Face Anti-spoofing based on Convolutional Neural Networks

2022 International Conference on Intelligent Data Science Technologies and Applications (IDSTA) | 978-1-6654-9960-6/22/$31.00 ©2022 IEEE | DOI: 10.1109/IDSTA55301.2022.9923172

Siyamdumisa Maphisa, Duncan Coulter
Academy of Computer Science and Software Engineering
University of Johannesburg
Johannesburg, South Africa
E-mail: siyamdu.m@gmail.com, dcoulter@uj.ac.za

***Abstract*—Biometrics technologies have gained increasing attention across different sectors in the past decade. Face recognition has proven to be one of these successful biometric technologies. For example, law enforcement uses face recognition for faster investigations, banks for identity confirmation, and different organisations for access control. However, face recognition has shortcomings regardless of its high successes, just like any biometrics technology. Face recognition technology is still susceptible to face spoofing attacks despite great efforts made by different researchers to combat such attacks. The study proposes an anti-spoofing model based on deep learning methods. Three different pipelines are implemented based on convolutional neural network (CNN) architecture. A hyper tuned baseline CNN, a convolutional neural network based on AlexNet architecture, and a neural network based on VGG16 architecture. The study benchmarked pipelines using the available face anti-spoofing detection datasets - the NUAA and CelebA datasets. The study measures these performance metrics for all the pipelines: accuracy, precision, recall, F1 score, AUC, and Roc curve. All three pipelines provided good results when tested against the selected datasets.**

Keywords— Face Spoofing Attacks, Face Recognition, Deep learning.

I. INTRODUCTION

Face recognition technology has grown in popularity over the last decade due to its ease of use, deployment, and non-invasive nature while producing acceptable results [1], [2], [3]. Corporations, the government, and society benefit from face recognition [4]. However, face recognition systems are susceptible to spoofing attacks such as print, replay, and 3D masks [1], [5], [6], [7]. According to [1], [3], [4], [5], [8], spoofing attacks occur when an invalid user impersonates a valid system user by falsifying a biometric trait and gaining unwanted access and advantage. As a result, several researchers and companies collaborate to build face anti-spoofing (FAS) technology to safeguard facial recognition systems. On the other hand, existing anti-spoofing systems do not adequately guard against spoofing assaults owing to variable circumstances [8].

Therefore, the study aims to implement an anti-spoofing model for a face-based authenticating system that will prevent a print/photograph attack. The study achieves its aim by implementing a neural network model based on CNN architecture, Baseline, AlexNet and VGG16 to identify such spoofing attacks. CNNs have been helpful for computer vision challenges image and video analysis.

The remainder of the paper is arranged as follows: Section II discusses the problem background. The related literature is elaborated in section III. Following that, the proposed model in section IV of the paper discusses the three different pipelines implemented by the study. The second part discusses the experiment setup demonstrating how the solution is implemented. The experiment setup and dataset are discussed in section V. Section VI contains the result and discussion section VII provides the conclusion to the paper and discusses possible future work.

II. PROBLEM BACKGROUND

There has been a surging interest in human automatic secure identification in the last decade, primarily based on unique personal physical and behavioural biometric data [1], [9]. The high number of security breaches and transaction frauds in non-biometric systems, which are prone to be cracked due to inherent vulnerabilities, such as stolen cards and shared passwords, to name a few, are one of the main reasons for such focus [9].

Fingerprint, hand geometry, palm print, voice, face, and a handwritten signature are all biometric data [9], [10]. Face stands out among these for its acceptability and recognition cost. Making it one of the best options for many applications [9]. From low-security uses such as social media and smartphone access control to high-security applications such as border control and video surveillance in critical areas [9].

However, this popularity comes at a cost: face recognition systems have become a popular target for spoofing attacks. A spoofing attack is a process by which an impostor subverts an identification process by displaying a forged bio-metric item of a registered user, therefore, gaining illegitimate access and advantage [1], [3], [9], [11]. Among the most well-known face spoofing attacks are:

- Photo Attack (Fig 1)
- Video / replay Attack (Fig 1)
- Mask attack (Fig 1)

A. Photo Attack

A photo attack occurs when the attacker acquires a genuine user's printed photo and uses it to spoof a face recognition system [12], [13]. The attacker presents this printed image to a sensor/camera of a face recognition system [13]. Alternatively, the attacker can use smartphones, tablets, or laptops with high device resolution to spoof a face recognition system [14]. Due to several factors, photo attacks

are the most critical attacks to protect against [14]. For example, printing colour images from a genuine user's face is highly inexpensive and straightforward [13],[14]. In addition, photo attacks mainly result from social media (Instagram, Twitter, etc.) since photos are available on such applications and are easily downloadable [13], [15].

B. Video Attack

With this type of face recognition attack, an attacker tries to spoof the system by replaying a video containing a face of a genuine person using digital devices and presenting it to a facial recognition sensor [16]. Like photo attacks, video acquisition of people intended to be impersonated is becoming increasingly accessible, thanks to the growth of public video sharing sites and social networks or even a hidden camera [14]. Another reason to use this type of attack is that it increases the likelihood of success by giving the displayed fake biometric sample a lifelike appearance [14].

C. Mask Attack

The artefact presented in this attack type is a 3D mask of the user's face. The attacker creates a three-dimensional reconstruction of the victim's face and shows it to the sensor/camera [14]. Mask attacks necessitate more skill than previous attacks and access to additional information to create a realistic mask of the genuine user [14]. Due to the depth elements in the facial features, the 3-D mask attack is a more advanced version of video and photo attacks [16].

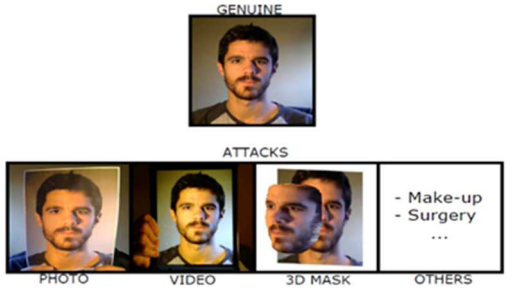

Fig. 1. Examples of face spoofing attacks: The top image represents genuine users while the bottom shows different presentation attacks [14].

III. Related Work

Face spoofing attacks have highlighted the need to continue researching effective ways to prevent such attacks. Several anti-spoofing techniques have been investigated to detect and possibly eliminate such vulnerability exploits over the past years. The following categories are used to classify anti-spoofing techniques.

- Face Anti-spoofing based on Motion Analysis Methods (A).
- Face Anti-spoofing based on Texture Analysis Methods (B).
- Face Anti-spoofing based on General Image Quality Assessment Methods (C).
- Face Anti-spoofing based on Hardware Methods (D).

A. Face Anti-spoofing based on Motion Analysis Methods

The methods in this category mainly focus on two-dimensional spoofing assaults like photos and videos [1]. They examine the motion characteristics of the input to assess the system's realism of the face sample [1]. For example, according to [1], it is generally known that humans blink once every 2–4 seconds. As a result, [15] developed a liveness detection method for photo-spoofing based on this spontaneous eye-blinking component.

Most motion-based methods work by extracting/ selecting a stable frame from a face recognition sensor (video camera/ infrared camera) and then segmenting it into foreground and background [17]. Each frame contains additional motion features that are extracted further [17]. When there is a lot of movement in the foreground and no movement in the background, the output is real; otherwise, the result is fake [17].

B. Face Anti-spoofing based on Texture Analysis Methods

The approaches in this category presume that real faces' surface qualities (e.g., pigments) differ from those of spoof prints; thus, looking at skin texture and reflectance can aid spoof detection [1]. Printing failures and blurring effects are the most prevalent visible texture patterns caused by artefacts [1]. Unlike approaches based on motion analysis, strategies in this category require a single static picture sample rather than video data [1]. As a result, these algorithms are typically quicker and easier to use.

C. Face Anti-spoofing based on General Image Quality Assessment Methods

The assumption with the use of IQM in face anti-spoofing is that "a false image taken in an attack attempt will have a different quality than a true sample obtained in the regular operating situation" [5], [18], [19]. Therefore, characteristics seen in the original pictures will be missing in the fake sample [5], [18], [19]. The use of IQMs for image quality characteristics allows for detecting quality differences between real and simulated samples [19].

Quality variations might include colour and brightness levels, information quantity, sharpness, structural deformities, or natural look [19], [20]. Image quality has been effectively used in past forensic work or image distortion identification and steg-analysis [5], [18]. Several face spoofing attacks, particularly those involving photographing a facial image exhibited in a 2-dimensional device (e.g., spoofing attacks with printed face images), can be considered image manipulation. Quality features can efficiently detect such spoofing attacks [5], [18].

D. Face Anti-spoofing based on Hardware Methods

Aside from software-based face anti-spoofing approaches, various hardware-based face liveness detection algorithms based on imaging have been developed [1]. According to [1], most techniques in this category use light spectrums that are not visible to the naked eye.

E. Face Anti-spoofing based on Deep learning.

Recent solutions to face anti-spoofing are CNN [21]. Different authors have trained CNN architectures to recognise which images are genuine and faked [21]. This technique assumes that the system would identify anything humans could not see with their naked eye [21].

F. Hybrid spoof-detection techniques

The literature contains several anti-spoofing techniques, but none are particularly effective in spotting the 3D face mask. In this category, two or more techniques are combined to enhance face anti-spoofing [22]. These techniques are known as hybrid techniques[22]. To identify attacks that use

2D faked faces, the depth information is typically combined with other anti-spoofing techniques[22].

IV. Model and Evaluation

This study aims to implement a face spoofing detection model based on deep learning. Three different pipelines are proposed based CNN architecture (Fig 2). Baseline CNN (the best hyperparameter-tuned CNN is chosen), Alexnet-based CNN, and a VGG16-based CNN are used. During training and testing, each pipeline's performance is evaluated. The performance metrics are confusion matrix, precision score, recall, accuracy score, F1 score, A receiver operating characteristic curve (ROC curve), and Area under the curve (AUC).

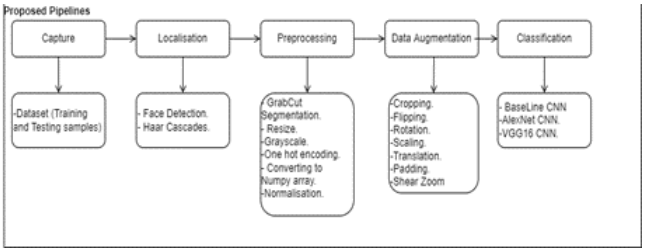

Fig. 2. The proposed model uses Convolutional Neural Networks.

A. Preprocessing

The study uses the NUAA database. Since the NUAA is an image dataset, the data in an image dataset is unstructured. There is a need to prepare and organise the data so that the proposed pipelines can read and analyse the dataset. For processing, the study uses different techniques (Fig 3). First, the study uses face detection using Haar cascades from the OpenCV library to capture only the face from the rest of the image. Then apply data augmentation to increase the amount of data to improve the model performance. Grab Cut algorithm is used to remove the remaining noise in the images. Finally, convert all the images to black and white and resize them to 240 by 240 pixels.

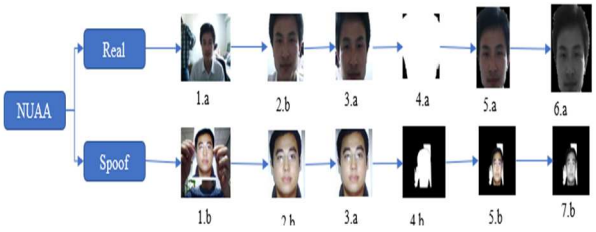

Fig. 3. Preprocessing steps: 1.a & 1.b are samples of the original NUAA dataset, 2.a & 2.b are samples after face detection process using Haar Cascade, 3.a & 3.b are samples after data augmentation process, 4.a and 4.b are samples of mask after applying GrabCut algorithm, 5.a & 5.b are samples of the output after applying GrabCut algorithm, 6.a & 6.b are samples of the final image after converting the images to black and white and resizing the images to 240X240 pixels.

B. Feature extraction and classification

Three different pipelines are implemented based on CNN architecture. A hyper tuned baseline CNN, a convolutional neural network based on AlexNet architecture, and a neural network based on VGG16 architecture. Since each pipeline is based on CNN architecture, each model's feature extraction is performed by the model based on the architecture of the respective model. Each pipeline then classifies the given image as either real or spoof.

V. Experiment Setup and Dataset

A. BaseCNN Architecture

The study used a Keras tuner to implement this architecture. A random hyperparameter study is conducted over several iterations to search the architecture's best parameters. For the input layer and hidden convolutional layers, the filter parameter search is between 32 minimum and 128 maximum values, and the stepping size is 32. Activation layer, the choice is between Relu, Tanh, and Sigmoid functions. The Hidden layers search 1 to 20., meaning the model can have between 1 and 20 hidden layers. After a sufficient number of iterations best parameters are saved and used to train and test the model.

B. Experiment Setup

In this section, the study discusses the performance of the proposed model and the adopted public available dataset: NUAA. The following section goes into the specifics of the dataset.

Face spoofing detection is a binary problem with two outcomes: positive (genuine) or negative (spoof) face. The study considered three CNN architectures to classify between positive and negative face spoofing samples. The baseline CNN, AlexNet, and VGG16 are used for training and testing using the adopted dataset. The goal is to train the model to classify whether the provided face in a biometric system is genuine or spoof. Accuracy, F1 Score, Precision, Recall, False Positive Rate (FPR), True Positive Rate (TPR), and Equal Error Rate (ERR) are all measured during training and testing. All the implementation was done in Python on Google Colaboratory and PyCharm.

C. NUAA Dataset

The NUAA Photograph Imposter Database (Fig 2), which is open to the public, incorporates pictures of each proper customer's right of entry and image assaults [23]. Each person's facial image graph is captured over the route of 3 periods separated using round weeks, with the ambient and light instances various from consultation to consultation. For every subject's recording, there are 500 photos [23]. The pictures within the database have been taken using well-known cameras and feature a decision of 640*480 pixels for 15 subjects [23].

Fig. 4. Face sample images of The NUAA Photograph Imposter Database [14]

For the experiment, all the images obtained from 15 different subjects in the original NUAA database are grouped

only into two classes, whether real or spoofs, since face spoofing is a binary classification problem. The real class contains 5105 images (Table 1), and the spoof class has 7509 (Table 1). The study then uses the train-test split model evaluation technique from the Scikit-learn library. Using a train-test split helps divide the dataset into a training set sample and a test set sample. The study uses the 60:40 split (Table 1) to split the data into 60% training and 40% for validation/ testing of the model.

D. CelebA- Spoof Dataset

A significant face anti-spoofing dataset called CelebA-Spoof has the following desirable qualities: 1) Quantity: CelebA-Spoof has 625,537 images of 10,177 patients, a significant increase over the datasets that are currently available. 2) [24]. For this experiment, a subset of the dataset was used (Table I).

TABLE I. DETAILS ABOUT THE NUAA DATABASE PARTITIONING FOR THE STUDY.

Database	Total Image set	Database Category/ Class name		Train-test split	
		Real (positive)	*Spoof (negative)*	*Training set (60%)*	*Testing set (40%)*
NUAA	12614	5105	7509	7568	5046
CalebA	10443	3503	6938	6266	4177

VI. RESULT AND DISCUSSION

A. Results

The results of the experiment are shown in Table II bel in terms of Accuracy, Precision, Recall, F1 score, and Area under the curve (AUC).

TABLE II. PERFORMANCE COMPARISON BETWEEN THE DIFFERENT PIPELINES

Dataset	*Pipeline*	*Accuracy*	*Precision*	*Recall*	*F1 Score*	*AUC*
NUAA Dataset	Baseline CNN	99.33 %	0.99	0.99	0.99	1.000
	AlexNet CNN	97.67 %	0.9745	0.9763	0.9789	1.000
	VGG16 CNN	98.6 %	0.9832	0.9876	0.9853	0.980
CelebA Dataset	Baseline CNN	78%	0.80	0.70	0.72	0.816
	AlexNet CNN	87%	0.89	0.83	0.85	0.936
	VGG16 CNN	92.7%	0.92	0.92	0.92	0.976

Secondly, this experiment uses a confusion matrix to measure how well each pipeline performed, classifying each class. The confusion matrix in Fig 5 – 7 is based on the NUAA dataset, and Fig 8 – 9 is based on CelebA dataset highlighting specific pictures that were incorrectly anticipated and vice versa. The false-positive observations allowed the experiment to determine how accurate each algorithm was compared to the others.

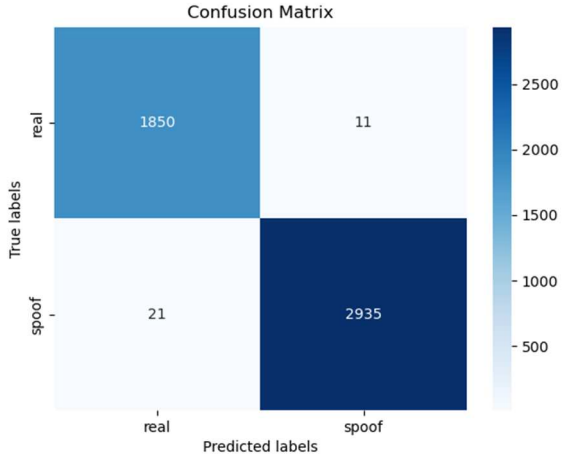

Fig. 5. Baseline CNN confusion matrix using NUAA dataset.

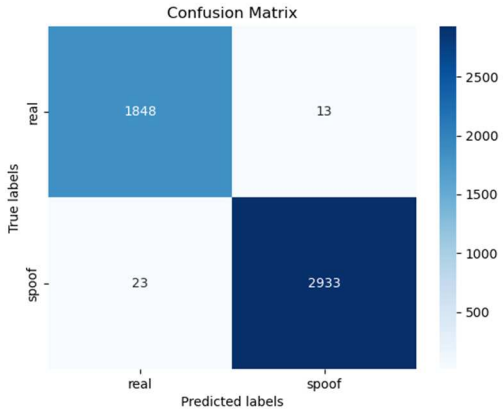

Fig. 6. AlexNet CNN confusion matrix using NUAA dataset.

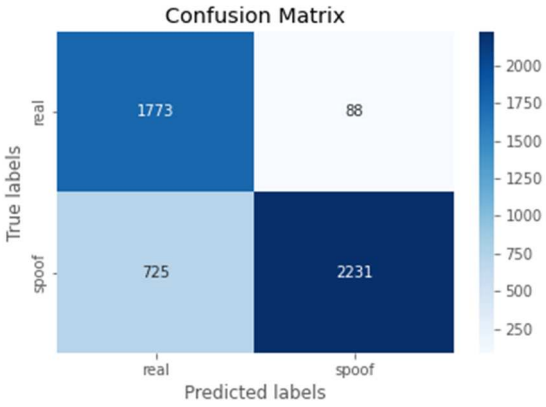

Fig. 7. VGG CNN confusion matrix using the NUAA dataset

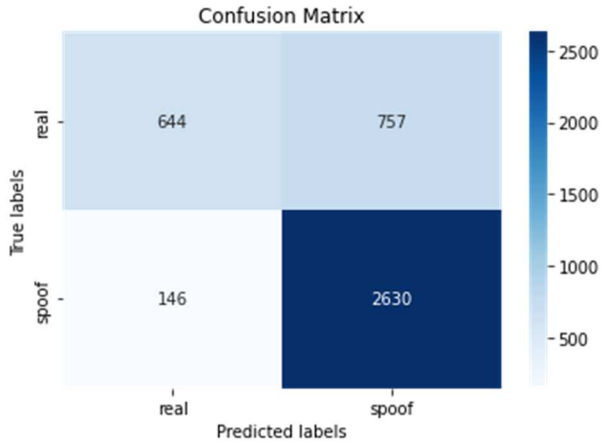

Fig. 8. Baseline CNN confusion matrix using CelebA dataset.

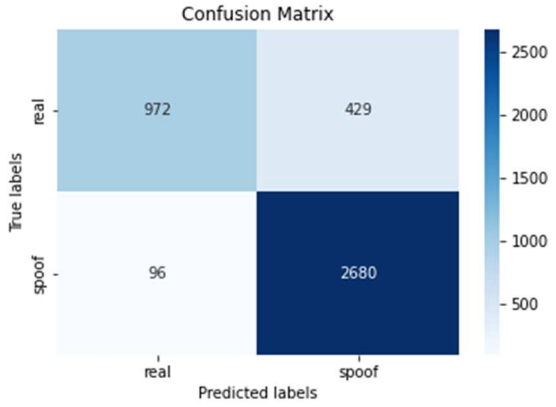

Fig. 9. AlexNet CNN confusion matrix using CelebA dataset.

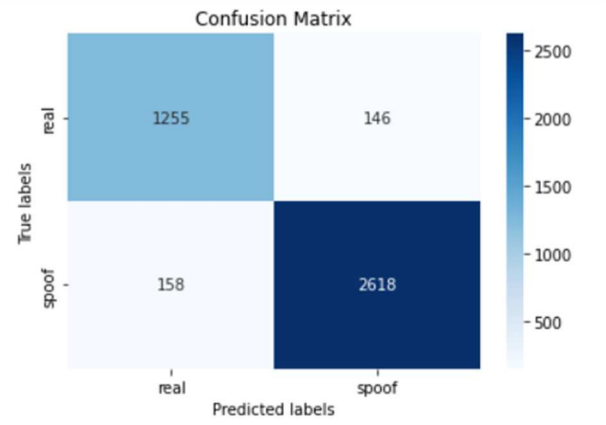

Fig. 10. VGG16 CNN confusion matrix using CelebA dataset.

The performance of each pipeline is also measured and compared using the Receiver operator characteristic curve (ROC curve). Figures 11 and 12 below show how each model performed during the experiment.

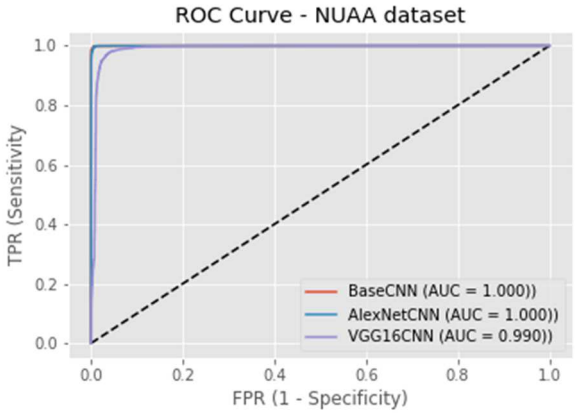

Fig. 11. Receiver operating characteristic curve using NUAA dataset.

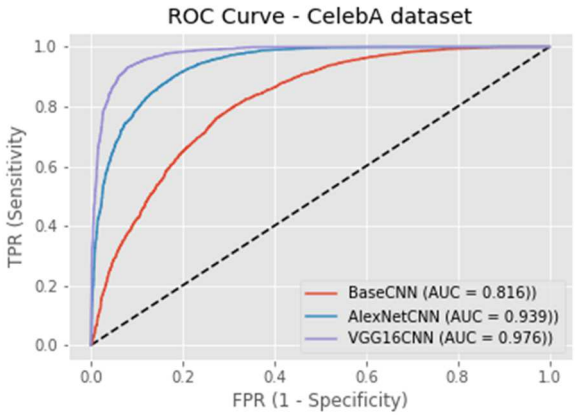

Fig. 12. Receiver operating characteristic curve using CelebA dataset.

B. Discussion

The performance of the model is given in table II. The table shows the accuracy of the three pipelines comparing the results between two adopted datasets. It is noted that the Baseline CNN performance is higher than the other pipelines when benchmarked using NUAA datasets. However, benchmarked using the CelebA dataset, the baseline CNN performance decreased drastically. When Observing other performance metrics (recall, precision, and f1 score), a similar trend is observed. The confusion matrix measures how each model is able to classify between live and spoof labels. All the models could predict the correct class for most of the images in NUAA datasets.

Regarding the CelebA dataset, the number of incorrectly classified images increased for all models. An assumption can be made for this observation, CelebA dataset is complex compared to the NUAA dataset. The ROC curve compares each pipeline against the other very well, and the illustrations can be used to conclude the study conducted by the paper. The VGG16 model performed relatively well in both datasets, which is evident in both ROC figures.

VII. Conclusion and Future Work

This study was able to train and compare three different CNN architectures. All the models proved to perform very well using the adopted datasets. The VGG16 architecture had excellent accuracy, confusion matrix, and ROC curve results

in both datasets. It came out top against the rest of the models implemented by the study.

The future work for this experiment is to use a more sophisticated database to benchmark the model. Doing this will help ensure that these face spoofing models can adapt to changing environments and learn and understand features that may be useful in determining whether a given face is genuine or spoofed. The study is also looking at using ensemble learning techniques to enhance the model performance in the future.

REFERENCES

[1] Z. Boulkenafet, Z. Akhtar, X. Feng, and A. Hadid, "Face Anti-spoofing in Biometric Systems," in *Biometric Security and Privacy*, Springer, Cham, 2017, pp. 299–321. DOI: 10.1007/978-3-319-47301-7_13.

[2] A. Harsoyo, M. C. Rezi, and P. H. Rusmin, "Design of face recognition system using local binary pattern and CLAHE on Smart Meeting Room System," in *2013 IEEE 3rd International Conference on System Engineering and Technology*, Aug. 2013, pp. 341–345. DOI: 10.1109/ICSEngT.2013.6650196.

[3] J. Maatta, A. Hadid, and M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis," in *2011 International Joint Conference on Biometrics (IJCB)*, Oct. 2011, pp. 1–7. DOI: 10.1109/IJCB.2011.6117510.

[4] G. M. Zafaruddin and H. S. Fadewar, "Face recognition: A holistic approach review," in *2014 International Conference on Contemporary Computing and Informatics (IC3I)*, Nov. 2014, pp. 175–178. DOI: 10.1109/IC3I.2014.7019610.

[5] J. Galbally and S. Marcel, "Face Anti-spoofing Based on General Image Quality Assessment," in *2014 22nd International Conference on Pattern Recognition*, Aug. 2014, pp. 1173–1178. DOI: 10.1109/ICPR.2014.211.

[6] R. B. Hadiprakoso, H. Setiawan, and Girinoto, "Face Anti-Spoofing Using CNN Classifier & Face liveness Detection," in *2020 3rd International Conference on Information and Communications Technology (ICOIACT)*, Nov. 2020, pp. 143–147. DOI: 10.1109/ICOIACT50329.2020.9331977.

[7] H.-Y. S. Lin and Y.-W. Su, "Convolutional Neural Networks for Face Anti-Spoofing and Liveness Detection," in *2019 6th International Conference on Systems and Informatics (ICSAI)*, Nov. 2019, pp. 1233–1237. DOI: 10.1109/ICSAI48974.2019.9010495.

[8] J. Komulainen, A. Hadid, M. Pietikainen, A. Anjos, and S. Marcel, "Complementary countermeasures for detecting scenic face spoofing attacks," in *2013 International Conference on Biometrics (ICB)*, Jun. 2013, pp. 1–7. DOI: 10.1109/ICB.2013.6612968.

[9] L. Souza, L. Oliveira, M. Pamplona, and J. Papa, "How far did we get in face spoofing detection?" *Engineering Applications of Artificial Intelligence*, vol. 72, pp. 368–381, Jun. 2018, DOI: 10.1016/j.engappai.2018.04.013.

[10] N. B. Sukhai, "Access control & Biometrics," in *Proceedings of the 1st annual conference on Information security curriculum development - InfoSecCD '04*, Oct. 2004, pp. 124–127. DOI: 10.1145/1059524.1059552.

[11] R. Ramachandra and C. Busch, "Presentation Attack Detection Methods for Face Recognition Systems," *ACM Computing Surveys*, vol. 50, no. 1, pp. 1–37, Jan. 2018, DOI: 10.1145/3038924.

[12] A. Alotaibi and A. Mahmood, "Enhancing computer vision to detect face spoofing attack utilising a single frame from a replay video attack using deep learning," in *2016 International Conference on Optoelectronics and Image Processing (ICOIP)*, Jun. 2016, pp. 1–5. DOI: 10.1109/OPTIP.2016.7528488.

[13] N. Daniel and A. Anitha, "A Study on Recent Trends in Face Spoofing Detection Techniques," in *2018 3rd International Conference on Inventive Computation Technologies (ICICT)*, Nov. 2018, pp. 583–586. DOI: 10.1109/ICICT43934.2018.9034361.

[14] J. Hernandez-Ortega, J. Fierrez, A. Morales, and J. Galbally, "Introduction to Face Presentation Attack Detection," in *Handbook of Biometric Anti-Spoofing*, 2019, pp. 187–206. DOI: 10.1007/978-3-319-92627-8_9.

[15] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based Anti-Spoofing in Face Recognition from a Generic Web camera," in *2007 IEEE 11th International Conference on Computer Vision*, 2007, pp. 1–8. DOI: 10.1109/ICCV.2007.4409068.

[16] S. Kumar, S. Singh, and J. Kumar, "A comparative study on face spoofing attacks," in *2017 International Conference on Computing, Communication and Automation (ICCCA)*, May 2017, pp. 1104–1108. DOI: 10.1109/CCAA.2017.8229961.

[17] S. Thepade, M. Dindorkar, P. Chaudhari, R. Bangar, and S. Bang, "The Comprehensive Review of Face Anti-Spoofing Techniques," *International Journal of Advanced Science and Technology*, vol. 29, pp. 8196–8205, Jun. 2020.

[18] J. Galbally, S. Marcel, and J. Fierrez, "Image Quality Assessment for Fake Biometric Detection: Application to Iris, Fingerprint, and Face Recognition," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 710–724, Feb. 2014, DOI: 10.1109/TIP.2013.2292332.

[19] P. Pravallika and K. S. Prasad, "SVM classification for fake biometric detection using image quality assessment: Application to iris, face and palm print," in *2016 International Conference on Inventive Computation Technologies (ICICT)*, Aug. 2016, pp. 1–6. DOI: 10.1109/INVENTIVE.2016.7823189.

[20] R. Karunya and S. Kumaresan, "A study of liveness detection in fingerprint and iris recognition systems using image quality assessment," in *2015 International Conference on Advanced Computing and Communication Systems*, Jan. 2015, pp. 1–5. DOI: 10.1109/ICACCS.2015.7324134.

[21] R. B. Hadiprakoso, H. Setiawan, and Girinoto, "Face Anti-Spoofing Using CNN Classifier & Face liveness Detection," in *2020 3rd International*

Conference on Information and Communications Technology (ICOIACT), Nov. 2020, pp. 143–147. DOI: 10.1109/ICOIACT50329.2020.9331977.

[22] L. Birla and P. Gupta, "PATRON: Exploring respiratory signal derived from non-contact face videos for face anti-spoofing," *Expert Systems with Applications*, vol. 187, p. 115883, Jan. 2022, DOI: 10.1016/j.eswa.2021.115883.

[23] S. Parveen, S. Ahmad, N. Abbas, W. Adnan, M. Hanafi, and N. Naeem, "Face Liveness Detection Using Dynamic Local Ternary Pattern (DLTP)," *Computers*, vol. 5, no. 2, p. 10, May 2016, DOI: 10.3390/computers5020010.

[24] Y. Zhang *et al.*, "CelebA-Spoof: Large-Scale Face Anti-Spoofing Dataset with Rich Annotations," 2020, pp. 70–85. DOI: 10.1007/978-3-030-58610-2_5.