Middle East Technical University

Informatics Institute

VULNERABILITIES OF FACIAL RECOGNITION AND COUNTERMEASURES

Advisor Name: Assist. Prof. Dr. Cihangir TEZCAN
(METU)

Student Name: Utku AKTAS
(CSEC)

February 2022

TECHNICAL REPORT
METU/II-TR-2021-

Orta Doğu Teknik Üniversitesi

Enformatik Enstitüsü

YÜZ TANIMA SİSTEMLERİNİN ZAFİYETLERİ VE ÖNLEMLERİ

Danışman Adı: Assist. Prof. Dr. Cihangir TEZCAN
(ODTÜ)

Öğrenci Adı: Utku AKTAS
(CSEC)

Şubat 2022

TEKNİK RAPOR
ODTÜ/II-TR-2021-

REPORT DOCUMENTATION PAGE

1. AGENCY USE ONLY (Internal Use)	**2. REPORT DATE** 08.02.2022
3. TITLE AND SUBTITLE **Vulnerabilities of Facial Recognition and Countermeasures**	
4. AUTHOR (S) Utku AKTAŞ	**5. REPORT NUMBER (Internal Use)** METU/II-TR-2021-
6. SPONSORING/ MONITORING AGENCY NAME(S) AND SIGNATURE(S) Non-Thesis Master's Programme, Department of Cyber Security, Informatics Institute, METU	
Advisor: Asst. Prof. Dr. Cihangir Tezcan	Signature:
7. SUPPLEMENTARY NOTES	

8. ABSTRACT (MAXIMUM 200 WORDS)

Due to the developments in deep learning, the use of face recognition systems started to spread rapidly. Face recognition systems, which are used for purposes such as unlocking phones, entering our offices, or tracking citizens of states, also bring several problems. Attackers can find a weakness of the face recognition systems and avoid detection by facial recognition systems in various ways. Additionally attackers may carry out an impersonation attack which is the act of tricking the system by looking like an authorized person. In this research, basic building blocks of face recognition algorithms, face recognition vulnerabilities, how the attacks occur and what precautions can be taken are examined. It has been understood that it is difficult to reach a generalizable result because of a variety of facial recognition systems. In addition, attacks and countermeasures may differ according to the target system.

9. SUBJECT TERMS **Face recognition, deep learning, vulnerabilities, anti-spoofing**	**10. NUMBER OF PAGES** **27**

Table of Contents

List of Abbreviations

BIM: Basic Iterative Method
BMIM: Binary Mask Iterative Method
C & W: Carlini & Wagner's Method
CNN: Convolutional Neural Network
CSD-S: Common Specific Decomposition for Specific
DL: Deep Learning
FAS: Face Anti-spoofing
FGSM: Fast Gradient Sign Method
LFW: Labeled Faces in the Wild
MIM: Momentum Iterative Method
RCNN: Region-based Convolutional Neural Network
ReLU: Rectified Linear Unit
SIFT: Scale Invariant Feature Transform
SURF: Speeded Up Robust Features
Tanh: Tangent Hyperbolic Function

ABSTRACT

Due to the developments in deep learning, the use of face recognition systems started to spread rapidly. Face recognition systems, which are used for purposes such as unlocking phones, entering our offices, or tracking citizens of states, also bring several problems. Attackers can find a weakness of the face recognition systems and avoid detection by facial recognition systems in various ways. Additionally attackers may carry out an impersonation attack which is the act of tricking the system by looking like an authorized person. In this research, basic building blocks of face recognition algorithms, face recognition vulnerabilities, how the attacks occur and what precautions can be taken are examined. It has been understood that it is difficult to reach a generalizable result because of a variety of facial recognition systems. In addition, attacks and countermeasures may differ according to the target system.

CHAPTER 1 - INTRODUCTION

Face recognition algorithms have been changed since the latest improvements in deep learning. Most of these algorithms use deep learning (DL) especially Convolutional Neural Networks (CNN). Face recognition system used in various areas such as:

- Unlocking phone
- Banking systems
- Airport security
- Surveillance
- Market payments

Unfortunately, deep learning-based models involve some vulnerabilities. Many studies show that face recognition models can be tricked by crafted inputs. Attackers often develop re-impersonation attacks, as well as evasion methods. Physical attacks such as using 2d or 3d masks and electronic displays are the most common attack types. These attacks are called presentation attacks, they also comply with the definition of a replay attack. Another attack method is to create a similar input by analyzing the input image or features extracted from the image [1]. In evasion attacks, manipulating input images can be a useful method [2]. In one study, an attacker was able to successfully apply an evasion attack on two popular deep learning models with a sticker on his hat [3]. Apart from this, crafting input image methods also gave successful results to deceive the

system [4]. As attacks emerged, these attacks and the measures that could be taken began to be studied in the academic field. A large part of the work done is on preventing a single type of attack. In one study, a successful anti-spoofing result was obtained by adding a 3-dimensional feature analysis module alongside the deep learning model [5]. In another study, an additional class was added to the deep learning model and it was tried to differentiate fake and genuine with the model [6]. Apart from providing security with the model, there is also a study that checks the liveness detection with additional hardware [7]. However, taking security measures with liveness detection can negatively affect usability [8]. In a recent study [9], it has been focused on not only one attack type but multiple at once. Authors proposed a deep learning based solution, MixNet, for presentation attacks including 13 different types of presentation attack [9]. However, that paper does not include input feature manipulation attacks.

In this study, different attack and defense methods that have been carried out so far will be classified and examined in detail. To be able to understand the vulnerabilities, key components of the face recognition algorithms will be explained in Chapter 2. Different types of attacks will be mentioned and why these attacks are working will be clarified in Chapter 3. Although weaknesses exist, there are countermeasures for some of the vulnerabilities. Chapter 4 explains these countermeasures in detail. Finally, we conclude with unresolved security weaknesses at Chapter 5.

CHAPTER 2 - FACE RECOGNITION METHODS

Face recognition systems can be developed with many algorithms. The most widely used and up-to-date method today is deep learning models. Deep learning methods extract information from the input, apply a variety of processes and try to match the result with the targeted result which is called ground truth. Deep learning methods also give higher accuracies compared to classic image processing algorithms.

2.1. Definition of Face Recognition

Facial recognition systems are used to determine the identity of a person using image or video data. The data of the people are kept in a predetermined database. The face recognition algorithm processes the input data and compares it with the data in the database. In this way, the match is made and the identity of the person is determined.

2.2. Face Recognition Steps

To be able to match given input to the database, we need to detect faces from the input. After finding the faces, different algorithms can be used to extract information. Human face has specific characteristics and algorithms

focus on these characteristics and try to extract features. These features may include the shape of the face, distance between eyes and mouth and/or nose as well as complex patterns in case of deep learning based feature extraction. Explainable artificial intelligence focuses on what features a deep learning model learns.

2.2.1. Face Detection

This is the first step of face recognition. Face detection aims to find the location of the face of a person. Location returns as a bounding box which is a rectangular shape including pixel coordinates of the rectangles corners.

Face detection can be done by template matching algorithms, feature extraction methods or rule based methods. Among different types of face detection methods, we are going to focus on deep learning based algorithms.

2.2.2. Face Identification

Face identification allows us to match detected faces or their extracted features with the images or features in the database. The match with the highest probability is returned if it is above a certain threshold and identification is provided.

2.3. Face Recognition with Deep Learning Models

Convolutional Neural Networks (CNNs) are widely used for object detection, localization and classification. Input of CNNs can be digital images. They can output features extracted from those images or coordinates of found faces.

Supervised deep learning models should be trained with a dataset including input images and target results. Models have hyperparameters that will be learned from the step called training. To be able to calculate the accuracy, a small portion of the dataset is separated and used for evaluation after training.

We are going to explain basic structures of CNN’s and some of the most common feature extraction methods. Vulnerabilities often occur at loss function of CNN or feature extraction part.

2.3.1. Feature Extraction Algorithms

Main objective of feature extraction is reducing the number of dimensions and obtaining main characteristics of the data.

Scale Invariant Feature Transform (SIFT) is an algorithm that aims to find the keypoint of the given image. It is prone to image scale and image transformations such as rotation and translation.

Speeded Up Robust Features (SURF) is a feature detector that can be used in object recognition. It is useful for locating key points of the objects in the scene.

VGG is a well-known CNN based deep learning architecture where it has been used for image classification tasks. It is also commonly used to extract

features from the images. The method is called VGG16 feature extraction. To do this, input is forwarded to a pretrained VGG model and results before the last layer are fetched. These results can be used as image features. The last layer of the model is used for predicting the class of these features which consist of Imagenet dataset classes.

2.3.2. Neuron

Neuron is the core building block of neural networks. It takes input from the previous layer, applying matrix operations with weights and biases which are hyperparameters of the network. It uses an activation function and passes the result to the next layer.

2.3.3. Activation Functions

It is a mathematical operation used to learn complex features from the given data. It also provides non-linearity.

Sigmoid is a function that returns value between 0 and 1. One of the use cases is binary classification where only two different classes exist.
Formula:
$S(x)=1/(1+e^{-x})$

Tangent Hyperbolic Function (Tanh) is similar to sigmoid. It is a function that returns value between -1 and 1.
Formula:
$S(x)=(e\|x-e^{-x})/(e^{x}+e^{-x})$

Rectified Linear Unit (RELU) is a function that returns value 0 for negative values and value itself for positive input. ReLU is also a widely used activation function.
Formula:
$max(0,x)$

2.3.4. Convolutional Layer

It is a mathematical process that takes both input data and a small size filter(kernel) and extracts features from the input. Some of the predefined filters include edge detection, sharpening and blurring. It involves multiple neurons.

2.3.5. Pooling

Helps to down sample while preserving information. It reduces computational cost and fastens the train duration of the model.

In max pooling, a sliding window used on input data and maximum value in that window is chosen. The main idea is that the biggest value includes most of the information.

In average pooling, a sliding window used on input data and average value in that window is calculated.

2.3.6. Loss Functions

To be able to determine how well the model is, loss functions are used. It compares predicted value and actual value which is called ground truth.

Some of the common loss functions are:

- Mean Absolute Error (L1 Loss)
- Mean Square Error (L2 Loss)
- Hinge Loss

2.3.7 Softmax

It is a function that is mostly used at the end of the neural network. It returns the probability of each class that our input can belong.
Formula:
$e^x / \sum \left(e^x \right)$

2.3.8. Backpropagation

After calculating the loss value, back propagation is used to update hyperparameters such as weights. Overall process is applied multiple times and each iteration is called epoch.

2.4. Common Datasets Used for Face Recognition

There are lots of datasets that can be used for face recognition tasks. To get better results for both recognition and anti-spoofing, combinations of the datasets can be used. We will share some of the datasets used in face recognition and face anti-spoofing.

2.4.1. Labeled Faces in the Wild (LFW)

Labeled Faces in the Wild is one of the most commonly used dataset in the field of face detection and recognition. It includes more than 13,000 face images. Images in the dataset retrieved from the web.

Although it is widely used by the research community, it is stated that it may not be effective in commercial usage. The reasons are following:

- Dataset is not balanced according to ethnicity, sexuality and age.

- Dataset does not have enough subgroups.
- Environmental changes are not included in the majority of the dataset.

2.4.2. Replay-Mobile Dataset

Replay-Mobile Dataset is also a commonly used dataset for face recognition. It is also used for anti-spoofing. It involves 1190 video clips and different environmental conditions. It includes attack types such as displaying victims' images.

2.4.3. CASIA-MFSD

It is a dataset including videos for face anti-spoofing (FAS). It has a total of 600 videos for 20 different people. Including attack types are:

- Cut print attacks
- Replay attacks
- Warp print attacks

2.4.4. MSU-MFSD

It has a total of 280 videos for 35 different people. Videos including different resolutions retrieved from different devices. It only includes replay attacks.

2.4.5. OULU-NPU

It has a total of 5940 videos for 55 different people. The dataset was created using 6 different mobile devices. It includes printed attacks and video replay attacks taken under different environmental conditions.

CHAPTER 3 - VULNERABILITIES AND KNOWN ATTACKS

As face recognition methods started to be used everywhere, adversarials aim to find weaknesses of face recognition modules and try to exploit those vulnerabilities. There are two different attack types based on domain:

- Digital domain attacks
- Physical domain attacks

We can inspect the purposes of attacks under three categories:

- Evasion attacks
- Impersonation attacks
- Multiple identity attacks

3.1. Attack Types in Terms of Domain

3.1.1. Digital Domain Attacks

Digital domain attacks are the attack types where attackers can directly communicate with the deep learning model without any other interference in between. There is no camera to get the input from the attacker to convert analog data to digital. For this reason, although these attacks are not very

applicable in real life, they are useful to have information about the vulnerabilities of the system.

3.1.2. Physical Domain Attacks

A physical domain attack requires the attacker to manifest in the physical world. Examples of this attack are the attacker showing a printed image to the camera, wearing a similar mask to the victim's face, or showing objects that could fool the facial recognition system. Capturing a victim's image or video and using it against the targeted system is also a physical domain attack called replay attack.

3.2. Attack Types in Terms of Target Knowledge

3.2.1. White-box Attacks

In white-box attacks, the attacker has all the information about the system. For this reason, the attack is difficult and unrealistic to implement in real life. However, it can be useful to learn about the vulnerabilities of the model.

3.2.2. Gray-box Attacks

In gray-box attacks, the attacker has limited information about the target system. They may know which deep learning model is being used and the parameters, but they may not have access to the database where users' information is kept.

3.2.3. Black-box Attacks

In black-box attacks, the attacker has no knowledge of the target system. Attackers do not know what the deep learning model is and what its parameters are. For this reason, it is the type of attack that is most suitable for real life.

3.3. Attack Types in Terms of Attack Purpose

3.3.1. Evasion Attacks

Evasion attack takes place when the attacker wants to avoid face recognition systems. By this method, confidentiality and untraceability can be ensured by the attacker. This attack type is also a non-targeted attack where the attacker does not care what the model predicts other than the correct id.

Escaping from face detection models is getting harder day by day. Current state-of-the-art deep learning models are able to detect the human face in most cases even if face is occluded. For that reason, attackers focus on the recognition part.

Komkov *et al.* [3] designed a physical domain based evasion attack where the attacker wears a hat and puts a crafted rectangular sticker on it. Proposed attack, AdvHat, designed for a specific face recognition model called ArcFace. It has been stated that it can be also transferable to other face recognition models. Attack includes applying off-plane image transformation, projection and adding perturbations on the sticker. Sticker also designed by loss value of the ArcFace model. It has been achieved that similarity between original image and adversarial image has less than 0.2 similarity in most cases. However, it has been tested with a very small dataset and this attack requires a face recognition system to be publicly available.

Agrawal *et al.* [2] proposed a binary mask iterative method (BMIM) where an attacker detects facial landmarks using face landmark detection and applies occlusion on randomly selected landmarks. The LFW dataset is used in the research. This approach is iterated through 100 times on each original image and occluded adversarial images are obtained. Attack considered successful if the system can not match the adversarial example with the real person. It is a black-box attack that has been developed under the idea that there is no knowledge about the target system. However, it is a digital domain based attack where an attacker can directly give input to the system. Attack tested for three different models (MobileFace, MobileNet, and SphereFace) and transferability can be provided although success rate is not too high.

Studies show that for evasion, the attacker must have prior knowledge of the target system or direct access to the system in order to be successful. In addition, targeted systems had no anti-spoofing capabilities. In the real world, it is not very feasible to get information about the target system or to get direct access. In this case, we can deduce that the attack is only valid for the target system that has been known, and has no protection. Hereby, proposed methods will not have much effect on the real world.

3.3.2. Impersonation Attacks

Impersonation attack occurs when the adversarial tricks the system such that the system perceives the attacker as an authenticated user. Replay attack is an impersonation attack that is done by using the victim’s image. This attack is targeted such that the attacker wants to be seen as a specific person.

Zhang *et al.* [10] designed a physical domain based attack that can exploit a face recognition system designed with the CNN-based anti-spoofing capability. He developed the attack such that the attacker has the knowledge about the target system and anti-spoofing methodology. Target

system has three steps; face detection, spoofing detection, and face recognition. Target system uses a face_recognition python library which has been trained with LFW dataset. The REPLAY-MOBILE database is also used for getting facial features and used in anti-spoofing models. Adversarial should successfully pass the face detection and face recognition steps and exploit the spoofing detection module. They proposed adding small perturbation on input images such that it will affect loss function of spoofing detection to classify input as real image. To attack the system, they created a duplicate of the target system to achieve the digital output of the camera. Traces of the achieved image are removed to be able to escape from spoofing detection. It is changed and feeded to the original system such that noise will exist and be sensitive to environmental change.

Nguyen *et al.* [1] proposed an attack on CNN based face recognition system without prior information about the system and without access to the system. It is a black-box and physical domain attack. Only interaction with the target system takes place in the real world as analog input via cameras. Tests are done on the LFW dataset. In that study, it is assumed that the victim's photo is captured. Attacker uses another person's photo and by applying noise and image transformation, the face recognition system matches the victim with the attacker's crafted input. To be able to generate an attack successfully, they observed the properties of physical domain attacks. In physical domain attacks, position of the person, environmental conditions like illumination and quality of the sensor that takes the input are important. They created a formula which tries to minimize distance between target image and the crafted adversarial sample. They applied transformations to the adversarial image to be able to mimic the environmental changes. To make the attack generalizable, two different methods, Momentum Boosting and Ensemble, are used. It is stated that the attack was successful regardless of CNN model or loss function.

In this attack, the attacker needs a photo of the victim. In order to attack a system that includes CNN-based anti-spoofing, it is necessary to have knowledge about the system. On the other hand, if there is no anti-spoofing

in the system, it can be deduced that impersonation attacks are generalizable.

3.3.3. Multiple Identity Attack

Amada *et al.* [4] designed an attack where crafted input matches with multiple real users in the target database. The method also uses perturbations on input image, but this time an adversarial example is registered to the target database. Suggested deep learning model is trained with a loss function that is calculated using differences from adversarial examples and all other genuine images. It is stated that on white-box system, the attack is 99% successful. However, this attack is not transferable between different neural network based face detection algorithms. Furthermore, it is not very feasible to register adversarial examples to the target system in a real world scenario.

3.4. Common Attack Methods

Depending on the purpose of the attack, adversarial aims to create an adversarial image such that the attacker wants to maximize the difference between adversarial image and the original image in case of evasion attack. In case of impersonation attack, the attacker needs to minimize this difference. To achieve the successful attack and craft an adversarial input, there are common algorithms used in facial recognition attacks [11].

Fast Gradient Sign Method (FGSM) used for changing each pixel of the original image by using gradients of the loss function. There is also a multiplier that keeps the changes small enough to be not noticed by the human eye.

Basic Iterative Method (BIM) takes the adversarial image and uses the FGSM multiple times.

Momentum Iterative Method is an addition to BIM. It leads to more generalizable attacks that can be used on different targets. Term momentum widely used in deep learning models which helps to learn better.

Besides these attacks, [2] used an attack where specific points on the facial image are occluded. It has been shown that attacking on pixels where the deep learning models focus also leads to successful attack.

CHAPTER 4 - COUNTERMEASURES

4.1. Anti Spoofing

4.1.1. Cnn based anti-spoofing

Liu *et al.* [12] proposed a CNN-based anti-spoofing solution where it differs from other solutions. He created a domain-invariant system. The problem with the Face anti-spoofing (FAS) is that they are focusing on a dataset where the model is trained with specific adversarial dataset. To overcome this issue, they gathered inputs from different domains and applied feature extraction. Proposed feature extraction method includes domain-invariant features. After the feature extraction step, both adversarial losses are calculated. Model with a Common Specific Decomposition for Specific (CSD-S) layer used. It is stated that the CSD-S layer ignores domain-specific features and focuses on common features. By this approach, they get better results from similar approaches. Results are tested on datasets including:

- CASIA-MFSD
- MSU-MFSD
- Replay-Attack
- OULU-NPU

Sanghvi *et al.* [9] also created a solution for unseen settings. They gathered different kinds of physical attack types and datasets and created a CNN-based solution called MixNet. Attacks include printed faces, 3D masks, and replay attacks. MixNet's result is not only binary classification where input is classified as real or fake. The model outputs different attack types which makes it harder to exploit. Multiple loss functions used for training the model. Overall loss including replay loss, mask loss, and print loss. Used datasets are:

- SMAD
- SiW-M
- Replay-Attack
- MSU-MFSD

In another study, Tang *et al.* [5] focused on only 3D mask attacks since there is a lack of study in this area. They proposed a method where 3D facial features are created and used for face recognition. The method is built upon principal curvature measures. The method extracts geometric information of the face and its surface. Using the Morpho database, they show the model's anti-spoofing capability. To be able to get close to the real-world case, they combine two different datasets where genuine faces are outnumbered. It is stated that the method is able to distinguish minor differences between real face and masks. Used datasets for verification and anti-spoofing accuracy are:

- Morpho Database
- FRGC v2.0

Chen *et al.* [6] proposed a method, face anti-spoofing region-based convolutional neural network (R-CNN), using a combination of face detection and anti-spoofing. They classified the input as three categories: real face, fake face and background. The system handles both face detection and feature extraction at the same step. While training the network, Crystal loss was used next to the main loss function. They also used Retinex based LBP

to handle different illumination cases. Following datasets are used for calculating the performance of the system:

- CASIA-FASD
- REPLAY-ATTACK
- OULU-NPU

Fatemifar *et al.* [13] proposed a client-based anomaly detection solution for face spoofing attacks. Most of the anti-spoofing classifiers use 2 or more classes that distinguish genuine and fake faces. They used one-class classifier for each person and threshold to determine identification is different from each other. It has been stated that proposed methods are not sufficient for different datasets.

The difference in the datasets used and the designs of deep learning models make it difficult for attackers to find vulnerabilities. These methods can also be supported by image processing algorithms. Combining more than one method for further work can make the system more secure.

4.1.2. Liveness Detection

Liveness is a property that shows the subject is alive. Most common usages in face detection systems are eye blinking and head motions.

Cindori *et al.* [7] developed an eye blinking method next to a deep learning based face recognition. It is stated that this method is successful against printed attacks. It is also mentioned that the system is designed using lightweight technologies considering speed and performance. However, the study has been tested on a relatively small dataset.

Tu *et al.* [14] proposed a motion based anti-spoofing detection. The method works on video data. After extracting features with CNN, features are forwarded to the LSTM model which can handle the sequential motion data.

To handle the small motion differences, they used motion magnification methods. Performance of the system tested on 2 different datasets and it is stated that method is generalizable.

CNN based liveness detection has been proposed [15]. Authors were inspired by the ResNet model and created a model where it estimates if input is a living object or spoofing attack. This method is prone to attack types including print attack and replay attack. It is a light model that can be run in real time since it is not working on video but single image. Model trained on following datasets:

- NUAA
- CASIA-FASD

Fourati *et al.* [16] also created a real-time solution for motion based anti-spoofing. They proposed an Image Quality Assessment method. They collected frames where the motion cues are and extracted quality indexes from those images. Collected indexes are fed to the deep learning model to classify the real faces and fake ones.

Hadiprakoso *et al.* [17] proposed liveness detection where information from users' eyes and lips were extracted. They used biological facts related to eye blinking. Time passes between two distinct eye blink and frequency of blinking is measured. This method was not sufficient against video replay attacks. For this reason the authors built another CNN based detection module after the first detection. Second module distinguishes differences between adversarial samples and original samples. It is stated that adversarial inputs while designing replay attacks loses some of the essential features due to digital to analog to digital conversion. To train and test the model, datasets including video samples are used.

CHAPTER 5 - CONCLUSION

In conclusion, there are many studies based on both attacking and defending the face recognition models. However most of the studies are often not interconnected, focused on specific attack or counter measurement. The general topics that can be drawn from the research are listed below:

- Attacks performed on deep learning models that have no anti-spoofing capability mostly result in success whether its black-box or white-box attack.

- Attacks performed on deep learning models that have anti-spoofing capability may result in success but it requires information on the target system and it's hard to transfer the attack to another system.

- It is more effective to use multiple classification on anti-spoofing phase instead of only using genuine or fake classification.

- Combining multiple datasets while designing the attack makes the attack generalizable. It is also useful while training anti-spoofing models.

- Although there are studies that ensure the security of 3d face recognition systems, there is not enough research on attacks.

- Liveness detection is an effective countermeasure to eliminate adversarial attack. On the other hand it reduces the usability and adds extra cost to the system.

Although there are many studies on face recognition vulnerabilities, there is a need for a study that will show the general validity of all these studies. Because too many face recognition algorithms and systems used are different from each other. So, as a future work, requirements of secure face recognition systems can be standardized and improved based on developed attacks.

REFERENCES

[1] Nguyen, D. M., Nguyen, A. T., Tran, H. M., Le, N. T., & Quan, T. T. (2021). Physical transferable attack against black-box face recognition systems. *2021 International Conference on Multimedia Analysis and Pattern Recognition (MAPR)*. https://doi.org/10.1109/mapr53640.2021.9585256

[2] Agrawal, K., & Bhatnagar, C. (2021). BMIM: Generating adversarial attack on face recognition via binary mask. *2021 International Conference on Intelligent Technologies (CONIT)*. https://doi.org/10.1109/conit51480.2021.9498370

[3] Komkov, S., & Petiushko, A. (2021). AdvHat: Real-world adversarial attack on Arcface Face ID system. *2020 25th International Conference on Pattern Recognition (ICPR)*. https://doi.org/10.1109/icpr48806.2021.9412236

[4] Amada, T., Liew, S. P., Kakizaki, K., & Araki, T. (2021). Universal adversarial spoofing attacks against face recognition. *2021 IEEE International Joint Conference on Biometrics (IJCB)*. https://doi.org/10.1109/ijcb52358.2021.9484380

[5] Tang, Y., & Chen, L. (2017). 3D facial geometric attributes based anti-spoofing approach against mask attacks. *2017 12th IEEE International Conference on Automatic Face & Gesture Recognition (FG 2017)*. https://doi.org/10.1109/fg.2017.74

[6] Chen, H., Chen, Y., Tian, X., & Jiang, R. (2019). A cascade face spoofing detector based on face anti-spoofing R-CNN and improved RETINEX LBP. *IEEE Access*, *7*, 170116–170133. https://doi.org/10.1109/access.2019.2955383

[7] Cindori, D., Tomicic, I., & Grd, P. (2021). Security hardening of facial recognition systems. *2021 44th International Convention on Information, Communication and Electronic Technology (MIPRO)*. https://doi.org/10.23919/mipro52101.2021.9596961

[8] Hofbauer, H., Debiasi, L., & Uhl, A. (2019). Mobile Face Recognition Systems: Exploring presentation attack vulnerability and usability. *2019 International Conference on Biometrics (ICB)*. https://doi.org/10.1109/icb45273.2019.8987404

[9] Sanghvi, N., Singh, S. K., Agarwal, A., Vatsa, M., & Singh, R. (2021). MixNet for generalized face presentation attack detection. *2020 25th International Conference on Pattern Recognition (ICPR)*. https://doi.org/10.1109/icpr48806.2021.9412123

[10] Zhang, B., Tondi, B., & Barni, M. (2020). Adversarial examples for replay attacks against CNN-based face recognition with anti-spoofing capability. *Computer Vision and Image Understanding*, *197-198*, 102988. https://doi.org/10.1016/j.cviu.2020.102988

[11] Yang, X., Yang, D., Dong, Y., Su, H., Yu, W., & Zhu, J. (2021, September 29). *RobFR: Benchmarking adversarial robustness on face recognition*. arXiv.org. Retrieved January 30, 2022, from https://arxiv.org/abs/2007.04118

[12] Liu, M., Mu, J., Yu, Z., Ruan, K., Shu, B., & Yang, J. (2021). Adversarial learning and decomposition-based domain generalization for face anti-spoofing. *Pattern Recognition Letters*. https://doi.org/10.1016/j.patrec.2021.10.014

[13] Fatemifar, S., Arashloo, S. R., Awais, M., & Kittler, J. (2021). Client-specific anomaly detection for face presentation attack detection. *Pattern Recognition*, *112*, 107696. https://doi.org/10.1016/j.patcog.2020.107696

[14] Tu, X., Zhang, H., Xie, M., Luo, Y., Zhang, Y., & Ma, Z. (2019, January 17). *Enhance the motion cues for face anti-spoofing using CNN-LSTM architecture*. arXiv.org. Retrieved January 30, 2022, from https://arxiv.org/abs/1901.05635

[15] Lin, H.-Y. S., & Su, Y.-W. (2019). Convolutional neural networks for face anti-spoofing and liveness detection. *2019 6th International Conference on Systems and Informatics (ICSAI)*. https://doi.org/10.1109/icsai48974.2019.9010495

[16] Fourati, E., Elloumi, W., & Chetouani, A. (2019). Anti-spoofing in face recognition-based biometric authentication using image quality assessment. *Multimedia Tools and Applications*, *79*(1-2), 865–889. https://doi.org/10.1007/s11042-019-08115-w

[17] Hadiprakoso, R. B., Setiawan, H., & Girinoto. (2020). Face anti-spoofing using CNN classifier & face liveness detection. *2020 3rd International Conference on Information and Communications Technology (ICOIACT)*. https://doi.org/10.1109/icoiact50329.2020.9331977