF*l* for all available resolutions of the data to cater for the false positives of image size and resolution changes. However, keeping in view the severity of threat being imposed by the DF, we can ignore the false positives.

Chapter 5

Conclusions and Future Work

5.1 Conclusions

Deep Fake videos and images are appearing to be the new threat in modern info warfare and security environment. The speed of evolution in AI and its applications is adding to the threat at very fast pace. As it is evident from this study as well, that a large no of researchers are contributing towards betterment of AI based GANs for creating DFs. Though the evolution is being done for creative and entertainment purposes however the nefarious use by both state sponsored and independent actors may result into high profile security incidents. Though some work has been done for detection of DF already propagated onto SM however most of the work is data set specific so it cannot be implied on billions of users spread across the world. Therefore, as it is said in IT that Prevention is cheaper than the cure, hence to take the first step today, there is a need to develop a mechanism for users as well as institutions to star protecting the data to prevent future misuse. However, this study has also found out that very less amount of work is being done in the prevention domain. Thus this study can provide a stepping stone for building an all-encompassing robust prevention detection and forensic system to address this emerging threat. The speed with which AI is getting out of hands may make it late by tomorrow even.

5.2 Future Work

This conceptual frame work for a DF forensic system that's based on DF*l*s may be implemented formally as a robust working system involving all steps of legislation and involvement by all stake holders. Use of AI in creating and archiving DF*l*s will further increase the dependability and effectiveness of the system. Following are a few further research directions that will achieve this for the system: -

5.2.1 AI based image selection from videos to calculate signatures.

5.2.2 Use of robust stegnographic techniques incorporating AI to achieve data locking.

5.2.3 HW based implementation including development of micro controllers to generate and embed DF*l*s into the data as and when captured by various devices i.e cameras, mob phones, commercial apps etc.

5.2.4 Implementing an all size all resolution based functions to cater for false positives.

5.2.5 Legislative steps to be taken to mitigate the threat including preventing un necessary social media laundering and meta data removal by various SM sites and apps.

5.2.6 Establishment of an international certification/ authentication authority incorporating stakeholders from all domains i.e users, governments, security organizations, optical device manufacturers and social media platform owners and developers.

References

[1]. Westerlund, Mika. "The emergence of DF technology: A review." Technology Innovation Management Review 9.11 (2019).

[2.] Tolosana, Ruben, et al. "DFs and beyond: A survey of face manipulation and fake detection." arXiv preprint arXiv:2001.00179 (2020).

[3.] BBC Bitesize, “DFs: What Are They and Why Would I Make One?” 2019. [Online]. Available: https://www.bbc.co.uk/ bitesize/articles/zfkwcqt

[4.] H. Farid, “Image Forgery Detection,” IEEE Signal Processing Maga- zine, vol. 26, no. 2, pp. 16–25, 2009 Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, “Generative Adversarial Nets,”

[5.] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." Proceedings of the IEEE conference on computer vision and pattern recognition. 2018.

[6.] DF Detection: Current Challenges And Next Steps by *Siwei Lyu.*published in IEEE Explore 2020.

[7.] DFs and Beyond: A Survey of Face Manipulation and Fake Detection by *Ruben Tolosana, Ruben Vera-Rodriguez , Julian Fierrez , Aythami Morales , Javier Ortega-Garcia*

[8.] Deep Learning for DFs Creation and Creation- Svy.

[9.] DeepTag_Robust_Image_Tagging_for_DF_Provenan.

[10.] The sate of DFs by Ajdr et al, Deeptracelabs.

[11] Celeb-DF - A large data set for DF Forensics by Yuezun Li, Pu Sun, Honggang Qi, and Siwei Lyu,.

[12.] Deep Fake Generation and Detection Issues, Challenges, and Solutions Sonia et al IEEE 23.

[13] DFs Deceptions, mitigations, and opportunities by Mekhail Mustak science direct 23.

[14] Deep Learning for Deepfaces and Detection A Survey.by Thanh Thi Nguyena, Quoc Viet Hung Nguyenb, Dung Tien Nguyena, Duc Thanh Nguyena, Thien Huynh-Thec, Saeid Nahavandid, Thanh Tam Nguyene, Quoc-Viet Phamf, Cuong M. Nguyeng at al

[15] Security Analysis of SHA-256 and Sisters. A Survey by Henri Gilbert1 and Helena Handschuh.

[16] The Evaluation Report of SHA-256 Crypt Analysis Hash Function A Survey by A.Arul Lawrence Selvakumar 1 ,C.Suresh Ganandhas 2 at al.

[17] Deep Fake Generation and Detection issues challenges and solutions A Survey by Sonia Salman and Jawwad Ahmed Shamsi at all.

[18.] Carlini, N.; and Farid, H. 2020. Evading DF-Image Detectors with White-and Black-Box Attacks. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 658–659.

[19.] He, Z.; Zuo, W.; Kan, M.; Shan, S.; and Chen, X. 2019. AttGAN: Facial attribute editing by only changing what you want. IEEE Transactions on Image Processing 28(11): 5464–5478.

[20.] Huang, Y.; Juefei-Xu, F.; Guo, Q.; Xie, X.; Ma, L.; Miao, W.; Liu, Y.; and Pu, G. 2020a. FakeRetouch: Evading DFs Detection jvia the Guidance of Deliberate Noise. arXiv preprint arXiv.

[21.] Huang, Y.; Juefei-Xu, F.; Wang, R.; Guo, Q.; Ma, L.; Xie, X.; Li, J.;Miao, W.; Liu, Y.; and Pu, G. 2020b. FakePolisher: Making DFs More Detection-Evasive by Shallow Reconstruction. arXiv preprint arXiv:2006.07533.

[22.] Huang, Y.; Juefei-Xu, F.; Wang, R.; Guo, Q.; Ma, L.; Xie, X.; Li, J.; Miao, W.; Liu, Y.; and Pu, G. 2020c. FakePolisher: Making DFs More Detection-Evasive by Shallow Reconstruction. In Proceedings of the ACM International Conference on Multimedia (ACM MM).

[23.] Huang, Y.; Juefei-Xu, F.; Wang, R.; Guo, Q.; Xie, X.; Ma, L.; Li, J.; Miao, W.; Liu, Y.; and Pu, G. 2020d. FakeLocator: Robust Localization of GAN-Based Face Manipulations. arXiv preprint arXiv:2001.09598.

[24.] Amir Jamaludin, Joon Son Chung, and Andrew Zisserman. 2019. You said that?: Synthesising talking faces from audio. International Journal of Computer Vision (2019), 1–13.

[25.] Tero Karras, Samuli Laine, Miika AiŠala, Janne Hellsten, Jaakko Lehtinen, and Timo Aila. 2019. Analyzing and improving the image quality of stylegan. arXiv preprint arXiv:1912.04958 (2019).

[26.] Hasam Khalid and Simon S Woo. 2020. OC-FakeDect: Classifying DFs Using One-Class Variational Autoencoder. In Proceedings of the IEEE/CVF Conference on Computer Vision and Paˆern Recognition Workshops. 656–657.

[27.] Xurong Li, Kun Yu, Shouling Ji, Yan Wang, Chunming Wu, and Hui Xue. 2020. Fighting Against DF: Patch&Pair Convolutional Neural Networks (PPCNN). In Companion Proceedings of the Web Conference 2020. 88–89.

[28.] Yuezun Li, Xin Yang, Pu Sun, Honggang Qi, and Siwei Lyu. 2019. Celeb-DF: A New Dataset for DF Forensics. arXiv preprint:1909.12962 (2019).

[29.] Yuezun Li, Xin Yang, Baoyuan Wu, and Siwei Lyu. 2019. Hiding Faces in Plain Sight: Disrupting AI Face Synthesis with Adversarial Perturbations. arXiv preprint arXiv:1906.09288 (2019).

[30.] Luca Guarnera, Oliver Giudice, and Sebastiano Battiato. Fighting DF by exposing the convolutional traces on images. IEEE Access, 8:165085–165098, 2020.

[31.] Todd K Moon. The expectation-maximization algorithm. IEEE Signal Processing Magazine, 13(6):47–60, 1996.

[32.] Jun-Yan Zhu, Taesung Park, Phillip Isola, and Alexei A Efros. Unpaired image-to-image translation using cycle-consistent adversarial networks. In Proceedings of the IEEE international conference on computer vision, pages 2223–2232, 2017.

[33.] Ke Li, Tianhao Zhang, and Jitendra Malik. Diverse image synthesis from semantic layouts via conditional imle. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 4220–4229, 2019.

[34.] Luciano Floridi. Artificial intelligence, DFs and a future of ectypes. Philosophy & Technology, 31(3):317–321, 2018.

[35.] Davide Cozzolino, Justus Thies, Andreas Rossler, Christian ¨Riess, Matthias Nießner, and Luisa Verdoliva. ForensicTransfer: Weakly-supervised domain adaptation for forgery detection. arXiv preprint arXiv:1812.02510, 2018.

[36.] Francesco Marra, Cristiano Saltori, Giulia Boato, and Luisa Verdoliva. Incremental learning for the detection and classification of gan-generated images. In 2019 IEEE International Workshop on Information Forensics and Security (WIFS),
pages 1–6. IEEE, 2019.

[37.] Shehzeen Hussain, Paarth Neekhara, Malhar Jere, Farinaz Koushanfar, and Julian McAuley. Adversarial DFs: Evaluating vulnerability of DF detectors to adversarial examples. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, pages 3348–3357, 2021.

[38.] Nicholas Carlini and Hany Farid. Evading DF-image detectors with white-and black-box attacks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 658–659, 2020.

[39.] H. Alexandros, K. Vougioukas, S. Petridis, and M. Pantic, "Lips don't lie: A generalisable and robust approach to face forgery detection," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., 2021, pp. 5037–5047.

[40.]. A. Shruti, H. Farid, O. Fried, and M. Agrawala, "Detecting deep-fake videos from phoneme-viseme mismatches," in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops, 2020, pp. 2814–2822.

[41.] N. Yuval, T. Hassner, and Y. Keller, "FSGANv2: Better subject agnostic face swapping and reenactment," IEEE Trans. Pattern Anal. Mach. Intell., vol. 45, no. 1, pp. 560–575, Jan. 2023.

[42.] J. S. Chung and A. Zisserman, "Out of time: Automated lip sync in the wild," in Proc. Asian Conf. Comput. Vis., 2016, pp. 251–263.

[43] Y. Peipeng, J. Fei, Z. Xia, Z. Zhou, and J. Weng, "Improving generalization by commonality learning in face forgery detection," IEEE Trans. Inf. Forensics Secur., vol. 0, pp. 547–558, 2022.

[44.] K. Janavi, C. Joshi, B. Yenarkar, S. Suratkar, and F. Kazi, "A deep learning framework for audio DF detection," Arabian J. Sci. Eng., vol. 47, no. 3, pp. 3447–3458, 2022.

[45.] L. Yuezun, X. Yang, P. Sun, H. Qi, and S. Lyu, "Celeb-DF: A large-scale challenging dataset for DF forensics," in Proc. IEEE/CVF Conf. Comput. Vis.Pattern Recognit., 2020, pp. 3204–3213.

[46] A. Rossler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies,and M. Nießner, "FaceForensicsþþ: Learning to detectmanipulated facial images," in Proc. IEEE Int. Conf. Comput. Vis., 2019, pp. 1–11.

[47.] A. Malik, M. Kuribayashi, S. M. Abdullahi, and A. N. Khan, "DF detection for human face images and videos: A survey," IEEE Access, vol. 10, pp. 18757–18775, 2022.