

Article

Face Anti-Spoofing Method Based on Residual Network with Channel Attention Mechanism

Yueping Kong ^{1,*}, Xinyuan Li ¹, Guangye Hao ² and Chu Liu ¹

¹ School of Information and Control Engineering, Xi'an University of Architecture and Technology, Xi'an 710055, China

² Informatization Technology Office, Shaanxi Provincial Public Security Department, Xi'an 710018, China

* Correspondence: annie_kyp@sina.com; Tel.: +86-82205599

Abstract: The face recognition system is vulnerable to spoofing attacks by photos or videos of a valid user face. However, edge degradation and texture blurring occur when non-living face images are used to attack the face recognition system. With this in mind, a novel face anti-spoofing method combines the residual network and the channel attention mechanism. In our method, the residual network extracts the texture differences of features between face images. In contrast, the attention mechanism focuses on the differences of shadow and edge features located on nasal and cheek areas between living and non-living face images. It can assign weights to different filter features of the face image and enhance the ability of network extraction and expression of different key features in the nasal and cheek regions, improving detection accuracy. The experiments were performed on the public face anti-spoofing datasets of Replay-Attack and CASIA-FASD. We found the best value of the parameter r suitable for face anti-spoofing research is 16, and the accuracy of the method is 99.98% and 97.75%, respectively. Furthermore, to enhance the robustness of the method to illumination changes, the experiment was also performed on the datasets with light changes and achieved a good result.



Citation: Kong, Y.; Li, X.; Hao, G.; Liu, C. Face Anti-Spoofing Method Based on Residual Network with Channel Attention Mechanism.

Electronics **2022**, *11*, 3056. <https://doi.org/10.3390/electronics11193056>

Academic Editors: Antonio Fernández-Caballero and Byung-Gyu Kim

Received: 18 August 2022

Accepted: 23 September 2022

Published: 25 September 2022

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2022 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: face anti-spoofing; secondary imaging; residual network; attention mechanism

1. Introduction

Facial recognition technology is widely used in our daily life, such as in access control systems, turnstiles, and financial payments. However, the face recognition system is vulnerable to spoofing attacks by photos or videos of a valid user's face, which leads to security threats or property losses. Therefore, researching the method of face anti-spoofing and adding a front-end "safety lock" to the face recognition system has been the research hotspot in recent years. This method aims to detect the different cues existing in live and non-live face images to form a criterion. Compared with the living face images imaged directly, the non-living face images in photos and videos are re-imaged. There is a certain loss of details and texture differences, mainly reflected in image blur, shadow changes, and local highlights [1]. In view of these clues, some researchers have used Local Binary Patterns (LBP) to analyze image texture information and use Support Vector Machines (SVM) to classify living and non-living objects [2,3]. There are also three-dimensional orthogonal plane local binary patterns LBP-TOP [4], Histogram of Oriented Gradient (HOG) [5], and other features that analyze the texture difference of face images. However, the degree of texture features of the face anti-spoofing method is not effective for low-quality non-living face images. Ref. [6] proposes a detection method based on color texture, and its experiments show that LBP features extracted from color space are better than grayscale features in face anti-spoofing. These artificially designed discriminative features perform significantly within a dataset. However, when detecting across datasets, the detection performance of these methods decreases due to lighting differences in the imaging environment. Ref. [7] introduced the Convolutional Neural Network (CNN) into

the field of face anti-spoofing, and ref. [8] proposed a transfer learning method using CNN for face anti-spoofing, designed the Face Anti-Spoofing Network (FASNet) by improving the top of the Visual Geometry Group (VGG) network framework, and achieved good detection results. Ref. [9] migrated the MobileNetV2 model to extract features from three images: RGB, HSV, and LBP for fusion, and achieved a low error rate on the SiW dataset. Some scholars improved the performance of the method through feature fusion. Ref. [10] integrated the depth information on the depth map, the dynamic information on the optical flow map, and the secondary imaging noise information on the residual noise map, and this achieved good results on datasets. Ref. [11] claimed that conventional convolution struggles to describe the intrinsic details of images and proposed a central difference convolution network to capture the intrinsic information of non-living face images in RGB, depth, and near-infrared modalities effectively. Ref. [12] proposed a model for face anti-spoofing based on CNNs and brightness equalization.

The above methods focus on the global feature descriptions of the face image. They do not consider the nose shadow mutation region and the cheek texture region that are fairly different between living and non-living faces. When the number of network layers of the CNN or FASNet model reaches a certain level, there will be a problem of gradient disappearance, which may cause the performance of the model to degrade. Ref. [13] proposed the channel attention mechanism named Squeeze and Excitation (SE). It adaptively learns the importance of each feature, reassigns weights to different features, highlights the importance of specific task features, and suppresses the features that are useless or harmful to the task. Therefore, the feature expression ability of convolutional neural networks is improved. Ref. [14] proposed an end-to-end bone age assessment (BAA) model based on lossless image compression and a squeeze-and-excitation deep residual network. Ref. [15] used the ResNet model as a base model for ECG signal classification and tried to improve performance through SE-ResNet. Ref. [16] established a ResNet with an attention layer CBAM, which achieved a signal recognition rate of more than 90% for ten different modulation types in a complex electromagnetic environment. The network performance was superior to other deep learning methods. This paper used a deep residual network as the backbone network, an embedded attention mechanism between the residual last layer and the residual skip connection in each convolution block, and it is designed to improve the residual convolution module. The attention mechanism can reweight the distribution of different features of face images and enhance the ability of the network to express key features with significant changes in details and textures in the nose and cheek regions. Experiments show that the detection accuracy of the proposed network method is higher than that of the neural network.

The remainder of this paper is organized as follows. Section 2 describes the construction of the face anti-spoofing model. Section 3 discusses the experimental results. Section 4 provides the conclusions.

2. Face Anti-Spoofing Model with Channel Attention Mechanism

Criminals use the face photos or videos of valid users to conduct identity authentication attacks, all of which need to be re-shot or resampled to form a non-living face. There is a process of re-imaging or resampling to form a non-living face. There is texture smoothing and a loss of details in the cheek texture region and the nose shadow mutation region of the face image, which is quite different from the direct imaging of living faces. This paper used this fact as the entry point. Combined with the attention mechanism, the feature extraction module of the deep residual network is modified so that the network pays more attention to the nose and cheek areas with large differences in shadows and textures in the face image.

2.1. Deep Residual Network

Learning the different features between live and non-living face images by deep neural network is important for recognition and classification. Scholars generally develop more

different feature matrices to improve the accuracy of network recognition by deepening the number of network layers. When the number of layers increases, the network performance does not necessarily improve; rather, the network performance decreases as the number of layers increases. In response to this problem, ref. [17] proposed a deep residual network and applied it to the image recognition task. The deep residual network contains multiple residual blocks with skip connections, and its structure is shown in Figure 1. Due to adding a direct connection of information between the input image x and the output feature $F(x)$, the residual block can provide the next network unit with the critical information lost in the convolution operation so that when the deep network degenerates, the shallow features are transmitted to the deep layer to ensure that the network performance is not lost to the shallow layer. Through the stacking of multiple residual blocks, a deep network structure is formed so that the residual network can effectively alleviate the gradient disappearance and performance degradation problems of the deep convolutional network.

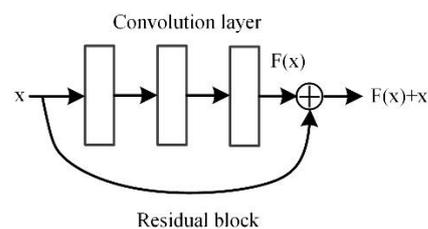


Figure 1. Residual unit.

In the conventional convolution operation, the spatial information and channel information of width and height are fused for superposition calculation. By extracting features globally and ignoring the different relationships between the channels, the features that are useless for the task will also be extracted and become redundant information classification criteria. When the face feature map is extracted by convolution, the example of the feature map is obtained, such as in Figure 2. The figure is a partial feature map of the output of the fourth activation layer in ResNet50. The reason for choosing this layer is that the unimportant features are discarded after the activation function ReLU to highlight the useful features, and at the same time, this layer is located in the superposition of shallow residual blocks. The shallow detail features are preserved. From the Figure 2b–d, we can see that when the input face image features are directly extracted by the residual block, the importance of different filter features is not distinguished. Features mainly focus on the edges of the face contour, nose, eyebrows, eyes, and other sudden color changes after multiple residual convolutions. However, the ceiling lines in the overhead background are also obtained as edge features. Although the skip connection operation in the residual network alleviates the gradient disappearance problem caused by the deepening of the network layer, it also transfers the useless feature information of the background area from the shallow network to the deep layer, which will not be discarded in the process of network optimization. Eventually, it becomes useless information in the living detection criterion.

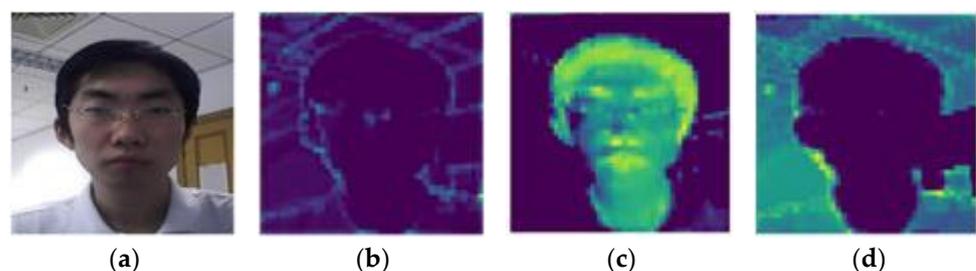


Figure 2. Face image feature map of the fourth ResNet50 convolution layer output. (a) Original image; (b–d) face feature map.

2.2. Channel Attention Mechanism

In order to make the detection criteria of the network model focus on the nose and cheek regions of the face with more differential feature information, the channel attention mechanism SE module can adaptively adjust the weight distribution of features, so we improved the ResNet50 network by embedding the channel attention mechanism SE module. The SE module can enhance the network’s ability to express different key features in the nose and cheek regions of the face, suppressing the expression ability of useless features in the background region of the face image. Thus, we used the SE module to measure the importance of each filter feature of the acquired face image.

Figure 3 shows the specific structure of the SE module. Its functions can divide into three parts: Squeeze, Excitation, and Reweight. We assume that the SE module receives the face filtermatrix U extracted by the residual network convolution layer. The Squeeze function performs a global average pooling operation on the feature matrix U , using Equation (1) to compress the spatial dimension of U from $H \times W \times C$ into $1 \times 1 \times C$ feature vector z_c , z_c expresses its global feature distribution, which is used to enhance key feature channels and suppress non-key feature channels.

$$z_c = \frac{1}{H \times W} \sum_{i=1}^H \sum_{j=1}^W U(i, j), \tag{1}$$

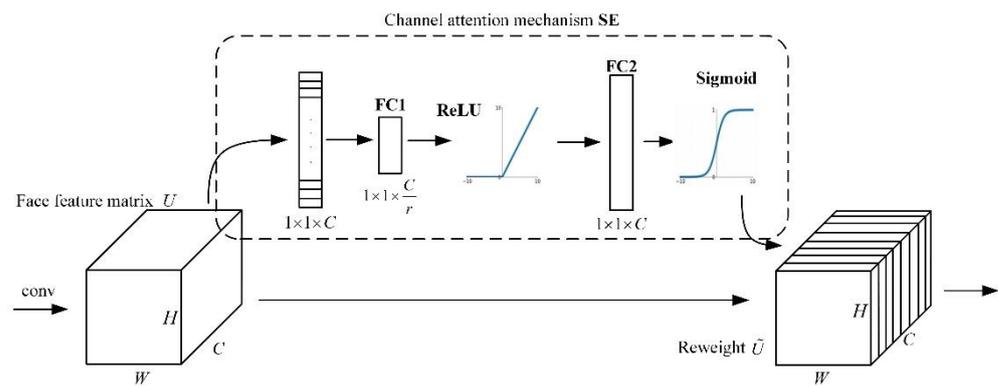


Figure 3. Channel attention mechanism SE module.

The Excitation function learns the weight of each feature channel to learn the correlation between the feature channels. The excitation reduces and increases the dimension by setting two fully connected layers. Finally, it uses the ReLU and Sigmoid activation functions to excite the global features, learning the nonlinear relationship between features and assigning the weight of each channel feature s_c , which can be expressed in Equation (2):

$$s_c = \sigma(w_2 F_{ReLU}(w_1 z_c)), \tag{2}$$

where σ is the Sigmoid function, and w_1 and w_2 are the weight of the global feature. s_c is the considered importance of each feature after feature weight assignment.

The Reweight function uses Equation (3) to multiply the channel weights, s_c , and the original face feature matrix σ to obtain the weights of adjusted features \tilde{U} , thereby enhancing the features of the nose and cheek regions of the face and reducing the importance of the background region features.

$$\tilde{U} = s_c \cdot U \tag{3}$$

2.3. Deep Residual Network with Channel Attention Mechanism

A residual network alleviates the problem of gradient disappearance by adding skip connections, but it also retains some useless feature information. This paper’s research uses an embedding channel attention mechanism to enhance the expressive ability of the facial nose and cheek regions features and suppress the expressive ability of background region

features. A residual network with an attention mechanism, referred to as SE-ResNet50, is established. Considering the parameters of the model and expecting the shallow layer of the network to reflect the details of the face, we embedded the SE module in the first residual block of each convolution block. The SE module is embedded between the residual last convolution layer and the residual skip connection in each convolution block, and this block is called SE-Res. Figure 4 shows the structure of the SE-Res and SE-ResNet50.

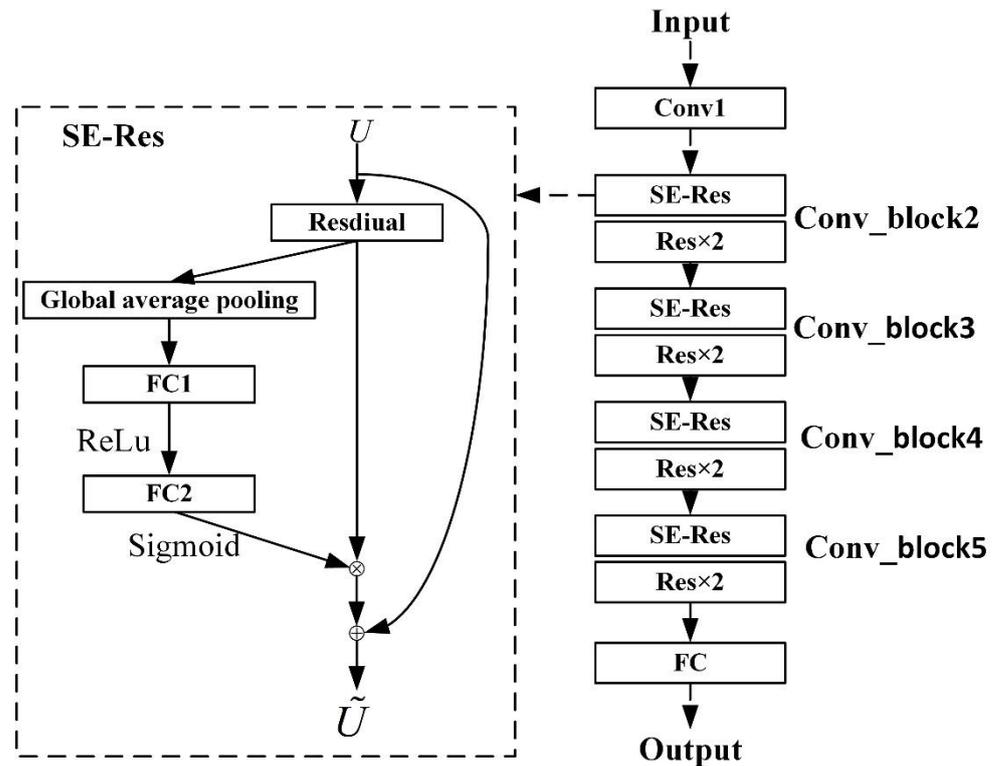


Figure 4. The structure of SE-ResNet50.

The SE-ResNet50 network consists of the first convolution layer, four convolution blocks, and a fully connected layer. Each convolution block consists of a SE-Res block and several residual blocks. For example, Conv_block2 consists of a SE-Res block and two residual blocks, and each residual block consists of a tree convolution layer. The first contains 64 convolution kernels with a scale of 1×1 , and the second convolution layer contains 64 convolution kernels with a scale of 3×3 . The third convolution layer contains 256 convolution kernels with a scale of 3×3 . Finally, to apply SE-ResNet50 to the face anti-spoofing task, the number of neurons in the final fully connected layer is adjusted to 2 to correspond to live faces and non-living faces. After the fully connected layer, this paper used softmax to calculate the predicted value of the final feature vector z prediction value of each class j , $y_j \in [0, 1]$, which can be expressed in Equation (4).

$$y_j(z) = \frac{e^{z_j}}{\sum_{j=1}^2 e^{z_j}} \tag{4}$$

Considering the probability value that y_j is between 0 and 1, thus, the loss function selected the cross-entropy error function, which can be expressed in Equation (5):

$$loss = -\frac{1}{N} \sum_i [l_i \cdot \log(y_i) + (1 - l_i) \cdot \log(1 - y_i)] \tag{5}$$

where N is the total sample, l_i is the actual label of sample i , and y_i is the predicted probability value of sample i .

3. Experiments and Analysis

3.1. Dataset

To verify the performance of the SE-ResNet50 model, experimental tests were performed on the public face anti-spoofing datasets Replay-Attack and CASIA-FASD, and this paper builds a data set of illumination changes to enhance the robustness of the model to changes in lighting. The Replay-Attack dataset contains 1200 videos of 50 people, including 200 living faces and 1000 non-living faces. The attack methods of non-living faces include handheld videos, printed photos, and electronic photos. The CASIA-FASD dataset contains 600 videos of 50 people, including 120 live faces and 480 non-live faces. The attack methods of non-live faces include handheld videos, curved photos, and cut photos. Since the influence of ambient lighting is not considered in the public dataset, this paper built a small dataset. The dataset consists of a total of 15 live face videos with a pixel size of 640×480 , and then the live face videos are played on a tablet computer for sub-sampling to obtain their non-living face data. Some data samples are shown in Figures 5 and 6.

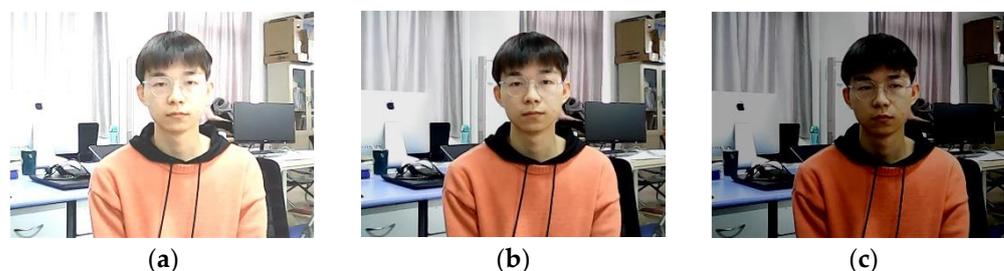


Figure 5. Examples of live faces in self-built dataset. (a) High light. (b) Normal light. (c) Dark light.



Figure 6. Examples of video faces in self-built dataset. (a) High light. (b) Normal light. (c) Dark light.

In this experiment, relevant parameters of the training network model were set as a learning rate of 0.00001, iterations were 100 times, the batch size was 32, and the images input to the network were normalized to 224×224 pixels.

3.2. Evaluation Metrics

A face anti-spoofing system is subject to two types of errors: either the real access is rejected (false rejection), or an attack is accepted (false acceptance). Its performance is often measured with the Half Total Error Rate (HTER), which is half of the sum of the False Rejection Rate (FRR) and the False Acceptance Rate (FAR). Since both the FAR and the FRR depend on a threshold τ , increasing the FAR will usually reduce the FRR and vice-versa. For this reason, results are often presented using the Receiver Operating Characteristic (ROC) curve, which plots the FAR versus the FRR. Equal Error Rate (EER) is defined as the point along the ROC curve where the FAR equals the FRR. Finally, we also used Accuracy (ACC) to evaluate the performance of the method by the percentage of correctly classified examples:

$$\text{Accuracy} = \frac{TP + TN}{FP + FN + TP + TN} \quad (6)$$

where TP , TN , FP , FN represent the number of true positives, true negatives, false positives, and false negatives, respectively.

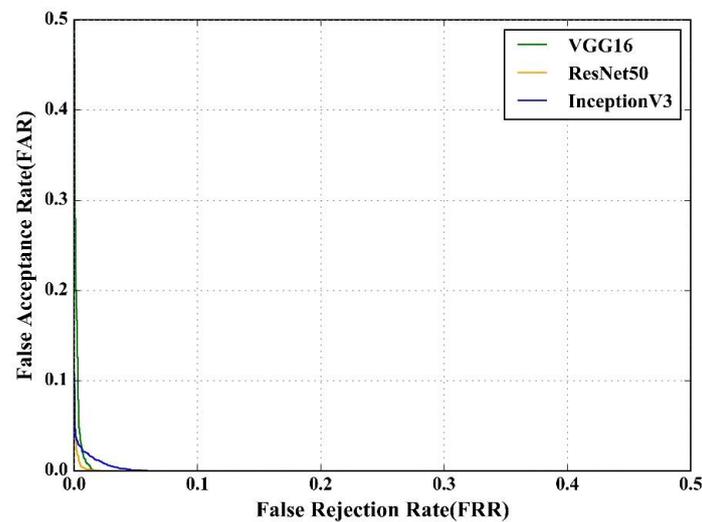
3.3. Experimental Results and Analysis

3.3.1. Comparison Experiment of Different Network Models

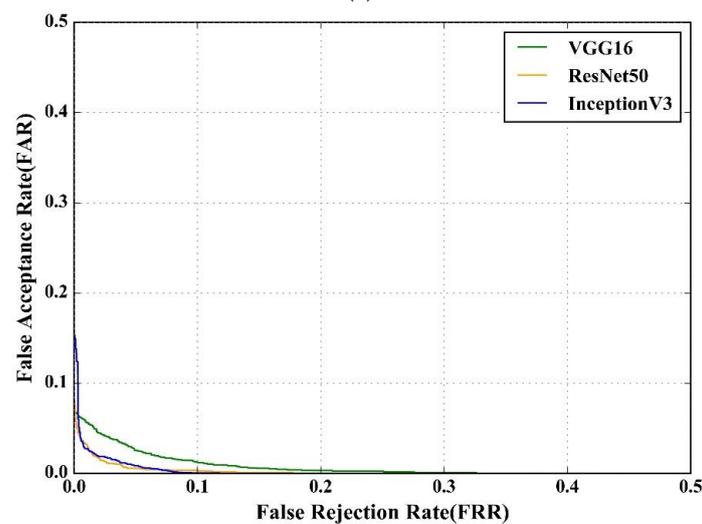
To verify whether the deep residual network performs better than other ordinary deep networks in the face anti-spoofing task, we conducted a comparative experiment. The experiment selected VGG16 [18], and InceptionV3 [19] for comparative experiments. In the results shown in Table 1 and Figure 7, ResNet50 has a higher detection accuracy than VGG16 and InceptionV3 on Replay-Attack and CASIA-FASD datasets.

Table 1. Comparison table of results for accuracy of network model detection.

Network Model	Replay-Attack	CASIA-FASD
VGG16	0.9408	0.9305
InceptionV3	0.9381	0.9232
ResNet50	0.9694	0.9555



(a)



(b)

Figure 7. ROC curves for different network models. (a) Replay-Attack. (b) CASIA-FASD.

In order to further analyze the experimental results, this paper adopts class activation mapping (CAM) [20] to visualize the network output. CAM can indicate where the predicted value of the network is more sensitive to the pixel value in a certain part of the input image. The more sensitive area represented by the CAM, the darker the color, which means that the area is more distinguishable. Figures 8–10 show the results. Comparing the CAM predicted by ResNet50, InceptionV3, and VGG16, the CAM predicted by the ResNet50 network has a deeper color and a larger region, indicating that distinctive feature regions are learned by the ResNet50 network, and the predicted probability value would also increase accordingly. According to the structure of the model, the analysis is because the VGG16 network does not have residual settings, and the effective features may be missing during the continuous convolution process. Thus, the CAM effective region is smaller than the ResNet50 network. ResNet50 contains residual structure in the continuous convolution of the network to extract features. The effective features of the shallow layer are retained and added to the features extracted from the deep layer, so the effective region of the CAM map is wider and concentrated. On the other hand, the InceptionV3 structure connects the convolution kernels in parallel, the effective region of both living and non-living faces is concentrated in the region below the mouth, and the difference in the mouth is not obvious. According to the deviation of the predicted probability, InceptionV3 is not very suitable for face anti-spoofing. Comprehensive analysis shows that ResNet50 with a deeper layer and residual structure has a better effect on face anti-spoofing than VGG16 and InceptionV3.

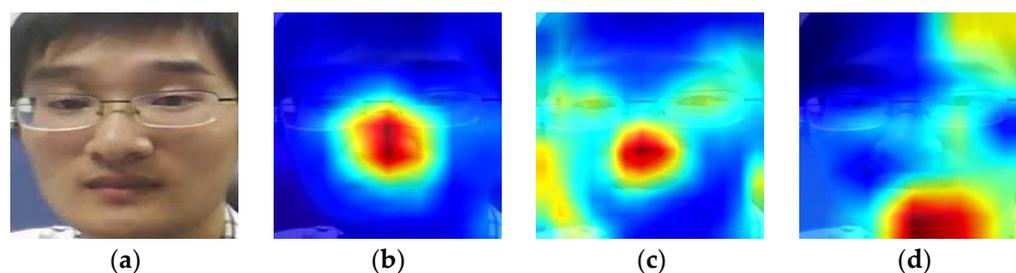


Figure 8. The visualization results of live face in the last layer network. (a) Original. (b) ResNet50. (c) VGG16. (d) InceptionV3.

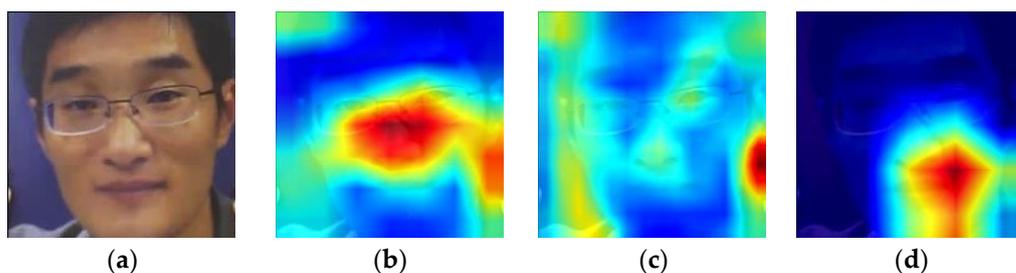


Figure 9. The visualization results of photo face in the last layer network. (a) Original. (b) ResNet50. (c) VGG16. (d) InceptionV3.

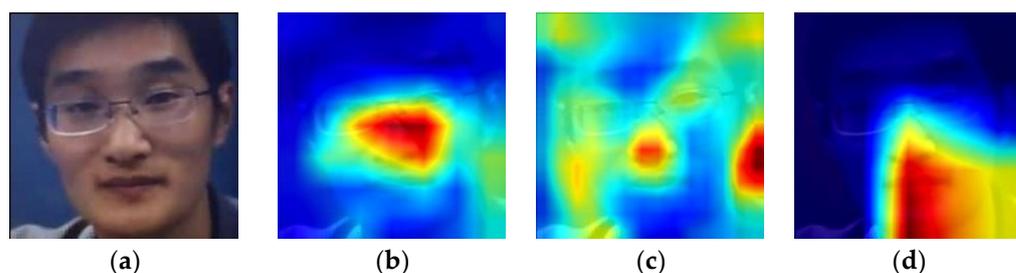


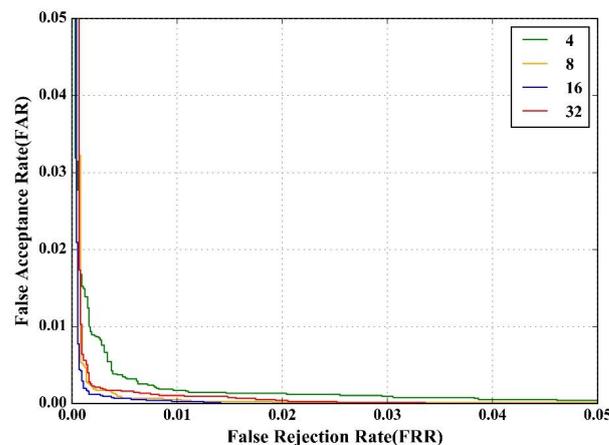
Figure 10. The visualization results of video face in the last layer network. (a) Original. (b) ResNet50. (c) VGG16. (d) InceptionV3.

3.3.2. Comparative Experiment for the Parameter r of SE Module

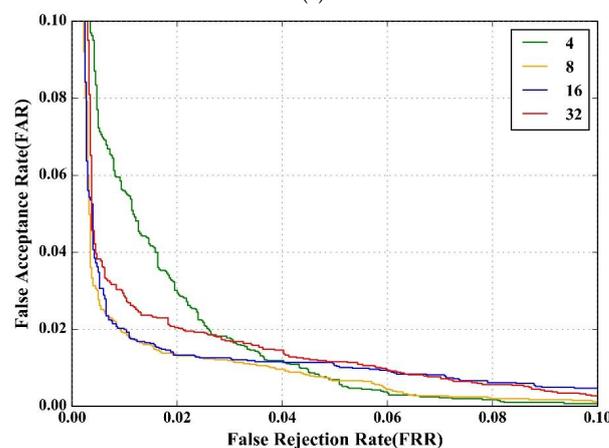
The reduction ratio r is an important parameter that can vary the capacity and computational cost of the SE module in the model, and the value of r can also affect the performance of the network model. So, in order to obtain the best value of the parameter r suitable for this face anti-spoofing research, we set the parameter r at the values of 4, 8, 16, and 32 for experiments, and compared experiments with the residual network without the SE module. The results are shown in Table 2 and Figure 11; the result of Table 2 shows that adding the SE module can effectively improve model performance. When the parameter r is smaller, the model is larger, but the model performance does not improve with the increase of the model parameters. Moreover, from Figure 11, we can see that when the value of r is 8 and 16, the model performance is relatively better. We Considered that the actual situation would make the model easier to deploy. The value of r of 16 can be a good choice between the performance and size of the model, so the parameter r is set to 16 in this paper.

Table 2. Comparison table of results for parameter r .

r	Replay-Attack		CASIA-FASD		Model Parameters (Ten Million)
	EER	ACC	EER	ACC	
-	0.46	0.9694	2.84	0.9555	-
4	0.22	0.9978	2.40	0.9697	2637
8	0.09	0.9991	1.76	0.9783	2498
16	0.02	0.9998	2.02	0.9775	2428
32	0.14	0.9986	2.28	0.9717	2393



(a)



(b)

Figure 11. ROC curves for different r values. (a) Replay-Attack. (b) CASIA-FASD.

In the same way, this paper wanted to further analyze the detection effect of the residual network with and without the introduction of the SE module on the living face. We visualized the activation layer output after the first residual block, conv_block2, which is in both SE-ResNet50 and ResNet50. The visualized results are shown in Figure 12. It can be seen that the effective region of the CAM without the SE module is concentrated on the ceiling line and the face contour line of the background, while the effective region of the CAM with the attention mechanism is more concentrated on the face region, and the effective region of the background is also reduced.

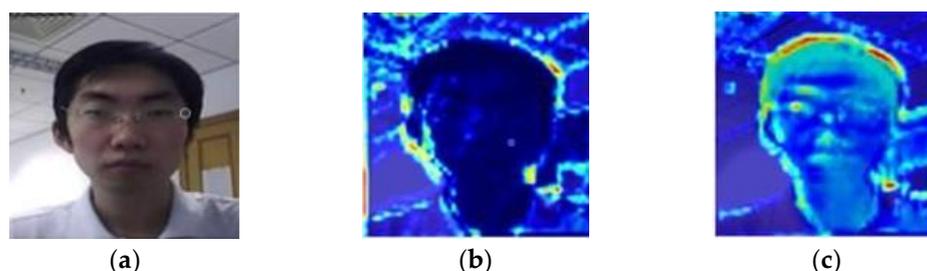


Figure 12. Visualized results. (a) Original. (b) ResNet50. (c) SE-ResNet50.

3.3.3. Comparison Experiment with Existing Methods

To demonstrate the advantages of the model established in this paper, we compared experiments with the LBP [2], LBP-TOP [4], CNN [7], FASNet [8], Patch+depthCNN [21], LiveNet [22] and MSR-Attention [18]. From Table 3, it can be seen that FASNet, Patch+depthCNN, LiveNet, MSR-Attention [23], and SE-ResNet50 are better than the single feature method of the literature [7], LBP [2], and LBP-TOP [4]. Our proposed method, SE-ResNet50, has a low detection error rate and stable performance on Replay-Attack and CASIA-FASD datasets. SE-ResNet50 can capture the key difference regions between living and non-living faces, and it can effectively resist the attacks of non-living faces.

Table 3. Comparison table of results with existing methods.

Method	Replay-Attack		CASIA-FASD	
	EER	HTER	EER	HTER
LBP	14.41	15.45	24.63	23.19
LBP-TOP	7.9	7.6	-	-
CNN	4.414	2.53	6.98	6.99
FASNet	0.16	0.04	5.25	11.34
Patch + depthCNN	0.79	0.72	2.67	2.27
LiveNet	-	5.74	-	4.59
MSR-Attention	0.21	0.39	3.15	-
SE-ResNet50	0.20	0.02	2.02	1.84

3.3.4. Enhance the Robustness of the Method to Illumination

In this experiment, the Lenovo notebook computer was used to adjust the video brightness to simulate the changes in ambient light: the high light was 150% of the normal light intensity, and the dark light was 50% of the normal light. The samples of ten people in the dataset were divided into a training dataset, and the remaining five people were used as the test dataset. The experimental results are shown in Table 4. It can be seen from Table 4 that although the accuracy rate decreases with the change of illumination, the detection accuracy rate of the model is still above 97.73%. Under the three lighting conditions, the performance of the model is not much different. When the training data is given enough samples with different lighting conditions, the model can learn attacks under different lighting conditions. The change of light affects the detection results of living faces much more; for example, the video attack is not affected by the light because of the brightness of the screen, and if the light is too dark or too bright, the paper photo

attack cannot detect the face. Therefore, when the training set is given to the network to learn more live face samples under different illumination, the network model can avoid the influence of illumination to a certain extent and enhance the robustness of the method.

Table 4. Comparison table of results under different ambient lighting.

Ambient Light Conditions	EER	ACC
High light	1.32	0.9792
Normal light	1.13	0.9798
Dark light	1.71	0.9773

4. Conclusions

Focusing on the differential features of facial shadows and details between living and non-living face images, this paper used the channel attention mechanism to transform the feature convolution module of the deep residual network, strengthening the network model's ability to extract and represent key differential features in the nose color mutation region and cheek texture region of the face. We embedded a channel attention mechanism module in the residual block in each convolution block with a total number of four. Thus, we proposed a face anti-spoofing method SE-ResNet50. Compared with the original ResNet50 network, SE-ResNet50 has 3.05% and 2.20% higher detection accuracy on Replay-Attack and CASIA-FASD datasets, respectively. Compared with other existing methods, it has a more stable and excellent detection effect on non-living face photo and video attacks. Finally, we can try to fuse other features to enhance our method, such as the face heart feature, to increase the ability to defend against more realistic 3D masked faces.

Author Contributions: Conceptualization, X.L. and Y.K.; methodology, X.L. and C.L.; software, C.L.; validation, X.L., and C.L.; formal analysis, X.L. and Y.K.; investigation, X.L.; resources, Y.K.; data curation, G.H.; writing—original draft preparation, X.L.; writing—review and editing, X.L. and Y.K.; visualization, C.L.; supervision, Y.K. and G.H.; project administration, Y.K.; funding acquisition, Y.K. and G.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by National Key Research and Development Plan, grant number 2019YFD1100901. This research also was funded by Social development project of Shaanxi provincial key Research and Development. Program: 2022SF-242.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Costa-Pazo, A.; Bhattacharjee, S.; Vazquez-Fernandez, E.; Marcel, S. The replay-mobile face presentation-attack database. In Proceedings of the 2016 International Conference of the Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 21–23 September 2016; pp. 1–7.
- Chingovska, I.; Anjos, A.; Marcel, S. On the Effectiveness of Local Binary Patterns in Face Anti-spoofing. In Proceedings of the 2012 BIOSIG Proceedings of the International Conference of Biometrics Special Interest Group (BIOSIG), Darmstadt, Germany, 6–7 September 2012; IEEE: Piscataway, NJ, USA, 2012; pp. 1–7.
- Cai, R.; Chen, C. Learning deep forest with multi-scale local binary pattern features for face anti-spoofing. *arXiv* **2019**, arXiv:1910.03850.
- Freitas Pereira, T.; Anjos, A.; Martino, J.M.D.; Marcel, S. LBP-TOP based countermeasure against face spoofing attacks. In Proceedings of the Asian Conference on Computer Vision, Daejeon, Korea, 5–9 November 2012; Springer: Berlin/Heidelberg, Germany, 2012; pp. 121–132.
- Kong, Y.; Liu, X.; Xie, X.; Li, F. Face Liveness Detection Method Based on Histogram of Oriented Gradient. *Laser Optoelectron. Prog.* **2018**, *55*, 237–243.
- Boulkenafet, Z.; Komulainen, J.; Hadid, A. Face Spoofing Detection Using Colour Texture Analysis. *IEEE Trans. Inf. Forensics Secur.* **2017**, *11*, 1818–1830. [[CrossRef](#)]
- Yang, J.; Lei, Z.; Li, S.Z. Learn Convolutional Neural Network for Face Anti-Spoofing. *Comput. Sci.* **2014**, *9218*, 373–384.
- Lucena, O.; Junior, A.; Moia, V.; Souza, R.; Valle, E.; Lotufo, R. Transfer learning using convolutional neural networks for face anti-spoofing. In Proceedings of the International Conference Image Analysis and Recognition, Montreal, QU, Canada, 5–7 July 2017; Springer: Cham, Switzerland, 2017; pp. 27–34.

9. Deng, X.; Wang, H.C. Face liveness detection algorithm based on deep learning and feature fusion. *J. Comput. Appl.* **2020**, *40*, 1009–1015.
10. Luan, X.; Li, X.S. Face anti-spoofing algorithm based on multi-feature fusion. *Comput. Sci.* **2021**, *48*, 409–415.
11. Yu, Z.; Qin, Y.; Li, X.; Wang, Z.; Zhao, C.; Lei, Z.; Zhao, G. Multi-modal face anti-spoofing based on central difference networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 13–19 June 2020; pp. 650–651.
12. Cai, P.; Quan, H. Face anti-spoofing algorithm combined with CNN and brightness equalization. *J. Cent. South Univ.* **2021**, *28*, 194–204. [[CrossRef](#)]
13. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; IEEE: Piscataway, NJ, USA, 2018; pp. 7132–7141.
14. He, J.; Jiang, D. Fully automatic model based on SE-resnet for bone age assessment. *IEEE Access* **2021**, *9*, 62460–62466. [[CrossRef](#)]
15. Yoo, J.; Jin, Y.; Ko, B.; Kim, M.S. k-Labelsets Method for Multi-Label ECG Signal Classification Based on SE-ResNet. *Appl. Sci.* **2021**, *11*, 7758. [[CrossRef](#)]
16. Tian, F.; Wang, L.; Xia, M. Signals Recognition by CNN Based on Attention Mechanism. *Electronics* **2022**, *11*, 2100. [[CrossRef](#)]
17. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 770–778.
18. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv* **2014**, arXiv:1409.1556.
19. Szegedy, C.; Vanhoucke, V.; Ioffe, S.; Shlens, J.; Wojna, Z. Rethinking the inception architecture for computer vision. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 2818–2826.
20. Zhou, B.; Khosla, A.; Lapedriza, A.; Oliva, A.; Torralba, A. Learning deep features for discriminative localization. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; IEEE: Piscataway, NJ, USA, 2016; pp. 2921–2929.
21. Atoum, Y.; Liu, Y.; Jourabloo, A.; Liu, X. Face anti-spoofing using patch and depth-based CNNs. In Proceedings of the 2017 IEEE International Joint Conference on Biometrics (IJCB), Denver, CO, USA, 1–4 October 2017; IEEE: Piscataway, NJ, USA, 2017; pp. 319–328.
22. Rehman, Y.A.U.; Po, L.M.; Liu, M. LiveNet: Improving features generalization for face liveness detection using convolution neural networks. *Expert Syst. Appl.* **2018**, *108*, 159–169. [[CrossRef](#)]
23. Chen, H.; Hu, G.; Lei, Z.; Chen, Y.; Robertson, N.M.; Li, S.Z. Attention-based two-stream convolutional networks for face spoofing detection. *IEEE Trans. Inf. Forensics Secur.* **2019**, *15*, 578–593. [[CrossRef](#)]