Journal of Information Hiding and Privacy Protection
DOI:10.32604/jihpp.2020.010467
Review

A Survey on Face Anti-Spoofing Algorithms

Meigui Zhang*, Kehui Zeng and Jinwei Wang

Nanjing University of Information Science and Technology, Nanjing, 210044, China
*Corresponding Author: Meigui Zhang. Email: tinyrosy@nuist.cn
Received: 21 May 2020; Accepted: 10 June 2020

Abstract: The development of artificial intelligence makes the application of face recognition more and more extensive, which also leads to the security of face recognition technology increasingly prominent. How to design a face anti-spoofing method with high accuracy, strong generalization ability and meeting practical needs is the focus of current research. This paper introduces the research progress of face anti-spoofing algorithm, and divides the existing face anti-spoofing methods into two categories: methods based on manual feature expression and methods based on deep learning. Then, the typical algorithms included in them are classified twice, and the basic ideas, advantages and disadvantages of these algorithms are analyzed. Finally, the methods of face anti-spoofing are summarized, and the existing problems and future prospects are expounded.

Keywords: Face anti-spoofing; feature extraction; deep learning

1 Introduction

With the rapid development of Internet technology, biometrics technology has attracted more and more attention and has been widely applied in intelligent security, public security criminal investigation, financial and social security, medical education and other fields. Due to its advantages of safety, naturalness and non-contact, face recognition technology is more easily accepted by users among existing biometric recognition technology, and has become a key research direction of academia and industry. However, face recognition system is vulnerable to malicious attack by illegal users, which brings great threat to the security performance of the system. Therefore, it is very important to design a face anti-spoofing system with high detection accuracy, short time and strong robustness.

Face anti-spoofing detection refers to the process of identifying whether the currently acquired face image is from a living person or a deceptive face. In view of the important academic value of the research on face anti-spoofing, the research on it is very active at home and abroad in recent years. The number of papers related to face anti-spoofing published in important international conference journals such as CVPR (IEEE conference on Computer Vision and Pattern Recognition), ECCV (Europe conference on Computer Vision) and IEEE Transactions on Information Forensics and Security has increased substantially. Human intelligence is infinite, and there are endless ways to create deceiving faces. The most common spoofing attacks include printing attack [1–2], video replay attack [3] and 3D mask attack [4]. There are some differences between real and deceptive faces, which are mainly reflected in image texture information, motion information and depth information. Taking advantage of these differences, we can design a variety of face anti-spoofing methods to determine the true and false faces. In recent years, the research on face anti-spoofing detection has developed rapidly, and many valuable research results have been obtained. This paper will focus on the two aspects: the method based on manual feature expression and the method based on deep learning, and analyze the advantages and disadvantages of various methods as well as the development trend of face anti-spoofing in the future.

2 Methods Based on Manual Feature Expression

Nowadays, face recognition technology is more and more widely used. As an important part of face recognition, face anti-spoofing detection has also received extensive attention and has become a relatively independent research field. In view of the important value of face anti-spoofing detection, many researchers have proposed corresponding detection algorithms. According to the difference of feature extraction, the face anti-spoofing method based on manual feature expression is divided into five aspects to elaborate: the method based on image texture feature, the method based on human-machine interaction, the method based on life information, the method based on image quality and the method based on depth information.

2.1 The Methods Based on Image Texture Feature

The image loses some information during the acquisition process and is accompanied by various noises. The image acquired twice has a certain texture difference from the image acquired once. The microscopic texture difference between the deceived face image and the real face image is mainly due to the local highlight, shadow change and blur degree of the image [5]. The texture-based approach primarily uses these differences to determine true and false faces.

Some early research work was to convert the collected images into grayscale images and extract the grayscale texture information of the images for the classification of living human faces and deceptive faces. These methods ignored the color texture information. Määttä et al. [6] used multiple uniform LBP operators of different scales to extract texture feature histograms from local blocks of grayscale images and global images, and then connected them to form a 531-dimensional feature histogram and sent it to SVM classifier with RBF as the core for training and testing of living human faces and deceptive face classification. Texture analysis algorithm based on gray scale map are effective for high-resolution, texture-clear deceptive face images, but for some low-resolution deceived face images, it is difficult to accurately distinguish them. In view of this, Boulkenaf et al. [2] proposed a face anti-spoofing method based on color texture analysis. They extract LBP histograms from a single image channel and connected them to form the final descriptor. The specific process is shown in Fig. 1. In order to analyze which color space is more discriminating, this method takes into account the three color spaces RGB, YCbCr and HSV. Experiments show that the method based on color texture is superior to the method based on gray texture in detecting various attacks. Boulkenaf et al. [3] also focused on the luminance and chrominance channels, and combined the multi-stage LBP features of human face in HSV space with the LPQ features of human face in YCbCr space by using the joint information of color and texture. Although good results have been achieved in the experiment, the low level of microtexture descriptors makes them sensitive to light changes and high-quality images. In order to improve the discrimination in the further, Boulkenaf et al. [7] with accelerated the steady characteristics (speeded-up robust features, SURF) to face anti-spoofing detection. Compared with previous methods, this method showed better and stable performance.

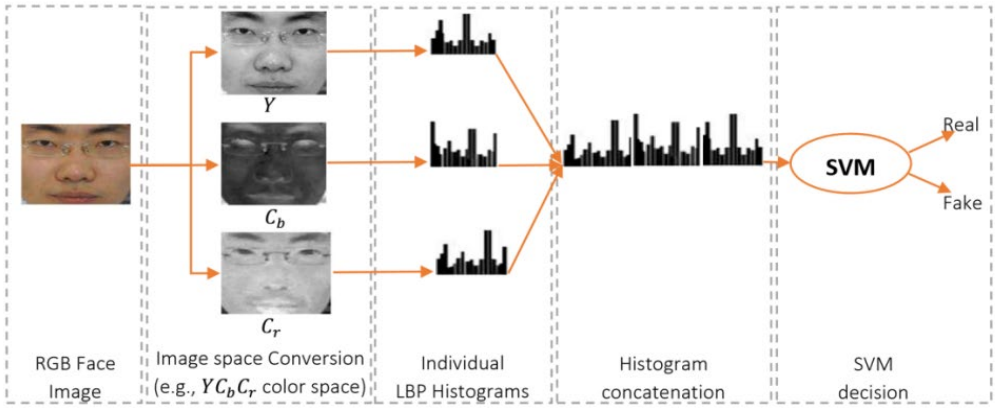

Figure 1: Face anti-spoofing method based on color texture analysis

The features extracted from the above methods based on texture features are all low-level, which will inevitably affect the robustness and generalization ability of the model. Since low-level features generally exist in high-dimensional space and are easily disturbed by noise, they are not conducive to direct classification [8]. In order to improve the expression ability of the image content, the recognition efficiency and generalization ability of the algorithm, and hope that the features within the class are more similar while the features between the classes are more differentiated, we need to express the lower-level features into higher-level features that are more differentiated through some coding algorithms. High-level features can better express the information of the whole image and facilitate classification.

Peixoto et al. [9] first used DoG filter to obtain the medium frequency band information in the image information, then extracted key features through Fourier transform, and finally differentiated and classified the extracted and processed feature information through logistic regression classifier, so as to achieve the goal of whether the recognized image is a real face or a deceptive face. Zhang et al. [10] proposed a face anti-spoofing detection scheme based on color texture Markov feature (CTMF) and support vector machine recursive feature elimination (SVM-RFE). The author analyzed the difference between adjacent pixels of real faces and deceived faces, and fully considered the texture information between color channels. First, the texture differences of real and false faces are captured by directional differential filter, which can be regarded as the low-level features of CTMF. Then Markov process is used to model the facial texture difference to form the high-level representation of the low-level features. Finally, SVM-RFE dimension reduction is used to make it suitable for real-time detection.

In general, the method based on image texture analysis has many advantages, such as low cost, simple algorithm and easy to implement. However, with the popularity of high-definition cameras and the application of high-quality 3D masks, the use of texture information alone can no longer meet the demand, so texture information often needs to be integrated with other information.

2.2 The Methods Based on Human-Computer Interaction

The host of a living face is a living human. Humans can make movements or make sounds as required, such as nodding, blinking, opening their mouths, smiling, tongues, reading a paragraph of text, etc., while deceiving faces is difficult to do. Based on this consideration, an interactive human face anti-spoofing detection method has been proposed.

The early interactive face anti-spoofing detection were designed to be fixed, which enables the pre-recorded completion of motion command video can break this kind of face anti-spoofing algorithm easily. In order to solve this problem, human-computer interaction detection based on random motion instructions comes into being. The randomness of motion instructions makes it difficult for attackers to pre-record video to attack the face anti-spoofing algorithm, which greatly improves the detection performance of the algorithm.

Wang et al. [11] conducted lip language recognition by detecting the range of changes in the region of the face's mouth, supplemented by voice recognition to obtain the voice information of the user's response to jointly judge whether the user read the randomly given statements according to the requirements. Singh et al. [12] used blinking and mouth movements to make living judgments. The area of the eyes and the HSV (hue, saturation, value) of the tooth were calculated to determine whether the eyes were open and the mouth was open. The subjects acted according to the phrase prompts randomly generated by the system, and completed the relevant action to prove that it is a real face. Ng et al. [13] designed a human-computer interaction system to guide users to complete random facial expressions. By calculating SIFT flow energy of multiple frames of images, the users could be judged whether the specified facial expressions were completed and whether they were real faces.

The human-computer interaction-based method can effectively reduce the influence of inter-class differences on the performance of the algorithm through carefully designed interaction actions. Therefore, it has a high recognition rate and a good versatility. Currently, it is widely used in practical business scenarios such as public security, medical treatment and finance. However, the face anti-spoofing detection method based on human-computer interaction needs to recognize whether the user completes the action

from the multi-frame image, and the calculation amount is large and the time required is long compared with the single frame-based algorithm. Moreover, it requires a high degree of cooperation of the subject, the detection process is cumbersome and the user experience is not good, so it violates the convenience and natural advantages of face recognition technology.

2.3 Methods Based on Life Information

An obvious difference between a living face and a deceiving face is that a living face have vital features such as heartbeat, blood flow, and micro-motion of involuntary facial muscles, and most types of deceptive faces are difficult to imitate such life features. The life information-based approach primarily uses the differences in these vital characteristics to distinguish between living faces and deceiving faces.

2.3.1 Facial Optical Flow Analysis

The concept of light flow was first proposed by Gibson in 1950. When the human eye observes the moving object, the scene of the object forms a series of continuously changing images on the retina of the human eye. This series of continuously changing information constantly "flows" through the retina (that is, the image plane), like a kind of light "flow", so it is called light flow. When an object is moving, its brightness pattern of the corresponding point on the image is also moving, we can use the optical flow to characterize the motion of the image brightness mode, as shown in Fig. 2. There are differences in the motion patterns of 3D face and 2D face. When the face rotates and swings, the living face produces different light flows due to the inconsistencies in the movements of the face. However, the movements of the photo face are basically the same, and the light flow is quite different from the living face. Based on these differences, optical flow information can be used to make judgments on true and false faces. Smiatacz et al. [14] calculated the optical flow values generated by face rotation and trained and classified these optical flow values by SVM. Bao et al. [15] used Optical Flowof Lines (OFL) to calculate the spatio-temporal difference of human face images from two dimensions, horizontal and vertical, and obtain the motion information of human face to detect such planar false face attacks as photos and video. This method is relatively simple, but sensitive to light, and has poor detection effect on video attack and 3D mask attack. Because the optical flow method of the two assumptions: (1) Brightness constant; (2) Small movements are difficult to satisfy in real life scenes, so they also have a certain impact on the detection effect.

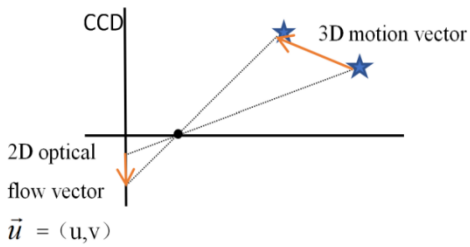

Figure 2: Optical flow method

2.3.2 Heart Rate Detection Analysis

Photoplethysmography (PPG) is a method of detecting the heart rate of a human body using photoplethysmography. These tests are generally contact-type. Non-contact heart rate measurement with a camera, commonly referred to as Remote Photoplethysmography (rPPG). Living faces are rich in capillaries, and the beating of the living heart will lead to changes in blood flow and velocity in the blood vessels, while the changes in blood flow will affect the absorption and reflection of facial light. Finally, such changes in blood will lead to changes in face color. The change of heart rate can be obtained by extracting the color change of the abundant capillary region of the face. The real face and the deceptive face have different heart rate distribution in frequency domain, using this, we can tell whether a face is real or fake.

Li et al. [3] were the first to apply rPPG to *in vivo* detection.Enter multiple frames of video, and first

extract the heart rate characteristics. If the discriminating result is living body, then LBP color and texture features should be further extracted to distinguish living body/screen attack since the distribution of human face heart rate in screen video is similar to that of living body. The specific process is shown in Fig. 3. Liu et al. [4] believed that although existing rPPG-based methods have achieved good results in the cross-database, they may not be robust enough when rPPG signals are polluted by noise. Therefore, they proposed a new feature -- rPPG correspondence feature (CFrPPG) to accurately identify heartbeat remnants from noisy rPPG signals. In order to overcome global interference, a learning strategy of introducing global noise into CFrPPG feature is proposed. The proposed feature is not only superior to the 3D mask attack method based on rPPG, but also able to deal with the actual scene of weak light and camera movement.

This kind of heart rate extraction method is mostly used in the detection of 3D mask deceiving human face. Under the condition of constant illumination, the object to be tested maintains posture and expression, this method has a high accuracy. However, their calculation process requires HD face video long enough to extract good enough rPPG signal, and the rPPG signal is easily affected by ambient light and motion of the object to be tested. The method has a general robustness, so it is often necessary to cascade other features and classifiers to realize face anti-spoofing.

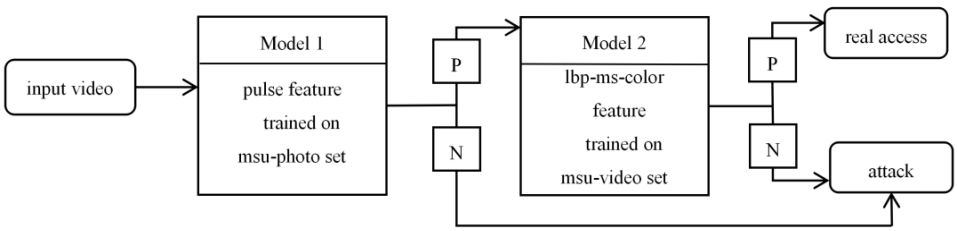

Figure 3: Cascading strategy structure

2.4 Methods Based on Image Quality

The presentation of deceiving human face needs certain media, whether it is photo paper, printing paper, electronic equipment, silica gel, plastic or other media, the material properties of which are different from the facial features and skin materials of living face. Differences in materials can lead to differences in reflection properties, such as photo paper, mobile phone display screens will have some specular reflections but live faces are basically not present. Although the manufacturing process of deceiving the face is excellent, most of the image quality after deceptive face secondary imaging is different from the living face, such as the distortion of the color distribution and the blurring of the prosthetic face image. Image quality based methods mainly use the difference between image distortion and reflection properties to distinguish true and false faces.

Galbally et al. [16] proposed an evaluation of image quality by analyzing 25 significant factors in image quality metrics. See Fig. 4 for details. Galbally et al. [17] also designed 14 general features for face anti-spoofing to extract differences in image quality. Inspired by literature [16], Wen et al. [18] proposed a face anti-spoofing deception detection algorithm based on image distortion analysis (IDA). Firstly, four different features (specular reflection, blur, color moment and color diversity) were extracted to form IDA feature vectors. Then, multiple SVM classifiers trained for different face spoofing attacks (such as photo attack and video attack) constitute an integrated classifier to distinguish real and false faces. Finally, the method is applied to video multi-frame face deception detection based on voting and good results are obtained. The quality of the image is highly dependent on the shooting equipment and the external conditions. External conditions such as low-quality shooting equipment and poor illumination can also cause distortion of the image of a living human face. Li et al. [19] considered the influence of different quality shooting equipment. First, images were clustered according to image quality dimensions by clustering method, and then an image classification guidance model based on image quality characteristics was per-trained for each quality level of images. For the test image, firstly determine its image quality level, and use the regression method to map

the image to its corresponding image quality level classification guidance model, and then use the classification guidance model to classify the living face and the deceived face.

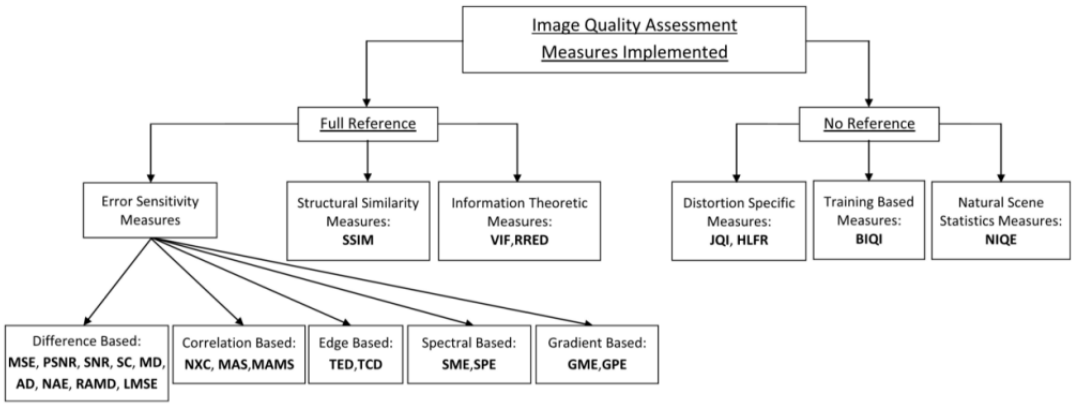

Figure 4: 25 image quality evaluation factors

The method based on image quality has low computational complexity and fast detection speed, which is beneficial to online real-time detection. But when the picture quality is high, this method is vulnerable to attack. Therefore, we need higher quality living human face and deceptive human face image as input in order to extract good enough image quality features, which requires higher requirements on face image acquisition equipment.

2.5 Methods Based on Depth Information

The real face is three-dimensional, with different depth information at different positions such as forehead, eyes, and tip of the nose, while the photo face and the video face are two-dimensional, and the depth information of different points is the same. Even if the photo is folded, it has different depth information from the real face, so the depth information can be used for face anti-spoofing.

Face anti-spoofing methods based on depth information usually require additional hardware devices. The material of the deceived face is different from the material of the skin, eyes, lips, and eyebrows of the living face, and the difference in the material causes the difference in the reflection properties. Although the deceived face looks very similar to the living face under visible light conditions, in the infrared spectrum, the appearance of the skin, eyes, nose and other areas of the living face is quite different from that of the deceived face. Some researchers used Gabor, HOG and lambert model to extract the reflection difference between living face and deceived face in near-infrared camera images for face anti-spoofing [20–22]. In the near infrared spectrum, the deceiving faces in photos and video are quite different from living faces, so this method is highly accurate, but the well-made masks are less different from living faces. To identify mask attacks, Steiner et al. [23] used short-wave infrared to distinguish face skin from mask. In addition, we can also use the depth image taken by the depth camera to record the depth information between objects for face anti-spoofing detection. Wang et al. [24] combined the depth information of Kinect camera and the texture features learned from the convolutional neural network to judge the true and false faces, and also obtained good results.

In general, the face anti-spoofing detection method based on depth information has obvious advantages: the depth information has the characteristics of illumination invariance, so the robustness of the face anti-spoofing detection is good; the real face depth map has the contour features of the three-dimensional face, and there is a significant difference between the depth map of the photo face and the video face; without excessive user interaction, it has a good detection effect on photos and video attacks, but the detection of 3D mask attack needs further research. However, it needs to add new hardware, which means new expensive hardware investment, and the new hardware will also limit the scope of the algorithm to some extent, so in some scenarios we will give up using this method.

The above mentioned methods are all based on artificial features. Although some of them can achieve a better recognition rate for face anti-spoofing, there are still some shortcomings, such as the detection effect depends on the extraction and expression of features, the need for additional hardware investment, and the algorithm's robustness and generalization capability are limited.

3 Face Anti-Spoofing Based on Deep Learning

With the continuous development and progress of deep learning and its outstanding performance in the field of face recognition, more and more researchers have applied it to face anti-spoofing to explore more effective methods to combat face deception. Different from the traditional manual feature extraction method, deep learning can automatically learn images, dig out more essential and abundant face features, and help to accurately distinguish real faces from deceptive faces.

In 2014, Yang et al. [25] first proposed the application of Convolutional Neural Network (CNN) to face anti-spoofing to extract features, which opened a new path of deep learning in the field of face anti-spoofing. As the technology was not yet mature, the detection effect was far less than that of traditional methods. However, the excellence of deep learning in feature extraction still attracted a large number of researchers to engage in face anti-spoofing based on deep learning. With the unremitting efforts and repeated attempts of many scholars, the ability of face anti-spoofing based on deep learning has been gradually improved through network updating, transfer learning, integration of multiple features, domain generalization, and has now surpassed the traditional method.

3.1 Network Updating

In deep learning, the construction of network structure plays a decisive role in algorithm performance. Menotti et al. [26] proposed a unified network framework for iris, face and fingerprint spoofing detection. The model learns representations directly from data through two optimizations, Architecture optimization (AO) and filter optimization (FO), which randomly search for the best convolutional neural network from a series of networks defined in the hyper-parametric search space. Then, linear support vector machine classifier is used to make the final decision. Learning a new network structure can maximize the generalization ability of the algorithm. However, at present, it is difficult to update the network structure in training to find the optimal algorithm for face anti-spoofing, and research results are still rather limited, so more researchers are needed to fill the gap.

3.2 Transfer Learning

Using deep learning to detect face authenticity often requires a large amount of training data to obtain more distinctive features. However, there is not enough data in the existing face anti-fraud database, and the neural network used by most methods only consists of several layers. It is difficult to train a large network classifier with high performance. When there is not enough data to train from scratch, transfer learning [27] can avoid over-adapting to large networks and save a lot of computing resources.

Oeslle et al. [28] constructed a network framework named FASNet and used pre-trained Convolutional Neural Network (CNN) to detect face spoofing. As shown in Fig. 5, FASNet fixed the previous network structure on the basis of VGG16 [29] and modified the last three layers of network to achieve transfer learning. For CNN, there are two methods for transfer learning. It is simpler to use the source model as a "ready-made" feature extractor, using the output of the selected layer as the input to the target model, which is the only model trained for the new task. A more complex approach is to achieve full or partial "fine-tuning" of the source model by retraining weights through back propagation.

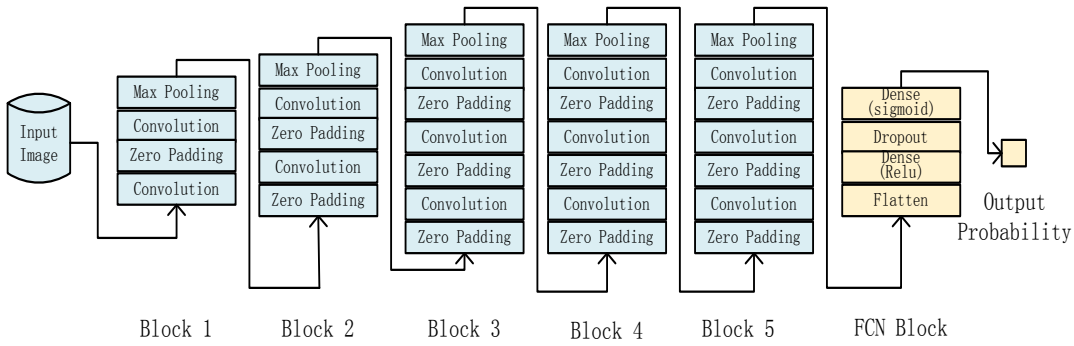

Figure 5: The architecture of FASNet

Tu et al. [30] proposed a fully data-driven hyper-depth model based on transfer learning. This model uses pre-trained deep residual network (ResNet-50) [31] to extract the spatial features of sequence frames, and then inputs the spatial features into the long and short term memory (LSTM) unit to obtain the temporal features that can be used for final classification, and finally judges the faces true or false.

Using transfer learning to train faces can solve the problem of large network over-adaptation caused by limited data set. While extracting the key features that can distinguish spoofing face from real face, it can reduce overfitting, obtain excellent detection effect and save the calculation cost. Just like network updating, the study on transfer learning in face anti-spoofing is not in-depth enough, and the effect has not reached the ideal state. The connection between the pre-training network and the real training network needs to be optimized.

3.3 Feature Integration

A feature is only sensitive to the change of the corresponding feature of the image. When there is little difference in this feature between real face and spoofing face, it is difficult for the classifier to distinguish them based on single feature training. Therefore, it is far from enough to focus on optimizing the network structure to extract more representative single features. By extracting multiple features from face images for integration, the difference between real face and spoofing face can be better highlighted, the robustness and generalization ability of the algorithm can be improved, and the accuracy of the test can be greatly improved.

3.3.1 Texture and Depth Information

The depth information of the images is an important basis for judging the authenticity of the face. Because the real face is three-dimensional, while the face attacked by photos and screens is flat. Even if the face is distorted, the depth map is still different from the real face. Atoum et al. [32] first took face depth map as the key information for discriminating face spoofing. In this paper, a two-channel CNN based face anti-spoofing method was proposed to integrate local features of face images with depth information. The first CNN extracts several local face blocks as training data, assigns scores to each block to represent the likelihood that the face is real, and calculates the whole face image with the average value. The second CNN adopts full Convolutional Neural Network to estimate the depth map of face images by the classification of pixel points, and provide an authenticity score according to the estimated depth map of face images. Finally, the scores of these two CNNS were integrated to judge the authenticity of the face. The algorithm flow chart is shown in Fig. 6. While this approach does its best to integrate features, it has yet to outperform traditional methods.

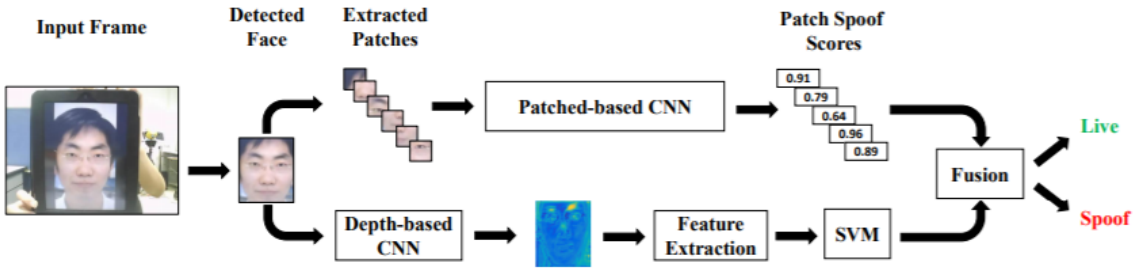

Figure 6: The flow chart of dual-channel CNN

3.3.2 Spatio-Temporal Information

Face images contain a large number of spatial features such as texture and depth, but temporal features also play a crucial role in face anti-spoofing. The analysis of human face from the perspective of time and space can dig out more effective information and improve the detection accuracy.

Liu et al. [33] used the integration of face depth information and rPPG signal to carry out face anti-spoofing, and pointed out that the binary classification problem was replaced by the targeted feature supervision problem. The depth of the face represents spatial information and the rPPG signal represents time, which can highlight the key differences between real and spoofing faces. From a spatial perspective, a real face is three-dimensional while a photo or screen face is two-dimensional; From a time perspective, real faces can detect normal rPPG signals but not spoofing faces. In order to achieve the two kinds of supervision, the author designed a deep learning method based on CNN - RNN structures. The CNN uses the depth image supervision to identify the subtle texture features, and then inputs the estimated depth and feature map into a new non-rigid registration layer and creates a new feature map, while the RNN uses the previously generated new feature map and rPPG for training. Finally, the depth information and heart rate statistics obtained by serial monitoring of rPPG signals are fused. Based on this, the real and spoofing faces were distinguished. The architecture is shown in Fig. 7. The experiment shows that this method has achieved ideal test results, which finally surpasses the traditional test method and it also reflects the importance of auxiliary supervision.

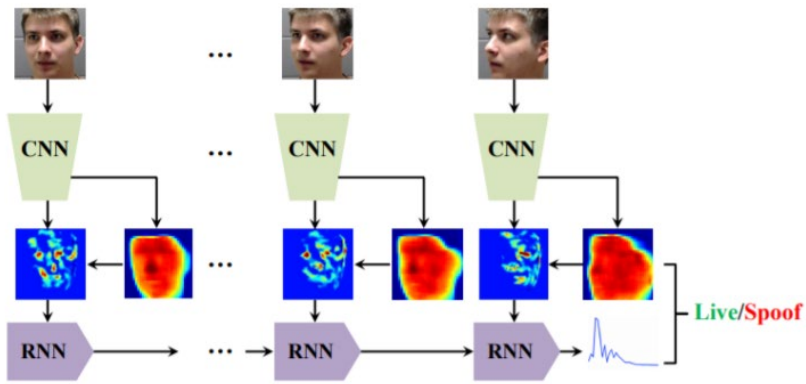

Figure 7: Face anti-spoofing based on CNN-RNN

However, the above method also has the following problems: (1) In the non-rigid registration layer, facial expressions and posture changes are removed, ignoring their differences between real and deceptive faces; (2) It is not convincing to use single frame image to predict depth. Relatively, reconstructing depth map through spatial micro-changes between multiple frames of images is more beneficial. Based on these two shortcomings, Wang et al. [34] constructed a depth supervision framework and used multiple RGB

frames as input to estimate the face depth, so as to make full use of spatio-temporal information to analyze the effect of motion and depth in the detection of presentation attack. It consists of two novel modules: The Optical Flow guided Feature Block (OFFB) and the ConvGRU module, which are designed to extract short-and long-term motions to distinguish faces from spoofing ones. This method can detect spoofing face efficiently and accurately with deep supervision.

Yang et al. [35] developed a novel Spatio-temporal Anti-spoofing Network (STASN), which takes into account global temporal and local spatial information to distinguish real faces from spoofing faces. The model consists of three parts: TASM, RAM and SASM. TASM is a CNN-LSTM, which takes frame sequence as input, first extracts the features of CNN, carries out LSTM propagation, and then predicts the result of binary classification. RAM learns the offset based on CNN features from TASM and outputs the participating regions related to sequence images. SASM inputs the participating region of RAM output into parameter sharing CNN, and finally integrates it for prediction, as shown in Fig. 8. The proposed model can automatically focus on the recognition area, which makes it possible to analyze the network behavior. By extracting features from different regions to find subtle evidence, such as edges, moire patterns and reflected artifacts, the model can effectively distinguish real and spoofing faces. At the same time, the authors say that for face anti-spoofing, not only to build a good network, data is also very important. Therefore, they propose a data collection solution and data synthesis technology to simulate digital media-based face spoofing attacks, which can help obtain a large number of training data reflecting real scenes.

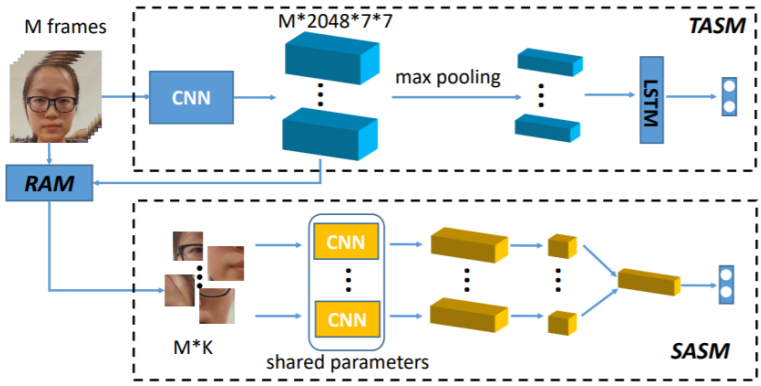

Figure 8: The constuction of STASN

The above articles show that integrating two or more features, especially spatio-temporal features, as a basis for judging whether a face is real or spoofing can highlight the difference between the two faces more comprehensively and effectively. Compared with face anti-spoofing based on single feature, multi-feature integration is superior in accuracy, and also improves the robustness and generalization of the algorithm.

3.4 Domain Generalization

With the extensive application of deep learning in face anti-spoofing, more and more methods have been proposed. However, these methods are usually limited to the detection of known spoofing attacks, and there is blindness to the unknown spoofing. In order to improve the generalization ability of detection methods under “invisible” attack, the following methods were developed.

Tu et al. [36] proposed a general face authentication Convolutional Neural Network (GFA-CNN). As shown in Fig. 9, the network proposed to use TPC (Total Pairwise Confusion) loss to balance the contributions of various attack modes to increase CNN's generalization ability for attack types. In addition, the Fast Domain Adaptation (FDA) will be integrated into CNN, and all input images will be converted into the target domain background to mitigate the impact of domain transfer. Finally, the training and testing of the network are realized by means of multi-task learning. In multi-task CNN, spoofing detection and face recognition share the same network and parameters. The overall objective loss function is:

$$L = Lanti + \lambda 1 Ltpc + \lambda 2 Lid \quad (1)$$

L_{anti} is the loss of face deception, L_{id} is the losecs of face rognition, λ_1 and λ_2 is the parameter.

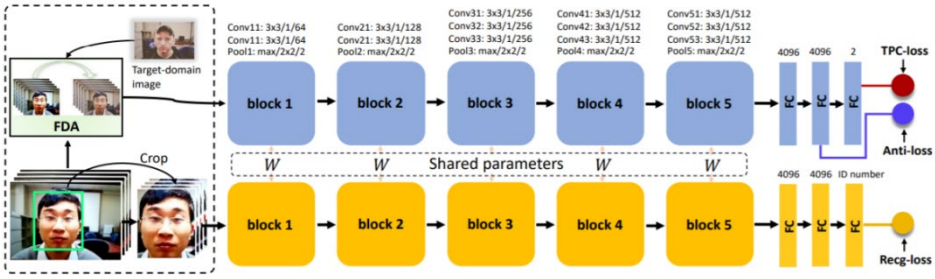

Figure 9: The structure of GFA-CNN

Liu et al. [37] defined the detection of unknown spoofing attack as Zero-shot Facial Anti-Spoofing(ZSFA), and proposed a novel Deep Tree Network (DTN), which was used to train trees in an unsupervised way to find the feature library with the greatest variation, so as to classify the spoofing samples into semantic subgroups. Whether faced with a known attack or unknown, DTN routes data samples to the most similar leaf nodes to produce real-time and spoofing binary decisions. In addition, in order to better study ZSFA, the author also created the first face anti-spoofing database SiW-M containing various types of deception.

Shao et al. [38] improved the generalization ability of face anti-spoofing method through a multi-adversarial discriminative deep domain generalization framework. In this framework, the generator that trains and generates domain shared features competes with multiple domain discriminators. Under the constraint of double-force triple mining, the shared and differentiated feature space is gradually learned, and the learned features are difficult to distinguish between multiple domain discriminators, as shown in Fig.10. Thus, when the feature generator successfully spoofs all domain descriptors, the feature space shared by all source domains can be automatically discovered. The model in this generalized feature space can extract the more generalized clues shared by all the source domains, and the prediction model based on the training data of the visible source domain can well deal with various invisible face presentation attacks. In order to further improve the generalization ability, face depth map is introduced into the network framework as an auxiliary supervision.

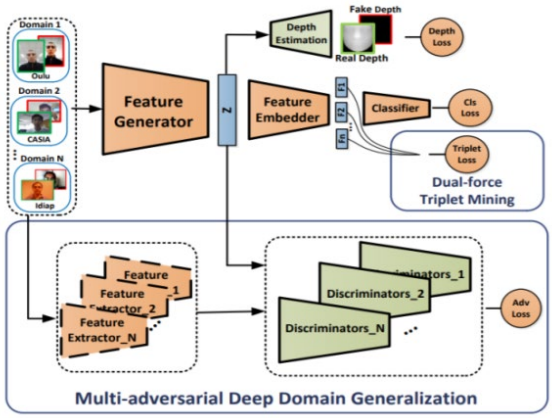

Figure 10: The architecture of multi-adversarial discriminative deep domain generalization

Daniel et al. [39] redefined generalized representation attack detection (GPAD) from the perspective of anomaly detection, and proposed a new face spoofing detection strategy based on depth-metric learning. This model uses a set of Siamese CNN's combined with Triplet Focal Loss as a regularization method to

implement a new "Metric-Softmax" Loss model, as shown in formula (2) and (3). The architecture is illustrated in Fig. 11.

$$Lanomaly = Lmetric\_soft + \lambda Ltf \tag{2}$$

$$Lmetric\_soft = -\sum_{i=1}^{b} log \frac{e^{Da_i,p_i}}{e^{Da_i,p_i} + e^{Da_i,n_i}} \tag{3}$$

λ are a hyperparameter that controls the balance between Triplet Focal loss and softmax loss. The loss model accumulates the highly separated probability distribution of each pair of triads in Euclidean space to ensure that the learned more robust and generalized feature representation is sufficient to distinguish real and spoofing faces and avoid generalization. Finally, a post-probability estimation is introduced to determine whether the image is a real sample or a simulation attempt, avoiding training any classifier for decision making and it shows the excellence of deep anomaly detection structure. An in-depth experimental evaluation of the challenging GRAD-GPAD shows that the approach based on anomaly detection is superior to the most advanced models.

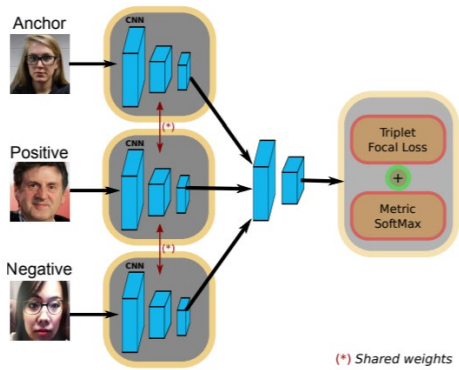

Figure 11: Face Anti-Spoofing based on anomaly detection

The experiments show that the networks can adapt to the cross detection of more unknown databases to some extent, weaken the over-fitting, and realize the effective detection of "invisible" spoofing attacks through the domain generalization under the training of limited attack types. However, there are problems. Domain generalization will make the training and retesting effect on a single data set unable to reach the optimal state, which requires the network framework to be improved.

4 Summary and Prospect

With the extensive application of artificial intelligence in real life, face recognition has become an important means to realize security. In order to avoid malicious attack, face anti-spoofing has become an urgent problem. From the beginning of manual feature extraction method based on image texture, human-computer interaction, life information, image quality and depth information, and then to using deep learning to automaticly extract feature, combined with network updating, transfer learning, feature integration and domain generalization, the study of face spoofing detection has been constantly updated and improved, and the efficiency and accuracy of detection have now reached a considerable state.

However, how to get rid of the influence of database size on detection accuracy, and how to ensure the accurate judgment of known spoofing while improving the generalization ability of detection of unknown spoofing are important problems that cannot be ignored in the follow-up work. At the same time,both network updating and transfer learning are good ideas to improve the performance of face anti-spoofing algorithms, which are worthy of further research by more researchers.

Funding Statement: This work is supported by National Natural Science Foundation of China(62072250).

Conflicts of Interest: We declare that we do not have any commercial or associative interest that represents a conflict of interest in connection with the work submitted.

References

[1] Z. Boulkenafet, J. Komulainen and A. Hadid, "Face antispoofing based on color texture analysis," in *Proc. of IEEE Int. Conf. on Image Processing,* Piscataway, NJ: IEEE Press, pp. 2636–2640, 2015.

[2] Z. Boulkenafet, J. Komulainen and A. Hadid, "Face spoofing detection using colour texture analysis," *IEEE Trans on Information Forensics and Security*, vol. 11, no. 8, pp. 1818–1830, 2016.

[3] Xiaobai Li, J. Komulainen and Guoying Zhao, "Generalized face anti-spoofing by detecting pulse from face videos," in *Proc. of IEEE23rd Int. Conf. on Pattern Recognition,* Piscataway, NJ: IEEE Press, pp. 4239–4244, 2016.

[4] S. Q. Liu, X. Y. Lan and P. C. Yuen, "Remote photoplethysmography correspondence feature for 3D mask face presentation attack detection," in *Proc. of the European Conf. on Computer Vision*, Cham: Springer, pp. 558–573, 2018.

[5] J. K. Huang, "Research on living detection technology of face recognition," Wuhan: Central China Normal University, 2018.

[6] J. Määttä, A. Hadid and M. Pietikäinen, "Face spoofifing detection from single images using micro-texture analysis," in *Pro. of Int. Joint Conf. on Biometrics*, Colorado State, USA: IEEE, pp. 1–7, 2011.

[7] Z. Boulkenafet, J. Komulainen and A. Hadid, "Face antispoofifing using speeded-up robust features and fifisher vector encoding," *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 141–145, 2017.

[8] C. Y. Xiang, "Research on feature extraction and classification method of RGB-D images," Xiangtan: Xiangtan University, 2017.

[9] B. Peixoto, C. Michelassi and A. Rocha, "Face liveness detection under bad illumination conditions," in *Proc. of 18th IEEE Int. Conf. on Image Processing*, Melbourne, Australia: IEEE, pp. 3557–3560, 2011.

[10] L. B. Zhang, F. Peng and L. Qin, "Face spoofing detection based on color texture markov feature and support vector machine recursive feature elimination," *Journal of Visual Communication and Image Representation*, no. 51, pp. 56–69, 2018.

[11] R. J. Wang, J. L. Li, H. Ni, Y. J. Wu and F. Y. Huang, "A face recognition method and system," *China*, 2015.

[12] A. K. Singh, P. Joshi and G. C. Nandi, "Face recognition with liveness detection using eye and mouth movement," in *Proc. of Int. Conf. on Signal Propagation and Computer Technology*, Piscataway, NJ: IEEE Press, pp. 592–597, 2014.

[12] E. S. Ng and Y. S. Chia, "Face verification using temporal affective cues," in *Proc. of the 21st Int. Conf. on Pattern Recognition*, Tsukuba Science City, Japan: IEEE, pp. 1249–1252, 2012.

[13] M. Smiatacz, "Liveness measurements using optical flow for biometric person authentication," *Metrology and Measurement Systems*, vol. 19, no. 2, pp. 257–268, 2012.

[14] W. Bao, H. Li, N. Li and W. Jiang, "A liveness detection method for face recognition based on optical flow Fifield," in *Proc. of Int. Conf. on Image Analysis and Signal Processing*, Cairo, Egyptpages: IEEE, pp. 233–236, 2009.

[15] 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 ,2014.

[16] J. Galbally and S. Marcel, "Face anti-spoofifing based on general image quality assessment," in *Proc. of 22nd Int. Conf. on Pattern Recognition*, Stockholm, Sweden: IEEE, pp. 1173–1178, 2014.

[17] D. Wen, H. Han and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 746–761, 2015.

[18] H. L. Li, S. Q. Wang and A. C. Kot, "Face spoofifing detection with image quality regression," in *Proc. of 6th Int. Conf. on Image Processing Theory Tools and Applications*, Oulu, Finland: IEEE, pp. 1–6, 2016.

[19] D. Yi, Z. Lei, Z. W. Zhang and Li. S. Z, "Face anti-spoofifing: Multispectral approach," *Handbook of Biometric Anti-Spoofifing*, Berlin: Springer, 2014.

[20] X. D. Sun, L. Huang and C. P. Liu, "Context based face spoofifing detection using active near-infrared image," in *Proc. of 23rd Int. Conf. on Pattern Recognition*, Cancun, Mexico: IEEE, pp. 4262–4267, 2016.

[21] X. D. Sun, L. Huang and C. P. Liu, "Multispectral face spoofifing detection using VIS–NIR imaging correlation," *International Journal of Wavelets, Multiresolution and Information Processing*, vol. 16, no. 2, 1840003, 2018.

[22] H. Steiner, A. Kolb and N. Jung, "Reliable face anti-spoofifing using multispectral swir imaging," in *Proc. of International Conf. on Biometrics*, Halmstad, Sweden: IEEE, pp. 1–8, 2016.

[23] T. Wang, J. W. Yang and Z. Lei, "Face liveness detection using 3D structure recovered from a single camera," in *Proc. of Int. Conf. on Biometrics*, Piscataway, NJ: IEEE Press, pp. 1–6, 2013.

[24] J. W. Yang, Z. Lei and S. Z. Li, "Learn convolutional neural network for face anti-spoofing," *arXiv preprint arXiv:1408. 5601*, 2014.

[25] D. Menotti, G. Chiachia and A. Pinto, "Deep representations for iris, face, and fingerprint spoofing detection," *IEEE Transactions on Information Forensics and Sec*urity, vol. 10, no. 4, pp. 864–879, 2015.

[26] J. Yosinski, J. Clune and Y. Bengio, "How transferable are features in deep neural networks?" *Advances in Neural Information Processing Systems*, pp. 3320–3328, 2014.

[27] O. Lucena, A. Junior and V. Moia, "Transfer learning using convolutional neural networks for face anti-spoofing," in *Int. Conf. Image Analysis and Recognition*, Springer, Cham, pp. 27–34, 2017.

[28] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[29] X. Tu and Y. Fang, "Ultra-deep neural network for face anti-spoofing," in *Int. Conf. on Neural Information Processing*, Springer, Cham, pp. 686–695, 2017.

[30] K. He, X. Zhang and S. Ren, "Deep residual learning for image recognition," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 770–778, 2016.

[31] Y. Atoum, Y. Liu and A. Jourabloo, "Face anti-spoofing using patch and depth-based CNNs," in *2017 IEEE Int. Joint Conf. on Biometrics*, IEEE, pp. 319–328, 2017.

[32] Y. Liu, A. Jourabloo and X. Liu, "Learning deep models for face anti-spoofing: Binary or auxiliary supervision," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 389–398, 2018.

[33] Z. Wang, C. Zhao and Y. Qin, "Exploiting temporal and depth information for multi-frame face anti-spoofing," *arXiv preprint arXiv:1811.05118*, 2018.

[34] X. Yang, W. Luo and L. Bao, "Face anti-spoofing: Model matters, so does data," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 3507–3516, 2019.

[35] X. Tu, J. Zhao and M. Xie, "Learning generalizable and identity-discriminative representations for face anti-spoofing," *arXiv preprint arXiv:1901.05602*, 2019.

[36] Y. Liu, J. Stehouwer and A. Jourabloo, "Deep tree learning for zero-shot face anti-spoofing," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 4680–4689, 2019.

[37] R. Shao, X. Lan and J. Li, "Multi-Adversarial Discriminative Deep Domain Generalization for Face Presentation Attack Detection," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition*, pp. 10023–10031, 2019.

[38] D. Pérez-Cabo, D. Jiménez-Cabello and A. Costa-Pazo, "Deep Anomaly Detection for Generalized Face Anti-Spoofing," in *Proc. of the IEEE Conf. on Computer Vision and Pattern Recognition Workshops*, 2019.