PUFshield: A Hardware-Assisted Approach for Deepfake Mitigation Through PUF-Based Facial Feature Attestation

Venkata K. V. V. Bathalapalli
University of North Texas
Denton, Texas, USA
vb0194@unt.edu

Venkata P. Yanambaka
Texas Woman's University
Denton, Texas, USA
vyanambaka@twu.edu

Saraju P. Mohanty
University of North Texas
Denton, Texas, USA
saraju.mohanty@unt.edu

Elias Kougianos
University of North Texas
Denton, Texas, USA
elias.kougianos@unt.edu

ABSTRACT

Deepfake has emerged as a threat to individual's privacy and identity. It uses advanced deep learning algorithms to synthesize visual, text, and audio from multimedia content in a realistic way. The advancement of Deepfake techniques is posing a question on the integrity of the digital content on social media. This work presents a novel hardware assisted Deepfake mitigation approach through the device and content integrity verification. In this work, the potential of hardware security primitive Physical Unclonable Functions (PUF) for mitigation of visual Deepfakes has been explored. The proposed framework presents a novel PUF-based image attestation technique that uses human facial features to create a unique pseudo-identity. The proposed architecture maps facial key point coordinates of each person in an image to PUF and creates a unique PUF generated key thereby having a unique pseudo identity for each image. Experimental evaluation uses Dlib facial detection model for facial attribute extraction and uses Arbiter PUF for image attestation.

CCS CONCEPTS

• **Security and privacy** → **Hardware-based security protocols**; **Hardware-based security protocols**; *Distributed systems security*; • **Social and professional topics** → Identity theft.

KEYWORDS

Security-by-Design (SbD), Hardware Assisted Security (HAS), Physical Unclonable Functions (PUF), Deepfake, Deepfake Mitigation

ACM Reference Format:
Venkata K. V. V. Bathalapalli, Venkata P. Yanambaka, Saraju P. Mohanty, and Elias Kougianos. 2024. PUFshield: A Hardware-Assisted Approach for Deepfake Mitigation Through PUF-Based Facial Feature Attestation. In *Great Lakes Symposium on VLSI 2024 (GLSVLSI '24), June 12–14, 2024, Clearwater, FL, USA.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3649476.3660394

GLSVLSI '24, June 12–14, 2024, Clearwater, FL, USA
© 2024 Copyright held by the owner/author(s). Publication rights licensed to ACM.
ACM ISBN 979-8-4007-0605-9/24/06…$15.00
https://doi.org/10.1145/3649476.3660394

1 INTRODUCTION

Deepfake is a term coined to define fake content creation and modification technique using deep learning algorithms. Deepfakes have evolved from synthetic media which uses computer generated artificial audio or video. Synthetic media has gained much prominence for its application in entertainment and media applications [13]. Deepfake leverages a Generative adversarial network (GAN) which enables the modification of human faces in a video or image. Deepfakes are evolving rapidly and their technological capability in modifying speech, perform face swapping thereby spreading misinformation and fake content is increasing [9]. The state-of-art Deepfake techniques are illustrated in Fig. 1.

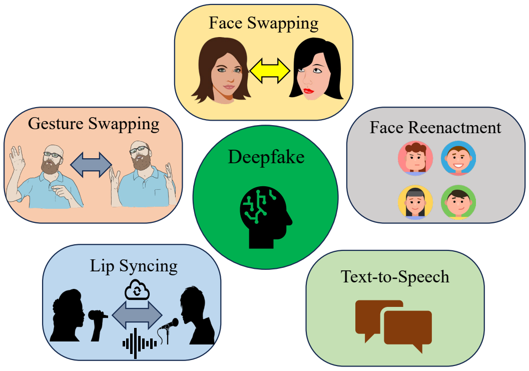


Figure 1: Deepfake Techniques

The disruptive technologies and applications have enabled faster and easier ways to compress and forge videos of individuals from social media. The mobile applications like 'FaceApp' have simplified the video forgery techniques which have become a serious threat to personalized digital content on social media [11]. The extent of implications this technology can have on identity protection, forgery, misinformation, and hate spreading has clearly shown the importance of Deepfake detection and mitigation techniques[21]. Emerging applications enabling easier ways to compress, edit, and crop images has eased the process of fake content creation with the highest degree of realism. Also, the human perceptibility of identifying Deepfakes is becoming a challenge due to the evolution of advanced techniques involving GAN, Deep learning, and machine learning techniques.

Popular video forgery techniques include auto encoders which consist of an encoder and decoder. Encoders map the facial features like eyes, nose, skin texture and face color of both the source and target faces into a set of latent vectors which provide an abstract of underlying features for the deep learning models to work. Decoders then decompress the compressed image to reconstruct the source face [4, 14, 18].

Deepfake detection and mitigation involves various techniques for identifying and countering fake digital content. Deepfake detection is the process of identifying the captured digital content as real or fake. This includes data classification techniques using machine learning to perform identification. Deepfake detection requires large amounts of datasets for training and validation to train the model for identifying real and fake content. Deepfake mitigation is an emerging technological advancement in countering Deepfakes. This includes awareness, legislation, and content integrity verification. These techniques help in countering or editing the video/image with proper mitigation techniques in place. However, the requirement for an energy efficient and scalable approach to protect multimedia content from any unauthorized modification is essential to counter any threat to individual privacy and content on digital media. As the influence of social media is increasing, any unauthorized modification to the original content uploaded and shared could have implications on society. This signifies the requirement for a reliable Deepfake mitigation solution.

This work ensures the integrity of captured digital content of an individual shared on social media to be secure and resistant to Deepfake. The proposed approach can effectively counter Deepfake and presents a sustainable solution to clearly identify the fake video/image of a person from the source using PUF. PUF is a hardware security primitive that taps onto the potential of device level uniqueness facilitated by micro variations during chip manufacturing process. A PUF can be simply referred to as a digital fingerprint that can generate random bit stream as unique response. The PUF response is obtained due to manufacturing variations propelling frequency and logical routing delay variations in an IC.

The proposed research focuses on enabling a video captured from a device to be shared with underlying PUF security. Also, this work focuses mainly on visual Deepfakes to protect individual user's privacy by performing PUF-based facial feature attestation in an image.

The rest of this paper is organized as follows: Section 2 presents an overview of related research in the areas of deepfake detection, mitigation and state-of-art PUF-based security applications. Section 3 presents the novel contributions of this research work. Section 4 presents the proposed novel PUF-based Deepfake mitigation technique, and its experimental validation and results are presented in Section 5. Finally, the conclusion and future research directions are presented in Section 6.

2 RELATED WORKS

This section discusses related research works on Deepfake detection and mitigation, and hardware security primitive PUF-based security solutions from the state-of-the art research.

Many Deepfake detection techniques were proposed which have been able to identify fake content. These techniques include identifying facial landmarks, and eye blinking. Another approach is to observe the misalignment of facial emotions on Deepfake [12]. Also, hardware assisted watermarking approaches that embeds a signature inside the multimedia object have been efficient for ensuring authenticity of digital content [7]. Exploring source-based content integrity verification schemes have been perceived as a sustainable approach for mitigating Deepfake and facilitating integrity to content [11, 19]. Deepfake detection for videos can also be done using eye movements, raw EEG signal, and blinking using frameworks like FakeEt [3, 6]. A secure digital identity framework for smart city applications to counter Deepfake and user identity is proposed in [10] which focuses on CNN model for feature extraction and a bio key generation using facial attributes and coordinates [10]. Table.1 presents a comprehensive analysis of state-of-art research on Deepfake detection and mitigation.

3 NOVEL CONTRIBUTIONS

This section will discuss the novelty of the proposed PUF-based approach for combating Deepfakes. Section. 3.1 and 3.2 provide a detailed overview of the threat model and novel contributions of the proposed work.

3.1 Problem Statement

Deepfake poses a serious threat to individual privacy and can also impact organizations. An example of this has been the usage of lip syncing to modify the original audio of a company's CEO addressing a conference. This type of instance could lead to misinformation [20]. Modifying videos through advanced deep learning techniques and performing face swapping is also another issue that requires serious attention. The threat model addressed in this paper is illustrated in Fig. 2.

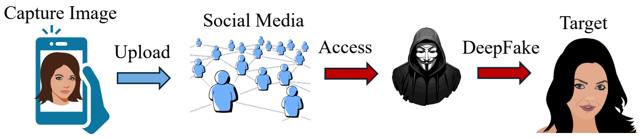


Figure 2: Overview of the Threat Model

Addressing visual Deepfake of individual content captured as a video/image is important and necessary to counter facial attribute manipulation which includes modifying facial attributes like eyes, nose, lips and replacing them with target's attributes.

3.2 Novelty of the Proposed Solution

This work focuses on addressing facial attribute manipulation of individuals in an image which is a very serious problem for individual identity. The novel contributions of the proposed work are:

- A secure digital content integrity verification scheme through hardware enabled attestation.
- Presenting a state-of-art PUF-based approach for digital content attestation.

Table 1: Related Research

Work	Approach	Technique	Methodology	Tools	Features
Kato et.al [5]	Mitigation	Visual	Scapegoat Image Generation	StyleGAN2	Privacy and Anonymity
Zheng et.al[23]	Mitigation	Visual	PUF-based device and data hash	CMOS Image sensor	Image content authenticity
Krause et. al[8]	Detection	Audio	Language and phoneme focused	Logistic regression	Detection using mouth movements
Pishori et.al[15]	Detection	Visual	Eye Blink rate	CNN+RNN, OpenCV	Efficient through eye blink rate detection
Wang et. al[17]	Mitigation	Visual	GAN based secret message embedding in an image	GAN	Personal photo protection
Zhao et.al[22]	Detection	Visual	Image watermarking	Neural network with encoder and decoder	Effective image quality preservation
Ashok et.al[16]	Detection	Visual	Training XceptionNet using faceforenscis++ dataset	XceptionNet Model	Identifying Deepfake from Original content
Doan et.al[2]	Detection	Audio	Identifying silence, breathing,talking in an Audio	RawNet2	Biological sounds based detection
PUFshield (This Work)	Mitigation	Visual	PUF-based Facial Feature Attestation	PUF, Dlib Facial detection and landmark prediction	Image and device integrity

- A state-of-art solution for countering facial attribute manipulation to prevent visual Deepfakes.
- A device security framework providing PUF-based pseudo identity for the camera capturing image/video.
- An approach to counter Deepfakes countering facial attribute manipulation.

3.3 Why PUF for Deepfake Mitigation?

The motivation for this research work is to explore the scope PUF to mitigate and counter Deepfakes. The proposed framework securely performs facial landmark coordinate attestation using PUF by mapping facial landmark pixel coordinates as inputs to PUF module at the device/ camera capturing video/image. Also, the device's pseudo PUF generated identity when bound with the image PUF key can provide authenticity to image since PUF ensures hardware generated randomness unique to each image. The same device with the PUF when capturing another image produces a completely different output since each image is captured with different angle and its coordinates are aligned differently thereby providing security for each image. The facial attribute attestation using PUF is a unique and novel framework to counter facial attribute manipulation which utilizes PUF to map these facial attributes which thereby provides hardware root of trust for captured personalized digital content.

4 PUFSHIELD: PROPOSED PUF-BASED DEEPFAKE MITIGATION TECHNIQUE

The architectural overview of proposed work is illustrated in Fig. 3. in the proposed work, it is assumed that any device capturing the image of a person will have a PUF embedded and has a unique PUF generated identity. Since all the modern digital cameras and smartphones are based on IC technology, the PUF embedded security can provide integrity to the device capturing image by generating a unique digital fingerprint. Once the image is captured, the image is preprocessed before performing facial detection. The preprocessing include resizing the image and converting it to gray scale.

Table 2: Facial Landmark Coordinates from Dlib

Facial Landmarks	Pixel Coordinates
Left Eye	36-41
Right Eye	42-47
Left Eyebrow	17-21
Right Eyebrow	22-26
Jaw	0-16
Nose Bridge	27-30
Lower Nose	31-35
Outer Lip	48-59
Inner Lip	60-67

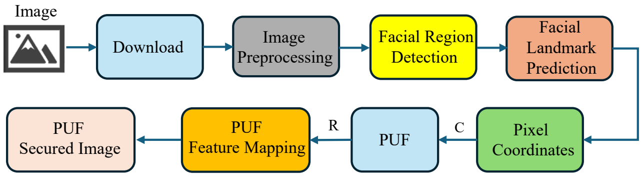


Figure 3: Working flow of Proposed Solution

For facial feature attestation, initially facial region has to be detected followed by identification of human facial features. Facial detection and landmark prediction models from Dlib were used for performing facial detection and landmark prediction which identifies 68 facial landmarks from the face in the captured image. The models were pretrained on i-Bug 300-W data set and the facial detection model is based on Histogram of Oriented Gradients (HOG) which is a static object prediction model. For image processing in the proposed facial detection model, HOG was used to identify the facial region. HOG is one of the most widely used feature descriptor for computer vision and image processing applications and is widely used for object detection in images. It analyzes the entire image by dividing it into small cells and tracks the varying intensity of an image pixel to perform object detection. The Haar cascade filter similarly is another feature descriptor that effectively performs object detection dynamically and is one of the most widely used feature descriptor for video processing [1] in computer vision. Haar cascade filter is preferred over HOG for video processing since it is difficult to track and analyze each frame in a video using HOG. HOG is a preferred choice for static object detection in image processing but might incur more computational resources and power for object detection in videos.

Once the HOG detects facial region of interest in an image, the predictor performs facial landmark prediction. The facial landmark coordinates from the landmark prediction model are given in Table. 2. 68 facial landmarks are identified by the model. 68 facial landmark pixel coordinates are extracted and a numpy array is created which consists of 136 elements. The numpy array consists of pixel coordinate values corresponding to 68 facial landmarks as each landmark has pixel coordinates values corresponding to x and y. The x coordinate represents the length of the pixel from the left edge of the image and y coordinate represents the length of the pixel from the top edge of the image. The facial coordinate array is then given as challenge to PUF module at the device. A chunk of n elements are given as challenge at a time. Totally j number of chunks are given as challenges to PUF and the responses are obtained for each chunk respectively. Finally, the responses for all the j chunks are XOR ed. The finally obtained XOR ed output will be the PUF attested digital fingerprint for the image. Any video uploaded onto social media can be easily identified as real or fake by performing the PUF attestation and comparing the finally obtained keys. If the keys are matching, the image can be considered as protected. The proposed approach can effectively counter facial attribute manipulation technique which is classified as visual Deepfake.

The methodology of the proposed work is illustrated in Algorithm 1.

5 EXPERIMENTAL RESULTS

For experimental validation, open source and attribution free images were downloaded and read using OpenCv library. Experimental prototype has an AI edge hardware from NVIDIA, a single board computer and FPGAs for PUF as outlined in Fig. 4 and Table. 3a. The 64-bit Arbiter PUF logic was programmed on Xilinx Artix-7 and Spartan-7 FPGAs. The image processing, facial region detection and landmark prediction were performed on 3 images with different facial attributes. Each image is read and resized to 600x500 for compatibility. Facial landmark pixel coordinate array is obtained for all the images. Landmark coordinate array is extracted for each image which consists of 136 elements corresponding to the pixel coordinates.

Universal asynchronous receiver and transmitter(UART) serial communication protocol was used for serially writing challenges to PUF module on FPGA. The responses from FPGA are serially read at a defined baud rate. Baud rate of 9600 was used in the proposed work for both Artix-7 and Sparton-7 FPGAs with 100 MHz clock frequency. The PUF module supports 64 bit Challenge Response pairs (CRPs). Since the landmark coordinate array has 136 numbers, a group of 8 numbers is given as challenge to the PUF module at a time. Totally, 17 groups of challenge inputs are processed by PUF module generating 17 64-bit responses. The XOR logic operation is performed on all the obtained 17 responses and finally, a unique bitstream is obtained. The computational time analysis was presented in Table. 3b for all the images which include

Algorithm 1: PUF-based Facial Feature Attestation

```
Capture Image $I_n$
Read Image
    • Resize $I_n \rightarrow$600x500
    • Image $I_n \rightarrow$ Grey Scale
Perform face region detection
    • HOG $\rightarrow$ RoI
Access the PUF at the camera/device $d_i$ capturing the video.
    • Obtain $D_I$ ID
Access PUF $PUF_{ID}$ at the device
    • Generate Pseudo Identity: Device ID $\rightarrow PUF_{ID} \rightarrow R_{ID}$
Obtain Facial landmark pixel coordinates
    • 68 Facial Landmarks $\rightarrow$ Pixel Coordinates
64 bit PUF Module
Generate PUF generated pseudo identity for facial landmarks 8 at a time
$F_1$,$\rightarrow$ (x,y) coordinates of 8 facial coordinates
$F_2$,$\rightarrow$(x,y) coordinates of first 8-16
$F_3$,$\rightarrow$(x,y) coordinates of first 16-24
.....
$F_{17}$,$\rightarrow$(x,y) coordinates of landmarks 128-136
First 8 facial landmarks $F_1$: Facial Key Points
$F_1 \rightarrow PUF_{ID} \rightarrow R_1$
Perform XOR operation of $F_1$ pseudo identity with $F_2$
$F_2 \rightarrow PUF_{ID} \rightarrow R_2$
Perform XOR of all 17 PUF generated Keys $F_1 \oplus F_2 \oplus ... F_{17}$
Generated Final image pseudo identity $F_n$ is the pseudo identity of image.
Generate pseudo identity for the device Edge node$\rightarrow$PUF$\rightarrow R_E$
Securely upload PUF protected Image using $R_{ID}$ and $F_n$
```

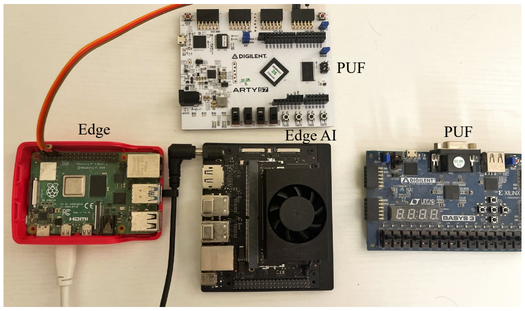
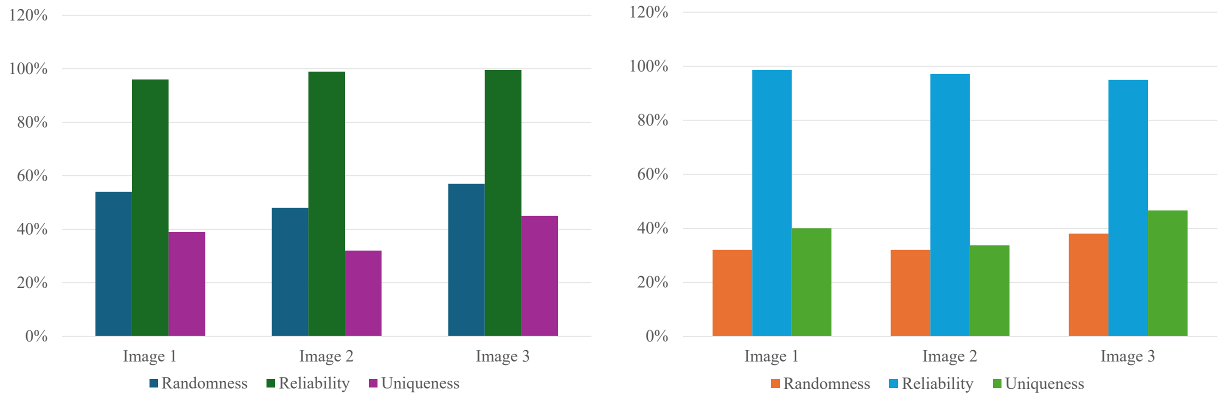


Figure 4: Edge Node with PUFs



(a) Artix-7 FPGA

(b) Spartan-7 FPGA

Figure 5: PUF Evaluation Results

Table 3: Performance Analysis of PUFshield

(a) Experimental Evaluation

Parameters	Details
Application	Deepfake Mitigation
Face Detection Model	Dlib
Security Module	PUF
PUF	Arbiter PUF
Edge	Single Board Computer & Jetson Orin Nano board
Tools	Jetpack SDK 6.0, Vivado 2023.2
PUF Hardware	Basys 3, Arty S7

(b) Computational Time Analysis

Content	Parameter	Results
Image 1	Facial Detection Facial Landmark Prediction	60 ms 3ms
Image 2	Facial Detection Facial Landmark Prediction	57 ms 2 ms
Image 3	Facial Detection Facial Landmark Prediction	56 ms 3 ms
All Images	Overall Attestation Time	300 ms

time taken for facial detection and landmark prediction for all the images.

The PUF keys were extracted using a single board computer corresponding to the facial landmark pixel coordinates.

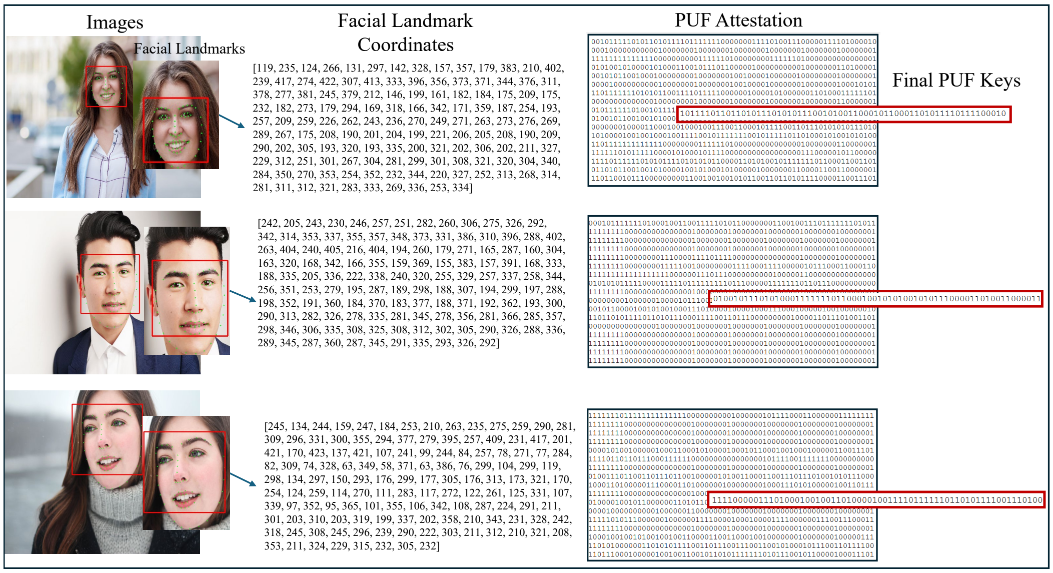
To evaluate the robustness of obtained results, the Figures-of-merit of PUF responses for all the images were evaluated. Initially, the randomness of obtained responses was evaluated by calculating the distribution of 1 and 0 in a bitstream. The ideal randomness is 50%. Similarly, the PUF response uniqueness is obtained by calculating the extent of variation of responses to different challenges. The average intra hamming distance for all the 17 responses was calculated to obtain uniqueness. Finally, reliability of a PUF module is its ability to regenerate the same response at varying environmental and operating conditions by comparing the obtained responses for same challenge input. In the work, the PUF metric evaluation was performed at ambient temperature and totally, PUF responses were evaluated for five instances. All the images PUF responses were regenerated with 100% reliability approximately. Fig. 5, and Fig. 6 presents PUF attestation and evaluation results for all the three images.

The power consumption analysis was performed for Jetson Nano and Raspberry pi board to evaluate the robustness of various edge driven AI hardware. The Jetson Orin Nano board has an idle power consumption of 5.9-6.4 watts and raspberry pi has an idle power consumption range of 3.1-3.5 watts. During the facial detection and landmark prediction, the power consumption of Jetson Orin Nano board was 7.3 watts and on Raspberry pi, the power consumption was 5.7 watts.

6 CONCLUSION AND FUTURE RESEARCH

This research work presented and validated a state-of-art Deepfake mitigation technique that utilizes the potential of PUF for secure facial feature mapping and attestation thereby securing multimedia content. The proposed work experimentally validated the PUF-based facial feature attestation process for an image. This work can effectively counter Deepfake particularly facial attribute manipulation technique. The proposed work clearly presents an approach to attest the device through PUF and facial feature attestation thereby ensuring PUF assisted digital content security for social media. The metrics evaluation results and computational time and power analysis on various hardware clearly demonstrates the potential of the proposed PUFshield.

As a direction for future research, countering other techniques of visual Deepfakes such as face swapping, lip syncing in video and audio Deepfakes using PUF can be potential areas for PUF-based Deepfake mitigation.



```
[119, 235, 124, 266, 131, 297, 142, 328, 157, 357, 179, 383, 210, 402,
239, 417, 274, 422, 307, 413, 333, 396, 356, 373, 371, 344, 376, 311,
378, 277, 381, 245, 379, 212, 146, 199, 161, 182, 184, 175, 209, 175,
232, 182, 273, 179, 294, 169, 318, 166, 342, 171, 359, 187, 254, 193,
257, 209, 259, 226, 262, 243, 236, 270, 249, 271, 263, 273, 276, 269,
289, 267, 175, 208, 190, 201, 204, 199, 221, 206, 205, 208, 190, 209,
290, 202, 305, 193, 320, 193, 335, 200, 321, 202, 306, 202, 211, 327,
229, 312, 251, 301, 267, 304, 281, 299, 301, 308, 321, 320, 304, 340,
284, 350, 270, 353, 254, 352, 232, 344, 220, 327, 252, 313, 268, 314,
281, 311, 312, 321, 283, 333, 269, 336, 253, 334]
```

```
1011110110110101110101011100101001100010110001101011110111100010
```

```
[242, 205, 243, 230, 246, 257, 251, 282, 260, 306, 275, 326, 292,
342, 314, 353, 337, 355, 357, 348, 373, 331, 386, 310, 396, 288, 402,
263, 404, 240, 405, 216, 404, 194, 260, 179, 271, 165, 287, 160, 304,
163, 320, 168, 342, 166, 355, 159, 369, 155, 383, 157, 391, 168, 333,
188, 335, 205, 336, 222, 338, 240, 320, 255, 329, 257, 337, 258, 344,
256, 351, 253, 279, 195, 287, 189, 298, 188, 307, 194, 299, 197, 288,
198, 352, 191, 360, 184, 370, 183, 377, 188, 371, 192, 362, 193, 300,
290, 313, 282, 326, 278, 335, 281, 345, 278, 356, 281, 366, 285, 357,
298, 346, 306, 335, 308, 325, 308, 312, 302, 305, 290, 326, 288, 336,
289, 345, 287, 360, 287, 345, 291, 335, 293, 326, 292]
```

```
010010111010100011111101100010010101001010111000011010011000011
```

```
[245, 134, 244, 159, 247, 184, 253, 210, 263, 235, 275, 259, 290, 281,
309, 296, 331, 300, 355, 294, 377, 279, 395, 257, 409, 231, 417, 201,
421, 170, 423, 137, 421, 107, 241, 99, 244, 84, 257, 78, 271, 77, 284,
82, 309, 74, 328, 63, 349, 58, 371, 63, 386, 76, 299, 104, 299, 119,
298, 134, 297, 150, 293, 176, 299, 177, 305, 176, 313, 173, 321, 170,
254, 124, 259, 114, 270, 111, 283, 117, 272, 122, 261, 125, 331, 107,
339, 97, 352, 95, 365, 101, 355, 106, 342, 108, 287, 224, 291, 211,
301, 203, 310, 203, 319, 199, 337, 202, 358, 210, 343, 231, 328, 242,
318, 245, 308, 245, 296, 239, 290, 222, 303, 211, 312, 210, 321, 208,
353, 211, 324, 229, 315, 232, 305, 232]
```

```
111100000111010001001001101000010011110111111011010111001110100
```

Figure 6: Facial Landmark Detection Model

REFERENCES

[1] Hoda El Boussaki, Rachid Latif, and Amine Saddik. 2023. Drowsiness detection using Dlib: an overview. In *Proc.7th IEEE Congress on Information Science and Technology (CiSt)*. IEEE, Morocco, 150–154. https://doi.org/10.1109/CiSt56084.2023.10409980

[2] Thien-Phuc Doan, Long Nguyen-Vu, Souhwan Jung, and Kihun Hong. 2023. BTS-E: Audio Deepfake Detection Using Breathing-Talking-Silence Encoder. In *ICASSP 2023 - 2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, Greece, 1–5. https://doi.org/10.1109/ICASSP49357.2023.10095927

[3] Parul Gupta, Komal Chugh, Abhinav Dhall, and Ramanathan Subramanian. 2020. The eyes know it: FakeET – An Eye-tracking Database to Understand Deepfake Perception. https://doi.org/10.48550/ARXIV.2006.06961

[4] Rahul Katarya and Anushka Lal. 2020. A Study on Combating Emerging Threat of Deepfake Weaponization. In *2020 Fourth International Conference on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*. IEEE, India, 485–490. https://doi.org/10.1109/I-SMAC49090.2020.9243588

[5] Gido Kato, Yoshihiro Fukuhara, Mariko Isogawa, Hideki Tsunashima, Hirokatsu Kataoka, and Shigeo Morishima. 2023. Scapegoat Generation for Privacy Protection from Deepfake. In *Proc. IEEE International Conference on Image Processing (ICIP)*. IEEE, Kuala Lumpur, 3364–3368. https://doi.org/10.1109/icip49359.2023.10221904

[6] Aynur Kocak and Mustafa Alkan. 2022. Deepfake Generation, Detection and Datasets: a Rapid-review. In *Proc. 15th International Conference on Information Security and Cryptography (ISCTURKEY)*. IEEE, Turkey, 86–91. https://doi.org/10.1109/iscturkey56345.2022.9931802

[7] Elias Kougianos, Saraju P. Mohanty, and Rabi N. Mahapatra. 2009. Hardware assisted watermarking for multimedia. *Computers and Electrical Engineering* 35, 2 (March 2009), 339–358. https://doi.org/10.1016/j.compeleceng.2008.06.002

[8] Jonas Krause, Andrei De Souza Inacio, and Heitor Silvério Lopes. 2023. Language-focused Deepfake Detection Using Phonemes, Mouth Movements, and Video Features. In *Proc. IEEE Latin American Conference on Computational Intelligence (LA-CCI)*, Vol. nill. IEEE, Colombia, 1–6. https://doi.org/10.1109/LA-CCI58595.2023.10409327

[9] Asad Malik, Minoru Kuribayashi, Sani M. Abdullahi, and Ahmad Neyaz Khan. 2022. DeepFake Detection for Human Face Images and Videos: A Survey. *IEEE Access* 10 (2022), 18757–18775. https://doi.org/10.1109/access.2022.3151186

[10] Alakananda Mitra, Saraju P. Mohanty, Peter Corcoran, and Elias Kougianos. 2021. iFace: A Deepfake Resilient Digital Identification Framework for Smart Cities. In *Proc.IEEE International Symposium on Smart Electronic Systems (iSES)*. IEEE, India, 361–366. https://doi.org/10.1109/ises52644.2021.00090

[11] Mekhail Mustak, Joni Salminen, Matti Mäntymäki, Arafat Rahman, and Yogesh K. Dwivedi. 2023. Deepfakes: Deceptions, mitigations, and opportunities. *Journal of Business Research* 154 (January 2023), 113368. https://doi.org/10.1016/j.jbusres.2022.113368

[12] Amal Naitali, Mohammed Ridouani, Fatima Salahdine, and Naima Kaabouch. 2023. Deepfake Attacks: Generation, Detection, Datasets, Challenges, and Research Directions. *Computers* 12, 10 (October 2023), 216. https://doi.org/10.3390/computers12100216

[13] Department of Homeland Security. 2021. Increasing Threat of DeepFake Identities. https://www.dhs.gov/sites/default/files/publications/increasing_threats_of_deepfake_identities_0.pdf

[14] Yogesh Patel, Sudeep Tanwar, Rajesh Gupta, Pronaya Bhattacharya, Innocent Ewean Davidson, Royi Nyameko, Srinivas Aluvala, and Vrince Vimal. 2023. Deepfake Generation and Detection: Case Study and Challenges. *IEEE Access* 11 (2023), 143296–143323. https://doi.org/10.1109/ACCESS.2023.3342107

[15] Armaan Pishori, Brittany Rollins, Nicolas van Houten, Nisha Chatwani, and Omar Uraimov. 2020. Detecting Deepfake Videos: An Analysis of Three Techniques. https://doi.org/10.48550/arXiv.2007.08517 arXiv:2007.08517 [cs.CV]

[16] Ashok V and Preetha Theresa Joy. 2023. Deepfake Detection Using XceptionNet. In *IEEE International Conference on Recent Advances in Systems Science and Engineering (RASSE)*. IEEE, India, 1–5. https://doi.org/10.1109/RASSE60029.2023.10363477

[17] Run Wang, Felix Juefei-Xu, Meng Luo, Yang Liu, and Lina Wang. 2021. FakeTagger: Robust Safeguards against DeepFake Dissemination via Provenance Tracking. In *Proceedings of the 29th ACM International Conference on Multimedia* (Virtual Event, China) *(MM ’21)*. Association for Computing Machinery, New York, NY, USA, 3546–3555. https://doi.org/10.1145/3474085.3475518

[18] Saima Waseem, Syed Abdul Rahman Syed Abu Bakar, Bilal Ashfaq Ahmed, Zaid Omar, Taiseer Abdalla Elfadil Eisa, and Mhassen Elnour Elneel Dalam. 2023. DeepFake on Face and Expression Swap: A Review. *IEEE Access* 11 (2023), 117865–117906. https://doi.org/10.1109/access.2023.3324403

[19] Mohammad Wazid, Amit Kumar Mishra, Noor Mohd, and Ashok Kumar Das. 2024. A Secure Deepfake Mitigation Framework: Architecture, Issues, Challenges, and Societal Impact. *Cyber Security and Applications* 2 (2024), 100040. https://doi.org/10.1016/j.csa.2024.100040

[20] Mika Westerlund. 2019. The Emergence of Deepfake Technology: A Review. *Technology Innovation Management Review* 9, 11 (January 2019), 39–52. https://doi.org/10.22215/timreview/1282

[21] Chaofei Yang, Leah Ding, Yiran Chen, and Hai Li. 2021. Defending against GAN-based DeepFake Attacks via Transformation-aware Adversarial Faces. In *Proc. International Joint Conference on Neural Networks (IJCNN)*. IEEE, Australia, 1–8. https://doi.org/10.1109/IJCNN52387.2021.9533868

[22] Yuan Zhao, Bo Liu, Ming Ding, Baoping Liu, Tianqing Zhu, and Xin Yu. 2023. Proactive Deepfake Defence via Identity Watermarking. In *Proc. IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*. IEEE, Hawaii, 4591–4600. https://doi.org/10.1109/WACV56688.2023.00458

[23] Yue Zheng, Yuan Cao, and Chip-Hong Chang. 2020. A PUF-Based Data-Device Hash for Tampered Image Detection and Source Camera Identification. *IEEE Transactions on Information Forensics and Security* 15 (2020), 620–634. https://doi.org/10.1109/tifs.2019.2926777