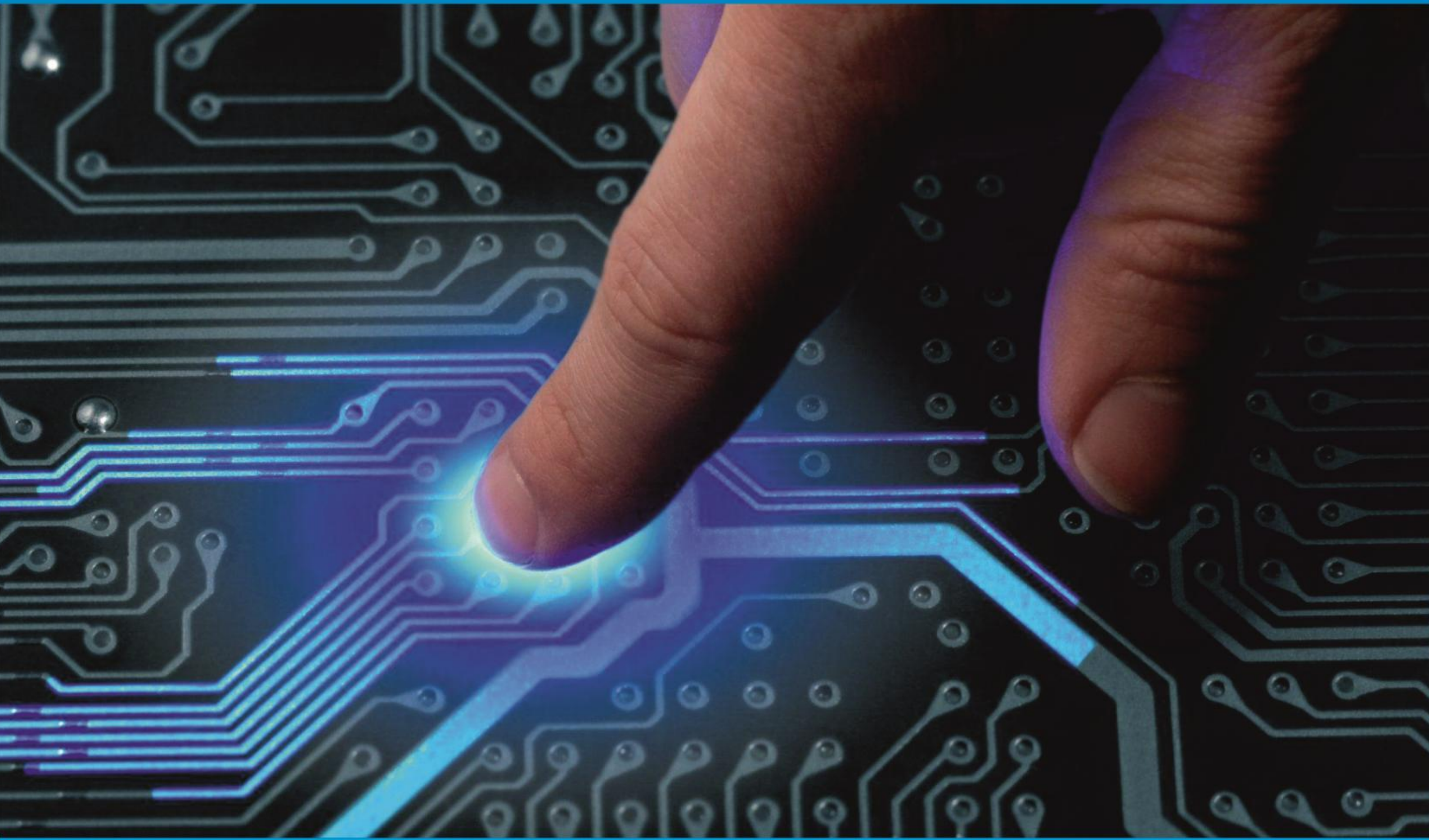
IJIRCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 9, Issue 5, May 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.488

International Journal of Innovative Research in Computer and Communication Engineering

| e-ISSN: 2320-9801, p-ISSN: 2320-9798| www.ijircce.com | |Impact Factor: 7.488 |

|| Volume 9, Issue 5, May 2021 ||

| DOI: 10.15680/IJIRCCE.2021.0905222 |

A Survey on Deepfake Detection Techniques

Akshay C Pednekar[1], Dr Ashwini K. B[2]

PG Student, Dept. of MCA, R V College of Engineering, Bengaluru, India

Associate Professor, Dept. of MCA, R V College of Engineering, Bengaluru, India

ABSTRACT: Recent technological advancements have made it simple to make what is currently called "deepfakes", hyper-sensible recordings utilizing face trades that leave little hint of control. Deepfakes are the result of man-made consciousness (AI) applications that union, consolidate, supplant, and superimpose pictures and video clasps to make counterfeit recordings that seem credible. Deepfake innovation can create, for instance, a diverting, obscene, or political video of an individual saying anything, without the assent of the individual whose picture and voice is included. The game-changing variable of deepfakes is the extension, scale, and complexity of the innovation in question, as nearly anybody with a PC can manufacture counterfeit recordings that are essentially vague from legitimate media. While early instances of deepfakes zeroed in on political pioneers, entertainers, jokesters, and performers having their appearances meshed into pornography recordings, deepfakes later on will probably be increasingly more utilized for retribution pornography, harassing, counterfeit video proof in courts, political harm, psychological militant promulgation, coercion, market control, and phony news.In this paper, I'll momentarily investigate a couple of methods for recognizing deepfakes.

KEYWORDS: Deepfake, Deep learning, Algorithm, SVM, CNN, GAN, PRNU

I.INTRODUCTION

Deepfakes are pictures or recordings that are artificially controlled or produced utilizing Deep learning. The term is filling in notoriety, yet an accurate definition is hard to nail down. It is in some cases utilized for "face trades,""demeanor/property control," and in any event, for pictures completely integrated by a profound learning calculation. The disclosure of deepfake caused enormous media consideration and countless new deepfake recordings started to arise from there on. In 2018, BuzzFeed delivered a deepfake video of previous President Barack Obama giving a discussion regarding the matter of State. The video was made utilizing the Reddit client's product (FakeApp), and it raised worries over wholesale fraud, pantomime, and the spread of deception via web-based media. In this paper, I'll utilize the term deepfake to allude to both media control and media age through profound learning.In spite of the fact that it's workable for people to recognize deepfakes, it's getting harder as the overall quality increments. Also, as it gets simpler and quicker to make excellent deepfakes, the volume of deepfake substance might be excessively incredible for human location alone. That is the reason we may have to depend on machine Learning Algorithms to decide if a piece of substance is a deepfake or not.While spreading false information is easy, correcting the record and combating deepfakes is harder. To fight against deepfakes, we need to understand deepfakes and the different techniques available to detect them. Luckily, deepfake research is progressing, and offers various likely advantages. In addition to the fact that algorithms are programmed, they can conceivably identify prompts that are difficult for people to discover all alone.

II.DEEPFAKE DETECTION TECHNIQUES

In this paper, I've coordinated deepfake identification strategies into the accompanying three general techniques:

- Hand created features
- Learning based features
- Relics

A. HAND CREATED FEATURES

While deepfake procedures make sensible looking substance, there are frequently clear imperfections that a human or calculation can spot after looking into it further. One model is unnatural facial highlights. In this class, we'll two or three procedures that concentrate and architect explicit highlights to perform deepfake identification. Shruti Agarwal et al. proposed a strategy including an oddity location model (here, a one-class SVM) that can recognize an individual of premium (POI) from others just as deepfake impersonators. This system is very fascinating as we just need legitimate recordings of the POI to prepare the model[2]. The creators' speculation is that as an individual talks, they have unmistakable (yet presumably not extraordinary) looks and developments. The scientists utilize the OpenFace2 tool compartment to separate facial and head developments in a given video. By gathering facial activity units, head revolutions about specific tomahawks and 3D distances between certain mouth tourist spots, they get 20 facial/head highlights for a given 10 second video cut. At last, by registering the Pearson relationship between over 20 facial/head highlights, they get a 190 dimensional component vector addressing the 10 second clasp. When the 190-D element vector is removed, a one-class support vector machine (SVM) is utilized to decide if the 10 second video cut is a genuine video of the POI.

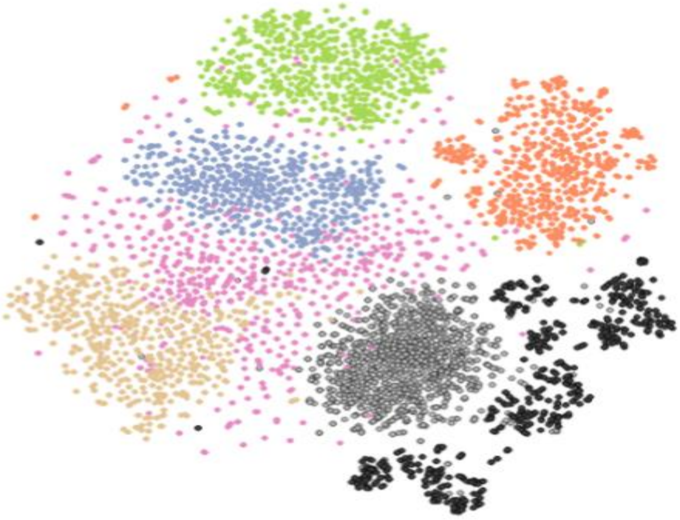


Fig 1: Visualization of 190D features for different people in different colors

One other illustration of utilizing hand created feature highlights for deepfake recognition is by Tackhyun Jung et al, The creators propose a structure to dissect an individual's flickering examples to identify deepfakes in video. They place that flickering examples are known to change dependent on states of being, intellectual exercises, organic factors, etc. Utilizing an assortment of calculations, the creators extricate the facial area and figure the Eye Aspect Ratio (EAR) for each casing in the video. The EAR an incentive for a shut eye is ordinarily more modest than the EAR an incentive for an open eye (for the specific system detailing, if it's not too much trouble, allude to the paper). By setting a fitting edge, one can identify squinting occasions dependent on the EAR esteem and dissect an individual's flickering examples in video[3].

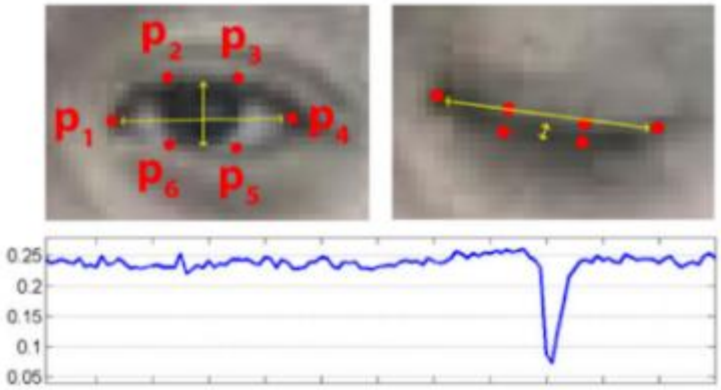


Fig 2: This equation shows the calculation of EAR in the frame Unit. Vertical-EAR and Time-Horizontal.

As a component of this system, four ascribes (sex, age, movement and time) were taken as contribution to portray the individual in the video. In view of these qualities, the creators question a pre-designed information base with normal flickering example information. At that point they contrast the deliberate squinting example information and the flickering example information questioned from the data set. In light of this they can learn whether a video is a deepfake.

B. LEARNING BASED FEATURES

In this technique we will focus on learning based deepfake detection methods. A large number of these strategies use convolutional neural organizations (CNNs) to become familiar with the highlights important for deepfake recognition.Andreas Rössler et al. in the FaceForensics++ paper evaluate several learning based methods and one hand-crafted feature extraction method on their ability to detect forgery at different video quality levels, They pre-extract the facial region from the input image (and extract a center-crop or resize for certain methods) and use the various methods to detect forgery, They noticed that all approaches achieved high accuracy on raw input data. However, performance dropped for compressed videos. Among the tested methods, the XceptionNet model achieved the highest performance across all video quality levels[4].

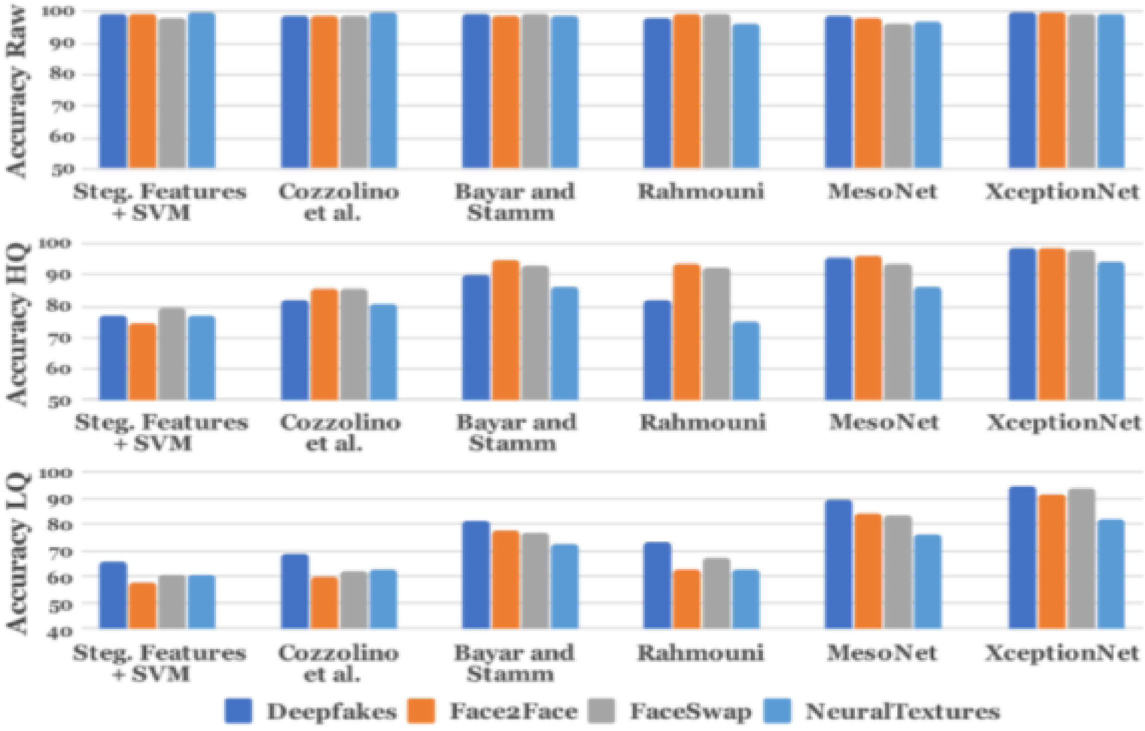


Fig 3: Accuracy of Different Algorithms

Some deepfake creation techniques can't produce recordings that are transiently reliable. Relics that emerge because of this irregularity may give great signals to video based deepfake detection. One approach to consolidate fleeting data alongside spatial data for preparing a model is by utilizing 3D CNNs, On that note, Yaohui Wang et al. examine the capacity of a couple of 3D CNNs (explicitly I3D, 3D ResNet and 3D ResNeXt) to identify controlled video[4].Irene Amerini et al. propose utilizing optical stream to recognize deepfakes. The overall thought is that deepfake recordings may have disparities moving across outlines (like uncommon developments of facial parts) which optical stream can catch, For a given casing f(t), they utilize the PWC-Net model to assess the forward optical stream OF(f(t), f(t+1)) which portrays the obvious movement of different components in the scene[5]. The extricated optical stream esteems are changed to a RGB picture design and is utilized by their Flow-CNN to distinguish deepfakes.

C. RELICS

As of now, deepfake creation strategies are not great. This implies there is regularly proof which we can examine to derive whether a piece of media was controlled and additionally is a deepfake. In this class, we will investigate some identification strategies that search for this proof, regularly called relics. Some falsification discovery strategies work by noticing the Photo Response Non Uniformity (PRNU) design on pictures,The creators of the paper Noiseprint notice that in the original paper by Lukas et al. they see that every individual gadget leaves a particular blemish on completely gained pictures (the PRNU design) because of defects in the gadget fabricating measure, The creators note that the absence of PRNU may show control. Obviously, PRNU-based strategies can be utilized to distinguish deepfakes[6]. Anyway these techniques do have a few downsides, one of which is you need an enormous number of pictures to make great assessments.

In another paper, Francesco Marra et al. show that each GAN leaves a particular "unique finger impression" in its created pictures, like certifiable cameras checking pictures with hints of their PRNU. The creators present a test showing proof of GAN fingerprints[7].For a picture Xi produced by a given GAN, they notice that the unique mark addresses an unsettling influence irrelevant with the picture semantics. To acquire the finger impression, they initially compute the commotion remaining Ri utilizing Ri = Xi – f(Xi), where the capacity f is a denoising channel. They accept that this remaining is an amount of a non-zero deterministic part (the unique finger impression, F) and an arbitrary clamor segment (Wi). Thus, Ri = F + Wi.

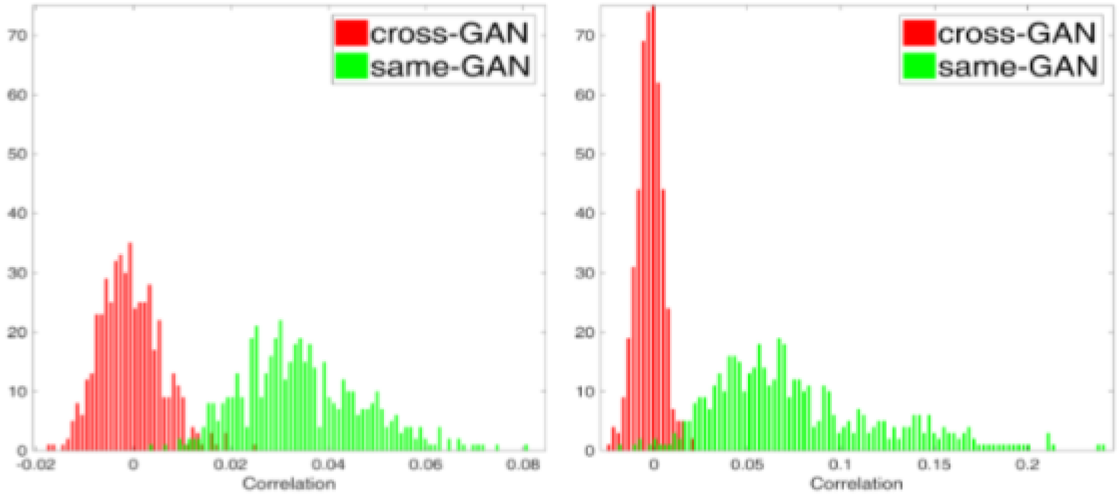


Fig 4: Correlation of Cycle-GAN(left) and Pro-GAN(right) residuals with same/cross-GAN fingerprints

The added substance commotion part for the most part gets counterbalanced on taking the normal of residuals processed from various pictures created by a similar GAN. Henceforth, by playing out the above activity, we get a gauge of the GAN unique mark.This outcome is fascinating on the grounds that we can possibly utilize it for distinguishing deepfakes, and in any event, recognizing the wellspring of a deepfake. Notwithstanding, the creators notice that more exploration is important to survey the qualities, ease of use and vigor of these fingerprints.

The work by Ning Yu et al. presents another intricate examination concerning learning and breaking down GAN fingerprints. One of their fascinating cases is that even minor contrasts in GAN preparing could bring about various fingerprints, which empowers fine-grained model confirmation, In general, their investigation gives a decent understanding into the reasonability and ability of GAN fingerprints.ate of the GAN finger impression[8].

III.CONCLUSION

Online influence campaigns will not decide to use deepfakes in a vacuum. Beyond the resources needed to produce a deepfake and the resulting quality, malicious actors will weigh their hoaxed media's risk of being identified as fake. To that end, progress in the field of ML on detecting deepfakes will influence how the technology is used to spread false narratives. Disinformation campaigns will avoid easily detectable deepfakes in favor of ones harder to identify. This spaper examines the current strengths and weaknesses in deepfake detection and how detection algorithms might be used in practice. These ML-based detection algorithms suffer from one significant drawback: they perform poorly when encountering novel means of creating faked media not incorporated into the original training set. Even beyond deepfake detection, ML models frequently perform well only on the data-set they were trained on, resulting in systems that fail when presented with new data. Deepfake detection is far from perfect. While detection systems succeed in identifying known deepfake creation methods, they lack in data for identifying new ones. In this paper we have gone through a few methods that have performed exceptionally well in detecting the deepfake. Still a lot more research needs to be done to detect the new types of deepfake.

REFERENCES

[1] Ruben Tolosana, Ruben Vera-Rodriguez, Julian Fierrez, Aythami Morales and Javier Ortega-GarciaBiometrics and Data Pattern Analytics - BiDA Lab, Universidad Autonoma de Madrid, Spain. “DeepFakes and Beyond: A Survey ofFace Manipulation and Fake Detection”, 2020

[2] Shruti Agarwal and Hany Farid “Protecting World Leaders Against Deep Fakes” University of California, Berkeley Berkeley CA, USA 2019.

[3] Tackhyun Jung; Sangwon Kim; Keecheon Kim "DeepVision: Deepfakes Detection Using Human Eye Blinking Pattern", IEEE Access (Volume: 8) 2020.

[4] Andreas Rossler,Davide Cozzolino, Luisa Verdoliva, Christian Riess,justus Thies,Matthias Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images" Technical University of Munich 2019.

[5] Yaohui Wang, Antitza Dantcheva, "A video is worth more than 1000 lies. Comparing 3DCNN", 020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020).

[6] Irene Amerini, Leonardo Galteri, Roberto Caldelli, Alberto Del BimboMedia Integration and Communication Center (MICC), University of Florence, Florence, Italy National Inter-University Consortium for Telecommunications (CNIT),”Deepfake Video Detection through Optical Flow based CNN” Parma, Italy.

[7] Davide Cozzolino, Luisa Verdoliva, "Noiseprint: a CNN-based camera model fingerprint", 2018.

[8] Jan Lukás, Jessica Fridrich, Miroslav Goljan, "Detecting digital image forgeries using sensor pattern noise - art. no. 60720Y" 2006.

[9] Francesco Marra, Diego Gragnaniello, Luisa Verdoliva, Giovanni Poggi DIETI – University Federico II of Naples Via Claudio 21, 80125 Napoli – ITALY, "Do GANs leave artificial fingerprints?" 2018.

[10] Ning Yu, Larry Davis, Mario Fritz, "Attributing Fake Images to GANs: Learning and Analyzing GAN Fingerprints", 2019.

[11] Joao C. Neves, Ruben Tolosana, Ruben Vera-Rodriguez, Vasco Lopes, Hugo Proenc¸a and Julian Fierre, "GANprintR: Improved Fakes and Evaluation of the State of the Art in Face Manipulation Detection", 2020.

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

☐ 9940 572 462 ☎ 6381 907 438 ✉ ijircce@gmail.com

www.ijircce.com