Face Anti-Spoofing Using Chainlets and Deep Learning

Sarah Abdulaziz Alrethea, Adil Ahmad
Information Technology Department, King Abdulaziz University, Jeddah, Saudi Arabia

***Abstract*—Now-a-days, biometric technology is widely employed for many security purposes. Facial recognition is one of the biometric technologies that is increasingly utilized because it is convenient and contactless. However, the facial recognition system has become the most targeted by unauthorized users to get access to the system. Most facial recognition systems are vulnerable to face spoofing attacks. With the widespread use of the internet and social media, it has become easy to get videos or pictures of people's faces. The imposter can use these documents to deceive facial authentication systems, which affects the system's security and privacy. Face spoofing occurs when an unauthorized user attempts to gain access to a facial recognition system using presentation attack instruments (PAIs) such as photos, videos, or 3D masks of the authorized users. Therefore, the need for an effective face anti-spoofing (FAS) system is increased. That motivated us to develop a face anti-spoofing model that accurately detects presentation attacks. In our work, we developed a model that integrates handcrafted features based on Chainlets (as motion-based descriptor) and the convolutional neural network (CNN) to provide a more robust feature vector and enhance accuracy performance. Chainlets can be computed from deep contour-based edge detection using Histograms of Freeman Chain Codes, which provides a richer and rotation-invariant description of edge orientation that can be used to extract Chainlets features. We used a benchmark dataset, the Replay-Attack database. The result shows that the Chainlets-based face anti-spoofing method overcome the state-of-art methods and provide higher accuracy.**

Keywords*—Presentation attacks; Chainlets; contour; handcrafted features; chain code; CNN; face anti-spoofing

I. Introduction

In recent years, there has been an increasing interest in human recognition using biometric systems. Facial recognition systems are among the most popular biometric systems on the market because of their ease of use, high performance, and high level of security compared to iris and fingerprint recognition [1]. It is widely used in real-world applications, such as online payment and e-commerce security, smartphone-based authentication, secure access control, and others [2]. However, because of the widespread use of face recognition technologies, they have increasingly been the focus of Presentation Attacks (PAs). These attacks can be categorized into two types: (1) Impersonation attacks are an unauthorized user attempt to spoof someone else's identity, and (2) Obfuscation attacks are a user attempt to avoid being recognized by the system [3]. Face spoofing is when an unauthorized user tries to gain access to the facial recognition system by using one of the Presentation Attack Instruments (PAIs) like a photo (Print attack), a digital video (Replay attack), or wearing a mask (3D mask attack) of the authorized user [4] [5]. The most popular PAs are photo and video because they are easy to obtain and inexpensive compared to 3D mask attacks.

Face anti-spoofing (FAS) methods can be categorized into four main types: liveness, texture, 3D geometric, and multiple cues-based. 1- Liveness-based methods utilize motion in video frames to distinguish between real and spoofed faces. 2-Texture-based methods analyze the micro-texture of the object in front of the camera. 3- Three-dimensional geometric methods rely on the 3D structure of the user's face. 4- Multiple cues-based methods combine various techniques, such as integrating texture features with motion features [6].

In prior works, most face anti-spoofing methods based either on traditional methods, which use machine learning or deep learning. The first-mentioned is by extracting the handcrafted features and training the classifier to differentiate between fake and real faces [7]. It provides good performance in some situations; however, the researcher should have prior knowledge to design the features. Also, it cannot guarantee that the technique will function well in an unknown environment. The methods based on deep learning are more robust and concentrate on designing the model's structure, and are better able to handle a wide range of frequency responses. In general, deep learning models tend to outperform shallow ones [8]. The Convolutional Neural Network (CNN) architecture was proven that it is very effective for face anti-spoofing and superior to the other algorithms in terms of feature extraction [7]. However, optimizing the parameters would need a large number of training samples, which is not feasible for the existing face anti-spoofing databases [8].

In our proposed strategy, we use both handcrafted features and deep learning to obtain more robust features and produce a face anti-spoofing model with high accuracy [8]. Although there are many attack types with different scenarios, until now, there is no outstanding technique for face anti-spoofing systems. There is little attention paid to the development of a fusion strategy that incorporates various liveness features with multiple visual cues. Integrating liveness features from multiple cues will improve the detection accuracy of the face anti-spoofing system [9]. Deep learning models, particularly Convolutional Neural Networks (CNNs), have shown encouraging outcomes in numerous visual recognition tasks, leading researchers to use them in face anti-spoofing [2].

Integrating of handcrafted features and deep features has been used in face anti spoofing with promising results. The paper [10] integrates the handcrafted features with deep learning. It proposes an approach to enhance motion cues for face anti-spoofing using a CNN-LSTM architecture. This architecture combines Convolutional Neural Networks (CNNs) for spatial feature extraction and Long Short-Term Memory (LSTM) networks for capturing temporal dynamics across video frames. While Eulerian Motion Magnification is used to enhance subtle motion cues, it also degrades image quality. This can sometimes impact overall feature extraction negatively. Also, the temporal features are sensitive to low-quality frames affected by blur or noise, reducing effectiveness.

Consideration of the cues of motion across video frames requires taking into account the changes in objects, such as the face for each frame in terms of orientation and shape boundaries, which play a major role in detecting fundamental changes in impersonation. The paper in [11] introduces Chainlets, a novel object detection and recognition descriptor in computer vision. Chainlets build upon edge-based representations, improving upon existing methods like Histograms of Oriented Gradients (HOG) by incorporating edge connectedness and ordered orientation changes. This enhancement enables Chainlets to handle challenges like orientation invariance. While Chainlets were used for ear recognition and pedestrian detection and achieved competitive results with simple classifiers SVM, this motivated us to use Chainlets with Deep learning in the face anti-spoofing field to differentiate the fake and real faces. More specifically, in our work, we developed a model that integrates Handcrafted features based on Chainlets[11] (as motion-based-descriptor) and deep learning (CNN network) to enhance the performance in terms of accuracy.

The contributions of this paper are:

1) Utilizing a novel, generalized feature descriptor, Chainlets, for face anti-spoofing.

2) Built a model that integrates handcrafted features with CNN to get more robust features to improve the accuracy.

3) The experimental results demonstrate that Chainlets deliver better performance in face anti-spoofing than other algorithms.

This paper is organized as follows: Section II presents the literature review, Section III presents the methodology, Section IV discusses the experiments, and Section V presents the conclusions.

II. Related Work

Nowadays, Face anti-spoofing FAS is important for all facial recognition systems to discriminate between real and fake faces. Most FAS methods are based on either traditional methods or deep learning. The following sections briefly review the relevant works performed in face spoofing detection based on traditional and CNN based FAS techniques [7].

A. *Traditional FAS Methods*

Traditional FAS methods involve extracting handcrafted features and training classifiers to differentiate between fake and real faces [7]. Consequently, all existing Traditional FAS methods can be categorized into face anti-spoofing methods depending on motion information, texture information, 3D geometric, or multiple information. The methods based on handcrafted features provides good performance in some situations; however, the researcher should have prior knowledge to design the features. Also, it cannot guarantee that the technique will function well in an unknown environment [8].

1) Motion cue-based methods: The motion-based methods endeavor to detect liveness based on movement, for example, mouth movement or eye blinking or others. This method effectively detects photo attacks however provides poor performance in detecting video attacks [6].

The study in [12] proposed a method for Counter-Measures which is based on the correlation of foreground and background movement. This technique detects motion correlation between the scene's background and the user's face to indicate spoofing attack existence depending on optical flow (OF) to estimate direction. Additionally, histograms, normalization, comparison, and OFC quantization are employed to analyze frames for correlation. In the results and conclusion, this study presents a photo-attack dataset, which includes spoofing attacks under various conditions, and successfully detects photo attacks using motion correlation, achieving a 1.5% Half Total Error Rate (HTER).

These previous studies [13] [14] aim to propose methods for detecting liveness. The study in [13] presented an algorithm to evaluate liveness in face image sequences, with the goal of adding liveness awareness in a non-intrusive manner. This system is based on new Optical Flow (OF) and employs local Gabor decomposition and SVM experts. The system operates in real-time, with an input of three frames and an output of a liveness score; if the value of the score is 0, that means no liveness (fake face), and if it is 1, that means the maximum liveness (real face). The study's method depends on integrating the detection of face and the estimation of Optical Flow (OP) in order to indicate the liveness score. This integration is a combination between Optical Flow of Lines (OFL) by means of 1D Gaussians employment including derivatives and implementation of model based on a logpolar grid Gabor and SVM features. The XM2VTS database has been used for evaluation. The results show the success and robustness in detecting the liveness. Furthermore, the rate of error was about 0.5%. However, this method is not effective against video replay attacks.

The objectives related to this study [11] fall under presenting a new approach for detecting pedestrian and recognizing ear, proposing utilize of Chainlets as a new object descriptor. In computer vision applications, Hand-crafted models have shown encouraging results for pedestrian recognition. Based on these findings, human identification and other applications have adopted the Histogram of Oriented Gradients (HOG) features. Unlike traditional methods like Histograms of Oriented Gradients (HOG), which fail to model edge connectedness, Chainlets use Histograms of Chain Codes to enhance object descriptiveness and provide orientation invariance. Chainlets, which we refer to as normalized histograms of chain codes, are used as new descriptors for detection/recognition in the suggested unique method. In addition, experiments demonstrate that Chainlets perform better than HOG-based algorithms and also deep networks in particular cases. Speaking of methodology, this study's method using an ordered oriented data which is computed from deep contour based on edge detection; Chainlets. Chainlets produce strong results for pedestrian detection, boosting performance in comparison to previous hand-crafted descriptors and setting a new standard for biometric ear recognition. However, the current state of the art uses a quicker R-CNN and an improved classifier to enhance deep features for pedestrian identification. Even greater performance may be achieved by combining these improved classifiers with Chainlets. Aside from pedestrian detection and recognition, we feel that Chainlets have a lot of potential for usage in other areas such as face anti-spoofing.

This paper in [15] also utilized Chainlets, introducing an algorithm for ear recognition called the Chainlet-Based Ear Recognition method, which incorporates multi-banding image processing and support vector machine (SVM) classification. The approach involves splitting an ear image into multiple bands based on pixel intensity and using the Canny edge detection algorithm to extract edges in each band. These edge maps are combined into a single binary edge map representing the ear's contours. It achieved high accuracy with optimal band counts on two benchmark datasets: IITD II and USTB I.

2) Texture cue-based methods: The methods based on texture took use of the fact that a genuine face has a distinct illumination pattern and texture than the 3D mask, paper, or LCD surface utilized to perform the presentation attack [16]. It aims to explore the micro-texture of the object in front of the camera [6]. Texture feature-based methods utilize two kinds of features: static or dynamic. The Static texture cues extract spatial texture features from a single image. In comparison, the dynamic texture cues extract spatiotemporal texture features from video sequences [6]. In this method, the computational complexity is low, and the implementation is easy; however, with the recent and high-resolution cameras and high-quality pictures, the texture information is not enough, and we need to combine it with other information [17].

These previous studies [18] [19] aim to propose methods based-on color texture analysis. This paper [18] proposed a novel face anti-spoofing approach based on analysis of color texture. This study focus on using chrominance information in order to discriminate fake faces from real ones, while other studies have focused on evaluating the luminance of the face pictures. Related to methodology and experiments, the information relative to joint color texture, which comes from the luminance and the chrominance channels, has been analyzed based on a color local binary pattern descriptor. Furthermore, the histograms of feature have been extracted separately from each picture band. For the classification, they utilized a SVM with RBF Kernel. In the result, the investigation of the useful color space among HSV, RGB, and YCbCr spaces has been indicated. The performed experiments on CASIA and Replay-Attack databases have delivered excellent results taking into account a very promising results related to capabilities of generalization.

This paper in [20] proposed a new approach to detecting spoofing attacks based on texture analysis regarding characterization of printing artifacts, image quality assessment, and differences in light reflection. In addition, this study aims to present a new model by evaluating facial image textures in order to distinguish between live face and face in a printed image. This study considers the NUAA Photograph Imposter Database. The evaluation of the facial images' texture has been done by multi-scale local binary pattern (LBP). Speaking of methodology, the live objects must appear fixed as much as possible by minimizing the blinking motion of the eye and other movements. For the classification, they utilized a SVM with RBF Kernel (LibSVM). The results, this approach is robust and fast in calculations and demands no cooperation from the user, where the accuracy achieves 98.0%.

3) 3D Geometric cue-based methods: The 3D geometric-based methods use the depth information of the face to differentiate between a 2D face (photo or video attack) and a 3D face (real face). This method effectively detects photo and video attacks; however, it is not adequate to detect 3D mask attacks and is also costly because it needs additional hardware [17].

This paper in [21] proposed a new approach for detecting face liveness in order to eliminate spoofing attacks based on recovering sparse 3D facial structures. The methodology involves capturing faces in videos or photo sequences from at least three viewpoints, detecting facial landmarks, and selecting key frames. After that, the recovering operation of the sparse 3D facial structure can be done based on the chosen key frames. Eventually, a training process for the Support Vector Machine (SVM) is performed in order to differentiate the real faces from the fake ones. They evaluate the proposed method using three datasets, CASIA, NUAA, and Idiap databases. In the results of experiments, the proposed method has achieved ideal performance in detecting liveness where the accuracy achieved 100% for Face Liveness Detection. It outperforms the state-of-the-art methods in differentiating the genuine, planar, and warped image faces. The disadvantages of this method are that it provides good performance just for detecting photo attacks, not others. Also, it needs various viewpoints where one image is not enough. In addition, if the number of key frames is not enough, it can affect the performance.

This paper in [22] proposed a multispectral approach for FAS with not only (VIS) spectra imaging but also with the near infrared (NIR) one in order to execute VIS–NIR image consistency for spoofing detection purposes. Related to methodology and tools, correlation analysis has been employed in order to achieve a more compatible anti-spoofing method. The overall system consisted mainly of: (i) trained correlation confidence map to have a robust system, (ii) VIS–NIR correlation model, (iii) illumination robust feature extraction, and (iv) final classifier to perform detection. First, to avoid unbalanced illumination terms, input has been preprocessed by histogram equalization and face to eye position normalization, then region features. In addition, the information of correlation is calculated based on features of NIR and VIS photo patches. The proposed method has achieved an effective performance in handling spoofing attacks with different illumination, and also intra and cross dataset testing protocols on the three datasets (Ms-spoof, self-collected and PolyU-HSFD) based on extracting complementary features on VIS and NIR photos. The ACER (%) is 1.76 in the self-collected dataset, and 0.241 in Ms-spoof dataset.

4) Multiple cues-based methods: The multiple cues-based methods consider combining multiple cues such as motion information with texture information. This method detects varied kinds of face presentation attacks [6]. The methods based on multiple cues are superior in accuracy; however, it needs high Computations [17].

This paper in [23] proposed a liveness detection system that integrates motion and texture cues, exploiting scene context and eye blinks in a real-time system to non-intrusively resist spoofing attacks. The proposed method utilized the conditional random field model for eye blinks and LBP to compare the scene context to the reference images. Two databases were created for evaluating the system: a photo-imposter video database and a live face video database. The experimental results verified the system's reasonable ability to counter spoofing attacks and its robustness against camera shifts as well as its achievement of 99.5% accuracy. However, this method is unsuitable for real-life applications as the camera must remain fixed in the same position.

This paper in [9] proposed a multi-cues integration framework for face anti-spoofing utilizing a hierarchical neural network in order to enhance the ability of generalization. The proposed method is meant to fuse image quality cues and related motion in order to detect liveness. They extract image quality features using shearlet (SBIQF) and motion features using optical flow. In addition, they fuse features from multi cues using bottleneck representations that can effectively integrate various features of liveness. They evaluate the method using three databases: Replay-Attack, CASIA-FASD, and 3D-MAD datasets. Effective performance was achieved, with Equal Error Rates (EERs) of 0% in the Replay-Attack and 3D-MAD databases and 5.83% in the CASIA-FASD database.

B. *CNN-Based FAS Methods*

This section discusses strategies that exclusively employ deep learning techniques or combine shallow machine learning and deep learning techniques based on their performance, strengths, and weaknesses.

This paper in [24] proposed a study utilizing CNNs to learn features with high discriminative ability, with the aim of enhancing FAS effectiveness. The OpenCV detector was employed for face detection, followed by the refinement of bounding boxes around facial landmarks to encompass the majority of facial regions. Images were resized to 128 × 128 pixels, and the Caffe toolbox was used for CNN training with a learning rate of 0.001, decay of 0.001, and momentum of 0.9. For classification, SVM with RBF kernel was employed. The proposed method was evaluated using Replay-Attack and CASIA datasets, achieving significant improvements over previous methods, with a 5% reduction in HTERs. However, the proposed method exhibited unsatisfactory performance in inter-test scenarios.

This paper in [10] proposed an FAS method based on a joint CNN-LSTM network, considering motion cues across video frames. This study extracted high discriminative properties and features from video frames using conventional CNN. The combination of these extracted features with Long Short-Term Memory (LSTM) was employed to capture temporal dynamics within video sequences. Additionally, Eulerian motion magnification was utilized to enhance facial expressions, while the attention mechanism in LSTM ensured the model focused on dynamic frames. The Adam optimizer was used for training the network with a learning rate of 0.00001. The proposed method was evaluated using Replay-Attack and MSU-MFSD datasets, achieving improved performance in generalization ability and fine-grained motion detection, with an HTER of 3.53% and an ACC of 96.47% in the Replay-Attack dataset. However, the experimental results indicated that dynamic variations in temporal features are beneficial for enhancing generalization ability.

This paper in [25] proposed an FAS method using CNN alongside a handcrafted technique (LBP-TOP) for feature extraction and classifier training. This method eliminates the need for detection, refinement, or enlargement steps. A cascading process for the LBP-TOP technique with CNN was performed to extract spatiotemporal features and obtain discriminative clues between genuine access and impersonation attacks. The proposed method was evaluated using the Replay-Attack and CASIA datasets, demonstrating significant potential for employing LBP-TOP with CNN for anti-spoofing purposes. An EER of 3.22% and HTER of 4.70% was achieved in the Replay-Attack dataset. Experimental results indicated competitive performance compared to previous methods.

This paper in [8] proposed a face anti-spoofing method regarding the deep local binary pattern. This study depends on extracting the handcrafted features from the convolutional responses related to tuned finely CNN model. First, the CNN has been finely tuned using a trained model of VGG-face. After that, the features of LBP are computed from the convolutional responses and combined in unique vectors of feature. These vectors are used by a SVM classifier in order to detect spoof faces. Related to samples, two databases are considered CASIA-FA and Replay-Attack. In the results, the investigation of the useful color space among HSV, RGB, and YCbCr spaces has been indicated. The experiments have delivered excellent results taking into account very promising results related to capabilities of generalization. The EER (%) is 0.1, and HTER (%) is 0.9 in Replay-Attack dataset. The EER (%) is 2.3 in CASIA-FA dataset.

This paper in [26] proposed a new approach for detecting face spoofing by extracting local features, specifically LBPs and simplified Weber local descriptors (SWLDs). The integration of Weber local descriptors (WLD) and LBP features ensures preservation of the local intensity information and edge orientations. The methodology involved two streams: an LBP-based CNN and a SWLD-based CNN. Initially, face regions were localized using Viola-Jones face detection. Subsequently, LBP and SWLD features were extracted by dividing the image into smaller parts. The features were encoded using CNN. Eventually, a non-linear SVM classifier with a radial basis function kernel was employed to determine whether a face was live or fake. The proposed method was evaluated using the Replay-Attack and CASIA datasets and achieved competitive results, with an EER and HTER of 2.62% and 2.14%, respectively, in the CASIA dataset and 0.53% and 0.69%, respectively, in the Replay-Attack dataset. This method's minimal complexity makes it suitable for real-time applications.

This paper in [27] proposed a new novel architecture consisting of a two-stream Frequent-Spatial-Temporal-Net for face anti-spoofing, which mainly combines the advantages of temporal, frequent, and spatial information. In addition, this study aims to achieve a multi-frame spectrum photo for the discriminative deep neural-network in order to distinguish between fake and live video. This study presents first the substantial properties related to the model. Then, elaborating the proposed network architecture is employed beside a strategy of learning in the pipeline. The implementation in MXNet and adoption of 4 blocks ResNet-v1b-18 for spatial part is used. While 0.3 rate of learning, 0.0001 weight decay, and 0.03 rate of Freq-Temp-Net stream learning are adopted. They evaluate the proposed method using three datasets: CASIA-SURF, OULU-NPU, and SiW datasets. Experimental results show a promising enhancement considering the proposed method besides the improved demonstrated ability of generalization compared with previous and existing approaches where the ACER (%) is 0.016 in CASIA-SURF, 1.1 in OULU, and 0.67 in SiW.

III. Methodology

To fulfill this study's objectives, we proposed designing a face anti-spoofing model capable of discriminating between real and fake faces with high accuracy. In Fig. 1, the proposed algorithm consists of five main stages: image pre-processing, Edge Detection, finding contours, extracting chainlets Features, and CNN model. The Chainlets were chosen as a motion-based descriptor due to their proven effectiveness in pedestrian detection and biometric ear recognition [11].

Chainlets is a concept that suggests that each object edge in an image can be represented as a histogram of chain codes, with each chain code representing the direction of connected edge segments. The label shifts between two absolute chain code-based Chainlets histograms can be used to determine the approximate rotation. Absolute chain code is dependent on the image’s edge orientation, which results in different codes for rotated versions of objects. The frequency of occurrence of each of the eight directions, which depicts the shape of the object, is calculated by a chain code histogram. The picture sequence’s Chainlets are first computed and concatenated with the sequence’s first image, which can characterize motion direction and orientation.

A. Image Preprocessing

Before extracting Chainlets, the image needs to be preprocessed. Fig. 2 shows the stages of image processing. First, the image will converted to grayscale, and then GaussianBlur will be applied to reduce noise from an image. cv2.GaussianBlur(img, (5,5), 0).

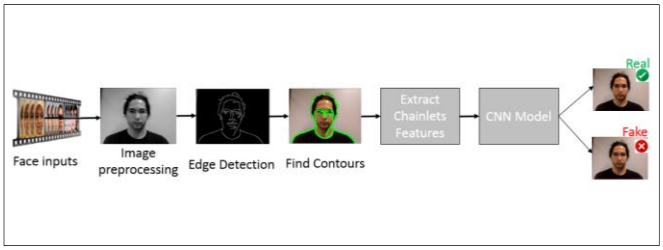

Fig. 1. The proposed algorithm.

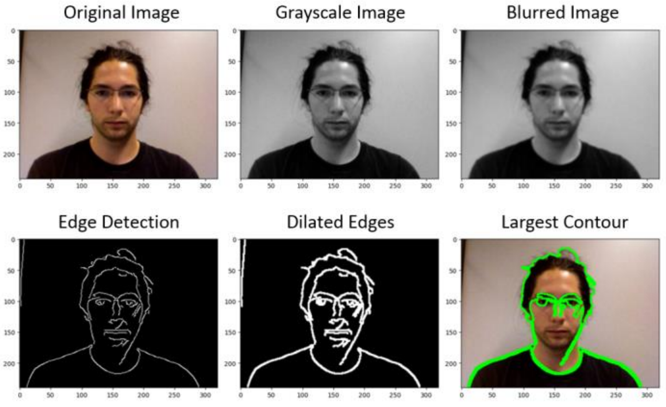

Fig. 2. Stages of image processing.

B. Edge Detection

The Canny edge detection algorithm [28] is a classic method that has proven its effectiveness over the years and remains widely utilized. Its approach is grounded in three main goals: (i) minimizing error rates, (ii) accurately locating edge points, and (iii) producing thin edges. We used canny edge detection with threshold values minVal and maxVal.

cv2.Canny(img, 50, 150)

Then the detected edges are dilated using a 3 x 3 kernel to close any gaps in the edges.

C. Finding Contours

Contours can be viewed as a plot or curve that links all the consistent points—typically found along the edges of a digital image—that exhibit similar colors or intensities. These connected points are associated with one another. Contours serve as useful tools for various applications, including structural analysis, object segmentation, object recognition, and others [29]. We used the find contours function from OpenCV and it has three parameters: the input image, contour retrieval mode, and contour approximation method.

cv2.findContours(img, cv2.RETR_EXTERNAL, cv2.CHAIN_APPROX_SIMPLE)

D. Extracting Chainlet Features

The representation of an image plays a crucial role in image processing and facial recognition. A key feature of an image is conveyed through the shapes created by the edges of an object. Histograms of Oriented Gradients HOG capture some of this information; however, it fails to account for important spatial relationships concerning the edges of an object. Specifically, HOG does not consider the connectedness of the edges, and the sequence and arrangement of orientation changes are not reflected in the histogram. The reason for using Chainlets as motion-based descriptor is that a face's appearance and shape can be effectively represented by the density of relative chain codes, which encode rotation-invariant edge detection. While this idea is related to HOG, it utilizes longer, connected edges, offering a more comprehensive and rotation-invariant representation of edge orientation [11].

We extract handcrafted features using the Chainlets which are built on the concept of Freeman chain codes [30] by making them rotation-independent and more localized, thus mitigating the negative impacts of partial occlusions. It is characterized by local edge orientations that are aggregated into a global histogram sequence representing the entire image.

Chain code [31] is a method for documenting a sequence of edge points along a contour, indicating the direction of each edge in the sequence. These directions are categorized into eight distinct orientations. The notation progresses clockwise around the contour, starting from the first edge. Each subsequent edge's direction is represented by one of the eight chain codes. The direction associated with the 8-neighbor of an edge is illustrated in Fig. 3.

A chain code histogram measures how often each of the eight directions occurs, indicating the shape of the face. Fig. 4 illustrates the steps of extracting Chainlets features. Firstly, divide the binary image into cells, then compute the chain code for each one, compute chain code histogram and normalized, the all blocks' collection represents the Chainlets features. Finally, the extracted chainlets features are passed to the CNN model.

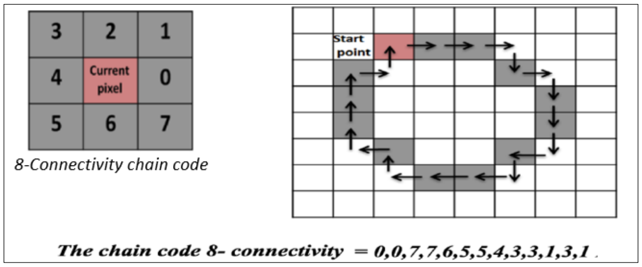
3
2
1
4
Current pixel
0
5
6
7
8-Connectivity chain code
Start point
The chain code 8- connectivity = 0,0,7,7,6,5,5,4,3,3,1,3,1 .

Fig. 3. Example of extracting chain code (clockwise).

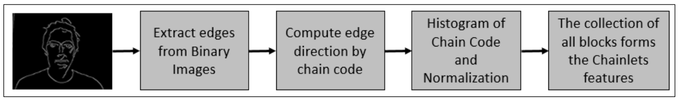
Extract edges from Binary Images
Compute edge direction by chain code
Histogram of Chain Code and Normalization
The collection of all blocks forms the Chainlets features

Fig. 4. Extracting chainlets features.

E. CNN Model

We build a CNN model for learning features. Our CNN model consists of three convolutional layers and 3 dense layers. The response normalization layers are applied to the outputs of the three Conv layers. Max pooling layers are utilized to generate the output for the three convolutional layers. Moreover, the 4th, 5th, and 6th layers are Dense layers. Further, we add the Dense layer with sigmoid activation to produce the binary output. We use Adam optimizer with a learning rate of 0.001. When compiling the model we use Binary Cross Entropy as the loss function to train the model for binary output, 1 is spoof face and 0 is real face.

IV. Experimental Results

We used a benchmark dataset, which is Replay-Attack. The Replay-Attack dataset contains 1,300 videos from 50 subjects, including real and fake face-spoofing attacks. For every subject, four genuine and eight fake sequences are obtained under various lighting terms with two support conditions (Fixed and Hand-held). Fig. 5 Shows example images from the Replay-Attack database. The images in the top row are genuine, while those in the bottom row are fake.

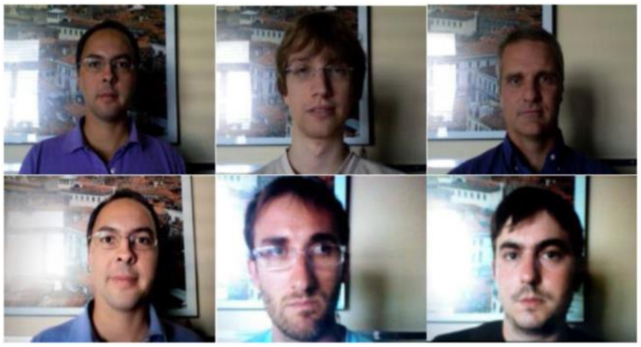

Fig. 5. Example of images from the Replay-Attack database.

The training set contains 93679 frames, the testing set 23395 frames, and the validation data size is 21055 frames. We evaluated the proposed model's performance to prove its effectiveness using the following metrics: Half Total Error Rate (HTER), Equal Error Rate (EER), and accuracy. The experimental results clearly demonstrate that the proposed Chainlets-based face anti-spoofing technique significantly surpasses state-of-the-art methods. This performance can be justified by the fact that integrating Chainlets as handcrafted features with deep learning will extract valuable features and produce a high-performance face anti-spoofing technique. Table I and Fig. 7 show that the proposed technique achieves the best performance compared with other state-of-art methods. Our model achieves high accuracy at 99.85%, the HTER is 0.15%, and the EER is 0.15%. Fig. 6 presents the training and validation accuracy and loss for the proposed algorithm.

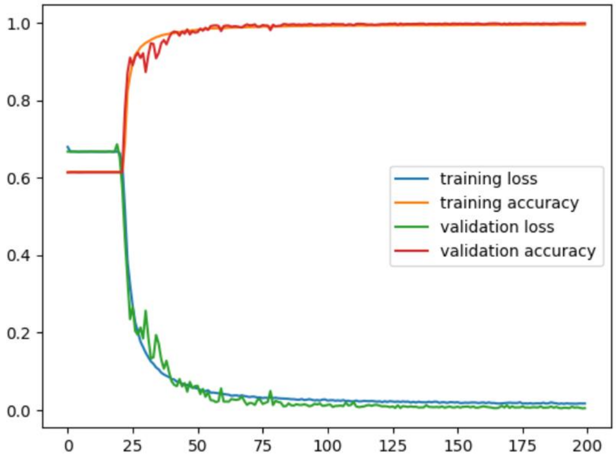

Fig. 6. Training and validation accuracy / loss curves.

TABLE I. Evaluation of Model's Performance Against State-of-the-art Methods

Approach	Replay-Attack Dataset		
	EER(%)	*HTER(%)*	*Accuracy(%)*
CNN + LBP-TOP [25]	3.22	4.70	-
DeepLBP + CNN + SVM [8]	0.1	0.9	-
CNN with LBP and WLD + SVM [26]	0.53	0.69	-
CNN-LSTM [10]	-	3.53	96.47
Chainlets + CNN (Our method)	**0.157**	**0.153**	**99.85**

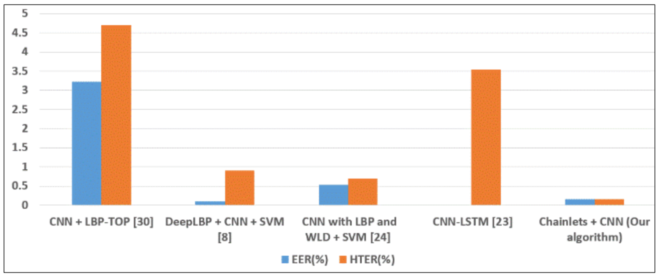

Fig. 7. The EER and HTER for the Replay-Attack dataset.

V. Conclusion

Given the current widespread use of the internet and social media, it is easy to obtain videos or pictures of individuals' faces, making most facial recognition systems vulnerable to spoofing attacks. An imposter can use such readily available materials to deceive facial authentication systems, thereby compromising their security and privacy. There are many attack types with different scenarios, so until now, there is no outstanding technique for face anti-spoofing. In our proposed algorithm, we depend on both handcrafted features (Chainlets) and deep learning (CNN) to obtain more robust features and high accuracy in the face anti-spoofing system. Thorough experiments were conducted on the Replay-Attack dataset to assess the effectiveness of the proposed Chainlets-based face anti-spoofing method. Our approach achieved perfect discrimination between real faces and spoof faces. The EER is 0.157%, the HTER is 0.153%, and the accuracy is 99.85%, which is better than the state-of-the-art methods. However, this work does not address the detection of 3D mask attacks. This aspect will be considered in future work.

References

[1] S. P. Borra, N. Pradeep, N. Raju, S. Vineel, and V. Karteek, "Face Recognition based on Convolutional Neural Network," International Journal of Engineering and Advanced Technology, vol. 9, May 2020, doi: 10.35940/ijeat.D6658.049420.

[2] R. G. Bhati and S. Gosavi, "A SURVEY ON FACE ANTI-SPOOFING METHODS," Jan. 2024, Accessed: Oct. 07, 2024. [Online]. Available: http://localhost:8080/xmlui/handle/123456789/16939

[3] Z. Yu, Y. Qin, X. Li, C. Zhao, Z. Lei, and G. Zhao, "Deep Learning for Face Anti-Spoofing: A Survey," IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 45, no. 5, pp. 5609–5631, May 2023, doi: 10.1109/TPAMI.2022.3215850.

[4] V. Sundharam, A. Sarkar, and A. L. Abbott, "Re-evaluation of Face Anti-spoofing Algorithm in Post COVID-19 Era Using Mask Based Occlusion Attack," arXiv.org. Accessed: Oct. 09, 2024. [Online]. Available: https://arxiv.org/abs/2408.13251v1

[5] Y. Wang, X. Song, T. Xu, Z. Feng, and X.-J. Wu, "From RGB to Depth: Domain Transfer Network for Face Anti-Spoofing," IEEE Transactions on Information Forensics and Security, vol. 16, pp. 4280–4290, 2021, doi: 10.1109/TIFS.2021.3102448.

[6] Z. Ming, M. Visani, M. M. Luqman, and J.-C. Burie, "A Survey on Anti-Spoofing Methods for Facial Recognition with RGB Cameras of Generic Consumer Devices," Journal of Imaging, vol. 6, no. 12, Art. no. 12, Dec. 2020, doi: 10.3390/jimaging6120139.

[7] T. Shen, Y. Huang, and Z. Tong, "FaceBagNet: Bag-Of-Local-Features Model for Multi-Modal Face Anti-Spoofing," presented at the Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, 2019, pp. 0–0. Accessed: Apr. 09, 2022. [Online]. Available: https://openaccess.thecvf.com/content_CVPRW_2019/html/CFS/Shen_FaceBagNet_Bag-Of-Local-Features_Model_for_Multi-Modal_Face_Anti-Spoofing_CVPRW_2019_paper.html

[8] L. Li and X. Feng, "Face Anti-spoofing via Deep Local Binary Pattern," in Deep Learning in Object Detection and Recognition, X. Jiang, A. Hadid, Y. Pang, E. Granger, and X. Feng, Eds., Singapore: Springer, 2019, pp. 91–111. doi: 10.1007/978-981-10-5152-4_4.

[9] L. Feng et al., "Integration of image quality and motion cues for face anti-spoofing: A neural network approach," Journal of Visual Communication and Image Representation, vol. 38, pp. 451–460, Jul. 2016, doi: 10.1016/j.jvcir.2016.03.019.

[10] X. Tu, H. Zhang, M. Xie, Y. Luo, Y. Zhang, and Z. Ma, "Enhance the Motion Cues for Face Anti-Spoofing using CNN-LSTM Architecture," arXiv:1901.05635 [cs], Jan. 2019, Accessed: Feb. 17, 2022. [Online]. Available: http://arxiv.org/abs/1901.05635

[11] A. Ahmad, D. Lemmond, and T. E. Boult, "Chainlets: A New Descriptor for Detection and Recognition," in 2018 IEEE Winter Conference on Applications of Computer Vision (WACV), Mar. 2018, pp. 1897–1906. doi: 10.1109/WACV.2018.00210.

[12] A. Anjos, M. M. Chakka, and S. Marcel, "Motion - based counter - measures to photo attacks in face recognition," IET biom., vol. 3, no. 3, pp. 147‑158, Sep. 2014, doi: 10.1049/iet-bmt.2012.0071.

[13] K. Kollreider, H. Fronthaler, and J. Bigun, "Non-intrusive liveness detection by face images," Image and Vision Computing, vol. 27, no. 3, pp. 233–244, Feb. 2009, doi: 10.1016/j.imavis.2007.05.004.

[14] W. Bao, H. Li, N. Li, and W. Jiang, "A liveness detection method for face recognition based on optical flow field," in 2009 International Conference on Image Analysis and Signal Processing, Apr. 2009, pp. 233–236. doi: 10.1109/IASP.2009.5054589.

[15] M. M. Zarachoff, A. Sheikh-Akbari, and D. Monekosso, "Chainlet-Based Ear Recognition Using Image Multi-Banding and Support Vector Machine," Applied Sciences, vol. 12, no. 4, Art. no. 4, Jan. 2022, doi: 10.3390/app12042033.

[16] C. Nagpal and S. R. Dubey, "A Performance Evaluation of Convolutional Neural Networks for Face Anti Spoofing," in 2019 International Joint Conference on Neural Networks (IJCNN), Jul. 2019, pp. 1–8. doi: 10.1109/IJCNN.2019.8852422.

[17] M. Zhang, K. Zeng, and J. Wang, "A Survey on Face Anti-Spoofing Algorithms," Journal of Information Hiding and Privacy Protection, vol. 2, no. 1, pp. 21–34, 2020, doi: 10.32604/jihpp.2020.010467.

[18] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face anti-spoofing based on color texture analysis," in 2015 IEEE International Conference on Image Processing (ICIP), Sep. 2015, pp. 2636–2640. doi: 10.1109/ICIP.2015.7351280.

[19] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face Spoofing Detection Using Colour Texture Analysis," IEEE Transactions on Information Forensics and Security, vol. 11, no. 8, pp. 1818–1830, Aug. 2016, doi: 10.1109/TIFS.2016.2555286.

[20] J. Määttä, A. Hadid, and M. Pietikäinen, "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.

[21] T. Wang, J. Yang, Z. Lei, S. Liao, and S. Z. Li, "Face liveness detection using 3D structure recovered from a single camera," in 2013 International Conference on Biometrics (ICB), Jun. 2013, pp. 1–6. doi: 10.1109/ICB.2013.6612957.

[22] X. Sun, L. Huang, and C. Liu, "Multispectral face spoofing detection using VIS–NIR imaging correlation," Int. J. Wavelets Multiresolut Inf. Process., vol. 16, no. 02, p. 1840003, Mar. 2018, doi: 10.1142/S0219691318400039.

[23] G. Pan, L. Sun, Z. Wu, and Y. Wang, "Monocular camera-based face liveness detection by combining eyeblink and scene context," Telecommun Syst, vol. 47, no. 3, pp. 215–225, Aug. 2011, doi: 10.1007/s11235-010-9313-3.

[24] J. Yang, Z. Lei, and S. Z. Li, "Learn Convolutional Neural Network for Face Anti-Spoofing," arXiv:1408.5601 [cs], Aug. 2014, Accessed: Mar. 25, 2022. [Online]. Available: http://arxiv.org/abs/1408.5601

[25] M. Asim, Z. Ming, and M. Y. Javed, "CNN based spatio-temporal feature extraction for face anti-spoofing," in 2017 2nd International Conference

on Image, Vision and Computing (ICIVC), Jun. 2017, pp. 234–238. doi: 10.1109/ICIVC.2017.7984552.

[26] M. Khammari, “Robust face anti-spoofing using CNN with LBP and WLD,” IET Image Processing, vol. 13, no. 11, pp. 1880–1884, 2019, doi: 10.1049/iet-ipr.2018.5560.

[27] Y. Huang, W. Zhang, and J. Wang, “Deep Frequent Spatial Temporal Learning for Face Anti-Spoofing,” arXiv:2002.03723 [cs, eess], Jan. 2020, Accessed: Mar. 25, 2022. [Online]. Available: http://arxiv.org/abs/2002.03723

[28] E. A. Sekehravani, E. Babulak, and M. Masoodi, “Implementing canny edge detection algorithm for noisy image,” Aug. 12, 2021, Social Science Research Network, Rochester, NY: 3904360. Accessed: Oct. 22, 2024. [Online]. Available: https://papers.ssrn.com/abstract=3904360

[29] Sakshi and V. Kukreja, “Segmentation and Contour Detection for handwritten mathematical expressions using OpenCV,” in 2022 International Conference on Decision Aid Sciences and Applications (DASA), Mar. 2022, pp. 305–310. doi: 10.1109/DASA54658.2022.9765142.

[30] H. Freeman, “On the Encoding of Arbitrary Geometric Configurations,” IRE Transactions on Electronic Computers, vol. EC-10, no. 2, pp. 260–268, Jun. 1961, doi: 10.1109/TEC.1961.5219197.

[31] J. Sun and X. Wu, “Shape Retrieval Based on the Relativity of Chain Codes,” in Multimedia Content Analysis and Mining, N. Sebe, Y. Liu, Y. Zhuang, and T. S. Huang, Eds., Berlin, Heidelberg: Springer, 2007, pp. 76–84. doi: 10.1007/978-3-540-73417-8_14.