FACE ANTI-SPOOFING BASED ON MULTI-LAYER DOMAIN ADAPTATION

Fengshun Zhou[1,2,*], *Chenqiang Gao*[1,2], *Fang Chen*[1,2], *Chaoyu Li*[1,2], *Xindou Li*[1,2], *Feng Yang*[1,2], *Yue Zhao*[1,2]

[1]School of Communication and Information Engineering,
Chongqing University of Posts and Telecommunications, Chongqing, China
[2]Chongqing Key Laboratory of Signal and Information Processing, Chongqing 400065, China
zhoufengshun1212@gmail.com

ABSTRACT

With the popularity of face recognition technology, people have put forward higher requirements for the security of face recognition system. Face anti-spoofing detection attracts extensive attention and many methods been proposed. However, these methods perform poorly in cross scenes. To solve this problem, we propose a face anti-spoofing detection algorithm based on domain adaptation. We apply Maximum Mean Discrepancy (MMD) to multi-layer network distribution adaptation, which improves the generalization ability of the model. To further improve the performance of face anti-spoofing detection, we fuse the low-level features with the high-level features of convolutional neural network for face anti-spoofing detection. Two widely used datasets are used to test the proposed method. The experimental results show that the proposed algorithm outperforms state-of-the-art approaches.

Index Terms— face anti-spoofing detection, deep learning, domain adaptation, Maximum Mean Discrepancy

1. INTRODUCTION

With the wide application of face recognition, the security of face recognition systems has become more and more important. Face recognition technology is mainly based on image information, so it is very vulnerable to attack. The attacker usually uses the fake face images or videos to cheat face recognition systems. Face anti-spoofing detection is to prevent the system from being cheated, which is an import part of the face recognition system.

Face liveness detection attracts extensive attention and many methods have been proposed. Traditional face anti-spoofing detection approaches include motion based method [1], texture based method [2] and multi-clue based method [3]. Deep learning based methods have also been adopted for face anti-spoofing detection [4].

Generally, the state-of-the-art methods have achieved satisfactory results in a specific scene. However, the trained model performs poorly in cross scenes [3, 4, 5] where the model is trained on one data set and tested on another. For example, the model [4] trained on Replay [6] has a Half Total Error Rate (HTER) of 2.1% on Replay, while has a HTER of 45.5% tested on CBSR [7]. The main reason for the poor generalization of the model is that the two datasets are collected under different environments, different devices and different distances from subject to camera, which results in inconsistent data distribution. Thus, cross scenes anti-spoofing detection is the focus of current research.

In order to improve the generalization ability in the cross scenes, domain adaptation is applied to face anti-spoofing detection. Domain adaptation can improve the generalization ability of the model by narrowing the data distribution difference of source domain and target domain. In [8], the authors proposed to apply MMD [9] to handle the face anti-spoofing detection. However, [8] only adopted MMD in the last full connected layer and ignored the impact of other full connected layers on data distribution. The fully connected layers will expand the distribution difference between the source domain and the target domain [10], which further result in poor generalization ability of the face anti-spoofing detection. To narrow the distribution differences between source domain and the target domain, we apply MMD to the multilayer full connected layers (ML-MMD). In this paper, ML-MMD is adopted to map the features of the source and target domains to the Reproducing Kernel Hilbert Space (RKHS) and minimize the distribution differences of two domains.

Besides, there are many differences in the details between the genuine face images and the fake face ones. To further improve the performance of face anti-spoofing detection, we fuse the low-level features with the high-level features of CNN for face anti-spoofing detection. Experimental results show that the proposed method have achieved advanced generalization ability.

The rest of this paper is organized as follows. In Section II, we briefly review the related works on face anti-spoofing detection. In Section III, the proposed multi-layer domain adaptation algorithm is introduced. Quantitative experimental results and comparative experimental results are shown in Section IV and Section V conclude this paper.

2. RELATED WORKS

Due to the diversity of spoofing attacks, in the past few years, numerous face anti-spoofing detection techniques have been proposed. we categorize existing face anti-spoofing detection methods into four categories: motion based, texture based, deep learning based and domain adaptation based.

Motion based Face Anti-spoofing: The main idea of the motion-based approach is to use the distinguished motion feature between genuine faces and fake faces. As compared to fake faces, genuine faces have subtle motions such as blinking, lip movements, and head rotation. In [11], the subtle movements of different facial parts were extracted as important features under the assumption that the genuine and fake faces can be distinguished by the movement cues. More recently, a novel motion-based countermeasure which exploits natural and unnatural motion cues is presented in [1].

Although motion based methods are effective against the attacks of video replay, they may suffer degraded performance when the spoofing attack is conducted by printing photo.

Texture based Face Anti-spoofing: In [12] the authors used a total-variation based decomposition method and the difference-of-Gaussian (DoG) filter to extract potential high-frequency features in the face image, and then a sparse low rank bilinear discriminative model was trained for the classification.

After that, [5] applied multi-scale local binary patterns features for face anti-spoofing, which performed better than most existing methods. Dynamic texture recognition using volume local binary count patterns is applied to 2D face spoofing detection in [13].

Though the above texture feature descriptors can effectively detect different manner of attacks, they could be very sensitive to different illuminations and other external noises.

Deep learning based Face Anti-spoofing: Recently, CNN has achieved great performance in computer vision tasks [14], and deep learning based methods have also been adopted for face anti-spoofing detection. In [4], the CNN was first utilized as a feature extractor for face anti-spoofig detection.

And in [15], the authors proposed an LSTM-CNN architecture which utilized temporal information to conduct a joint prediction for multiple frames of a video and achieve remarkable improvements in the intra-test. Although the CNN-based face anti-spoofing detection research has achieved excellent results, it is difficult to retrain as the existing face anti-spoofing database is small.

Domain adaptation based Face Anti-spoofing: To solve this problem of poor generalization ability of cross scenes detection, domain adaptation methods are used in the field of face anti-spoofing detection, but few related articles have been published. [16] proposed a person-specific face anti-spoofing approach. They assumed that for an individual subject, there was a linear relationship between the genuine and fake samples. In [17], the generalization ability of face anti-spoofing is improved by mapping the extracted features in the similar distributed subspace and narrowing the data distribution. Considering the spatial and temporal characteristics of samples, [8] proposed a 3D CNN for face anti-spoofing detection. They further improved the generalization of the face anti-spoofing detection by adding regular term in the loss function.

3. PROPOSED METHOD

The proposed algorithm framework is shown in Figure 1. Our method consists of two parts: feature fusion and domain adaptation. For the feature fusion part, we train an improved CNN to extract the useful information of low-level which are capable of discriminating genuine and fake face images. For the domain adaptation part, we apply ML-MMD to narrow data distribution differences and improve the generalization ability of the model in cross scenes.

3.1. Feature fusion

There are many differences in the details between the real face images and the fake face ones. To extract more detail features, we add two skip connections after the first two pooling layers. The skip connections help the network to utilize extracted features from layers with different depths, whiche is similar to the FCN structure [18].

Face detection is first conducted to obtain the face region. Then, the face region is fed into the convolution layers which is extended by feature fusion. We pooled the low-level features to ensure that the features of the fusion layer have the same dimensions. Next, we add the features of the last three layers of convolution. Finally, the fused features are fed into fully connected layer and we apply the domain adaption.

3.2. Domain adaptation

To solve the problem of poor performance of face anti-spoofing detection in cross scenes, we apply unsupervised domain adaption to minimize the distribution distance between the two data sets. In this paper, the source domain $D_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$ are given with labeled examples and the target domain $D_t = \{x_j^t\}_{j=1}^{n_t}$ are unlabeled examples. The samples of the source domain and the target domain are sampled from the probability distributions of p and q, respectively.

3.2.1. Maximum mean discrepancy

The MMD function [9] can measure the distance between two probability distributions. To improve the generalization ability of the model, we narrow the distribution differences be-

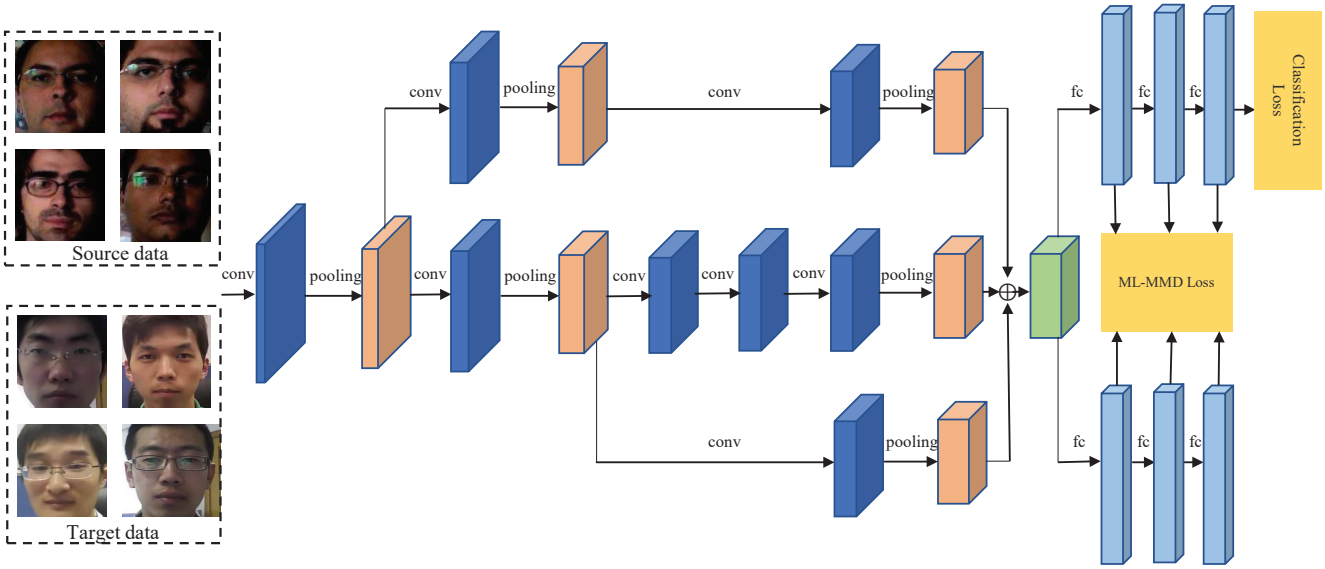

Fig. 1. The framework of the proposed method.

tween source and target domains by minimizing the MMD distance.

$$MMD[F,p,q] := \sup_{f\in F}(E_p[f(D_s)] - E_q[f(D_t)]), \quad (1)$$

where F is the set of functions that map the eigenspace to the set of real numbers. If $p = q$, it means that D_s and D_t are distributed in the same eigenspace with the same expectation, thus, $MMD = 0$. If $p \neq q$, we need to give F a constraint. In [9], the authors proved that F belongs to the unit sphere in RKHS. In the RKHS space,

$$f(x) =< f, \varphi(x) >_h, \quad (2)$$

where $\varphi(x)$ refers to the embedding function which maps the data from feature space to the RKHS.

In this paper, we map the sample features to the RKHS space and then calculate the distribution distance. Hence the distance of MMD in RKHS is expressed as:

$$\begin{aligned} MMD[F,p,q] &= \sup_{||f||_h\leq 1}(E_p[f(D_s)] - E_q[f(D_t)]) \\ &= \sup_{||f||_h\leq 1}(E_p[<\varphi(D_s), f>_h] - E_q[<\varphi(D_t), f>_h]) \\ &= \sup_{||f||_h\leq 1}(<u_p - u_q, f>_h) \\ &= ||u_p - u_q||_h. \end{aligned} \quad (3)$$

To be specific, u_p means $E_p[\varphi(D_s)]$ and u_q means $E_q[\varphi(D_t)]$.

3.2.2. *Multi-layer domain adaptation*

The MMD can improve the generalization ability of the model in cross scenes and previous work [8] showed its effectiveness on face anti-spoofing detection. However, [8] only applied MMD to the last fully connected layer, without considering the impact of other layers on the data distribution. The adaptation for the last full connected layer can not eliminate the distribution differences between the source and target domains, since there are other fully connected layers that are not transferable.

In [10], the authors proved that the fully connected layers will expand the distribution difference between the source domain and the target domain, which further results in poor generalization ability of the face anti-spoofing detection. Therefore, to narrow the distribution differences between source domain and the target domain, we apply ML-MMD in the network. The advantage of multi-layer adaptation is that by combining the representation layer with the classifier layer, we can essentially build the bridge for the domain discrepancy underlying both the marginal and conditional distributions, which is crucial for eliminating distribution differences.

Our architecture (see Figure 1) consists of five convolution layers and three full connected layers. The convolution layers weights are shared by source domain and target domain. We use the source domain data to compute the classification loss and all data to compute the domain loss. The total loss function is composed of domain loss and the classification loss.

$$L = L_c + \lambda \sum_{l=l_1}^{l_2} MMD_k^2(D_l^s, D_l^t), \quad (4)$$

where $\lambda > 0$ is a penalty parameter, l_1 and l_2 represent indexes for the network layer. In our implementation, we set l_1 = 6 and l_2 = 8. L_c is the classification loss function. D_l^s and D_l^t represent the data of source domain and target domain at layer l, respectively.

3.3. Implementation details

In this paper, we use the face detection algorithm [19] to detect the face localization in each video frame. To eliminate the interference of background information we cut out the face part and resize it to 227×227 as the network input.

The weight λ of the ML-MMD regularization term is set in the way where at the end of training, the classification loss and ML-MMD regularization term loss are approximately the same. Such setting is reasonable since the feature representation has both discrimination and generalization ability can be learned. More specifically, the weight λ is selected in $\{0.1, 0.4, 0.7, 1, 1.4, 1.7, 2\}$. The learning rate is 0.0001 and decays 10% every 10 epochs for both training and fine-tuning procedure. The network is trained with the adaptive moment estimation (Adam) method. The batch size of training data is 16 and the weight decay is set to 0.00005.

4. EXPERIMENTS RESULTS

4.1. Datasets and evalution criteria

In this paper, we validate our proposed method with extensive experiments on two public face anti-spoofing databases: CASIA Face Anti-Spoofing Database (CBSR) [7] and Replay-Attack Database (Replay) [6].

The CBSR, which has a total of 50 human subjects, consists of 600 video recordings of genuine and fake attacks and each subject has 12 sequences (3 genuine and 9 fake ones). Three fake attacks were designed: warped photo attacks, cut photo attacks, and video attacks, as shown in Figure 2. The genuine and the fake attacks were recorded using three camera devices with: low, normal and high resolutions. The 50 subjects were divided into two subject-disjoint subsets for training and testing (20 and 30, respectively).

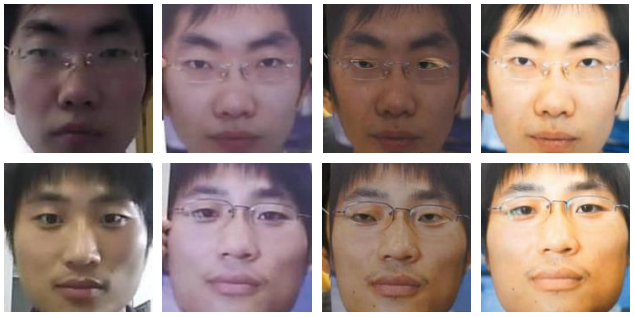

Fig. 2. Samples from the CBSR. From the left to the right: genuine faces and the corresponding warped photo, cut photo and video replay attacks.

The Replay also contains 50 subjects, which were divided into 3 subject-disjoint subsets for training, development and testing (15, 15 and 20, respectively), and consists of 1300 video clips of genuine and fake attacks. The genuine videos are recorded under two different lighting conditions: controlled and adverse. Two type of attacks are created: replay attacks and print attacks. In the replay attacks, high quality videos of the genuine client are replayed on iPhone 3GS and iPad display devices. For print attacks, the genuine images are printed on A4 papers or reacquired by cameras.

Fig. 3. Samples from the Replay database. The first row presents images taken from the controlled scenario, while the second row corresponds to the images from the adverse scenario.

To make a fair comparison with other methods, we followed the overall protocol associated with the two databases. On CBSR database, the results are evaluated in term of Equal Error Rate (EER). The Replay provides a development set to adjust the model parameters. Thus, the results are reported in term of EER on the development set and HTER on the test set.

$$HTER = \frac{FRR(\upsilon, D) + FAR(\upsilon, D)}{2}, \tag{5}$$

where D denotes the used database and the value of υ is estimated on the *EER*. $FRR(\upsilon, D)$ means the false rejection rate for the genuine face and $FAR(\upsilon, D)$ means the false acceptance rate for the fake faces.

4.2. Results of intra-test

We perform intra testing on Replay and CBSR databases and compare the proposed method with the state-of-the-art methods. Table 1 shows the EER and HTER of advanced face anti-spoofing methods: the LBP+HOOF based mothod [20], the IDA based method [3], the color analysis based methods [5, 2] and the dynamic texture based mothd [13]. From Table 1, it can be seen that our method outperforms many state-of-the-art algorithms on the two challenging databases.

More specifically, for CBSR, our performance exceeds most results with an EER of 3.7%, close to the best results of [2]. On the Replay database, our proposed method outperforms other methods in EER and HTER, with only 0.3% and 0.6%, respectively. This shows that the features extracted by our proposed method contain more discriminant information.

Table 1. Results of intra tests of different methods on the datasets.

	Replay		CBSR
Method	EER	HTER	EER
LBP+HOOF [20]	-	-	3.1
IDA [3]	-	7.4	-
LBP [5]	1.5	5.1	8.8
Color texture [2]	0.4	2.8	2.1
Dynamic texture [13]	1.7	0.8	6.5
Ours	**0.3**	**0.6**	3.7

4.3. Results of inter-test

To demonstrate the generalization of our method, we conducted a cross scenes evaluation. Table 2 shows the HTER of inter-test on CBSR and Replay. When the model is trained on Replay, the HTER on CBSR is 34.3%. And when the model is trained on CBSR, the cross scenes performance on Replay is 33.1%. The results show that the inter-test performance of our method is better than other baseline methods. Especially when the model is trained on Replay, the HTER on the CBSR achieves the best, which demonstrates that the proposed model possesses the capability of generalization.

Table 2. Results of HTER% inter tests of different methods on the datasets.

	train	test	train	test
Method	CBSR	Replay	Replay	CBSR
LBP+HOOF [20]	35.4		44.6	
IDA [3]	**26.9**		43.7	
LBP [5]	37.9		35.4	
Color texture [2]	30.3		37.7	
Motion-based [1]	33.7		49.3	
Ours	33.1		**34.3**	

4.4. Ablation Study

To analyze effects of each components of the proposed method, we conduct ablation studies on the Replay and CBSR dataset. The qualitative results of intra-test and inter-test are listed in Table 3 and Table 4, respectively.

First, we fine-tune the pre-training model with spoofing data, and get acceptable results in the intra-test. This is possibly due to the fact that the small difference between genuine and fake faces in these two data sets. With a small amount of data training, a network with classification capabilities can be achieved.

In addition, we extend the network structure by fusing low-level information (Only-fusion) to ensure adequate texture information and information integrity. From Table 3, it can be seen that the low-level information are important for face anti-spoofing detection. In particular, for the CBSR dataset, EER decreased from 4.8% to 3.4%, which exceeds most advanced methods. Although satisfactory results have been obtained in the intra-test, the performance drops dramatically while dealing with cross scenes testing, as shown in Table 4. This could be due to the fact that the model is overfitting and the distribution differences between the data sets result in the learned features without generalization ability.

Table 3. Results of intra tests of different strategies on the datasets.

	Replay		CBSR
	EER	HTER	EER
Finetune	1.9	2.1	4.8
Only-fusion	0.9	0.8	3.4
Only-ML-MMD	1.7	3.3	3.8
All	**0.3**	**0.6**	3.7

Table 4 shows that adding the ML-MMD regular term (only-ML-MMD) can significantly improve performance of inter-test. Specifically, when using Replay for training, the HTER for the tested CBSR dropped from 58.8% to 41.1%, and the HTER dropped from 49.4% to 37.9% on Replay when using CBSR for training. On the opposite, as shown in Table 3, the intra-test performance decreased after the addition of ML-MMD. This indicates that the regularization overemphasizes the similarity of the two data sets and ignores the classification ability.

Table 4. Results of HTER% inter tests of different strategies on the datasets.

	train	test	train	test
	CBSR	Replay	Replay	CBSR
Finetune	49.4		58.8	
Only-fusion	43.9		60.1	
Only-ML-MMD	37.9		41.1	
All	**33.1**		**34.3**	

The best results can be obtained in inter-test when combining all strategies (All). In particular, the HTER is further reduced to 33.1% and 34.3% on two datasets. This shows that the features learned from the proposed model possess generalization abilities.

5. CONCLUSION

In this paper, we propose a novel face anti-spoofing detection based on multi-layer domain adaption. This method extends AlexNet by fusing low-level features to get more texture information which is beneficial to distinguish subtle differences. In addition, ML-MMD distance loss is adopted to eliminate distribution differences cross scenes. Experimental results reveal that our method outperforms state-of-the-art methods, and has obvious advantage on generalization ability.

6. ACKNOWLEDGMENT

This work is supported by the National Natural Science Foundation of China (No. 61571071), Chongqing Research Program of Basic Research and Frontier Technology (No. cstc2018jcyjAX0227).

7. REFERENCES

[1] Taiamiti Edmunds and Alice Caplier, "Motion-based countermeasure against photo and video spoofing attacks in face recognition," *Journal of Visual Communication and Image Representation*, vol. PP, no. 50, pp. 314–332, 2018.

[2] Zinelabidine Boulkenafet, Jukka Komulainen, and Abdenour Hadid, "Face spoofing detection using colour texture analysis," *IEEE Transactions on Information Forensics & Security*, vol. 11, no. 8, pp. 1818–1830, 2017.

[3] Di Wen, Hu Han, and Anil K. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics & Security*, vol. 10, no. 4, pp. 746–761, 2015.

[4] Jianwei Yang, Zhen Lei, and Stan Z Li, "Learn convolutional neural network for face anti-spoofing," *Computer Science*, vol. 9218, pp. 373–384, 2014.

[5] Z Boulkenafet, J Komulainen, and A Hadid, "Face anti-spoofing based on color texture analysis," in *IEEE International Conference on Image Processing*, 2015, pp. 2636–2640.

[6] Ivana Chingovska, Andr Anjos, and Sbastien Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *Biometrics Special Interest Group*, 2012, pp. 1–7.

[7] Zhiwei Zhang, Junjie Yan, Sifei Liu, Zhen Lei, Dong Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *Iapr International Conference on Biometrics*, 2012, pp. 26–31.

[8] Haoliang Li, Peisong He, Shiqi Wang, Anderson Rocha, Xinghao Jiang, and Alex C. Kot, "Learning generalized deep feature representation for face anti-spoofing," *IEEE Transactions on Information Forensics & Security*, vol. 13, no. 10, pp. 2639–2652, 2018.

[9] A. Gretton, K M Borgwardt, Malte Rasch, B. Schlkopf, and A. Smola, "A kernel two-sample test," *Journal of Machine Learning Research*, vol. 13, no. 1, pp. 723–773, 2012.

[10] Shuwen Qiu and Weihong Deng, "Deep local descriptors with domain adaptation," in *Pattern Recognition and Computer Vision*, 2018, pp. 344–355.

[11] K. Kollreider, H. Fronthaler, and J. Bigun, "Evaluating liveness by face images and the structure tensor," in *IEEE Workshop on Automatic Identification Advanced Technologies*, 2005.

[12] Xiaoyang Tan, Li Yi, Jun Liu, and Jiang Lin, "Face liveness detection from a single image with sparse low rank bilinear discriminative model," in *European Conference on Computer Vision*, 2010.

[13] Xiaochao Zhao, Yaping Lin, Janne Heikkila, Xiaochao Zhao, Yaping Lin, Janne Heikkila, Xiaochao Zhao, Yaping Lin, and Janne Heikkila, "Dynamic texture recognition using volume local binary count patterns with an application to 2d face spoofing detection," *IEEE Transactions on Multimedia*, vol. PP, no. 99, pp. 1–1, 2017.

[14] Lan Wang, Chenqiang Gao, Luyu Yang, Yue Zhao, Wangmeng Zuo, and Deyu Meng, "Pm-gans: Discriminative representation learning for action recognition using partial-modalities," in *The European Conference on Computer Vision (ECCV)*, September 2018.

[15] Zhenqi Xu, Li Shan, and Weihong Deng, "Learning temporal features using lstm-cnn architecture for face anti-spoofing," in *Pattern Recognition*, 2016.

[16] Jianwei Yang, Lei Zhen, Yi Dong, and Stan Z. Li, "Person-specific face antispoofing with subject domain adaptation," *IEEE Transactions on Information Forensics & Security*, vol. 10, no. 4, pp. 797–809, 2017.

[17] Haoliang Li, Li Wen, Cao Hong, Shiqi Wang, Feiyue Huang, and Alex C. Kot, "Unsupervised domain adaptation for face anti-spoofing," *IEEE Transactions on Information Forensics & Security*, vol. 13, no. 7, pp. 1794–1809, 2018.

[18] Jonathan Long, Evan Shelhamer, and Trevor Darrell, "Fully convolutional networks for semantic segmentation," *IEEE Transactions on Pattern Analysis & Machine Intelligence*, vol. 39, no. 4, pp. 640–651, 2014.

[19] Kaipeng Zhang, Zhanpeng Zhang, Zhifeng Li, and Qiao Yu, "Joint face detection and alignment using multitask cascaded convolutional networks," *IEEE Signal Processing Letters*, vol. 23, no. 10, pp. 1499–1503, 2016.

[20] Talha Ahmad Siddiqui, Samarth Bharadwaj, Tejas I. Dhamecha, Akshay Agarwal, Mayank Vatsa, Richa Singh, and Nalini Ratha, "Face anti-spoofing with multifeature videolet aggregation," in *International Conference on Pattern Recognition*, 2017.