2018 24th International Conference on Pattern Recognition (ICPR)
Beijing, China, August 20-24, 2018

Face Anti-Spoofing with Multi-Scale Information

Shiying Luo*†, Meina Kan*‡, Shuzhe Wu*†, Xilin Chen* and Shiguang Shan*‡
*Key Lab of Intelligent Information Processing of Chinese Academy of Sciences (CAS),
Institute of Computing Technology, CAS, Beijing 100190, China
†University of Chinese Academy of Sciences, Beijing 100049, China
‡CAS Center for Excellence in Brain Science and Intelligence Technology

***Abstract*—Face anti-spoofing has encountered increasing demand as one of the key technologies for reliable and safe authentication with faces. Current face anti-spoofing methods generally take a single crop of face region as input for classification, i.e. exploiting information at only one scale. This single-scale scheme mainly focuses on facial characteristics but not utilize the surrounding information, causing poor generalization for different scenarios with varied means of attacks. Besides, it is tedious or highly empirical to determine an optimal scale of face crops. To overcome the limitations of single-scale methods, in this work we propose to integrate Multi-Scale information for better Face ANti-Spoofing (MS-FANS). Specifically, the proposed MS-FANS method takes multiple face crops at different scales as input followed by a convolutional neural network (CNN) for feature extraction. Then the features from different scales form as a sequence, which are fed into a Long Short-Term Memory (LSTM) network for adaptive fusion of multi-scale information, constructing the final representation for classification. Benefited from this multi-scale design, MS-FANS can adaptively utilize context information from multiple scales, leading to promising performance on two challenging face anti-spoofing datasets, Idiap REPLAY-ATTACK and CASIA-FASD, with significant improvement compared with the existing methods.**

I. Introduction

Face anti-spoofing has become a necessary task in most application of daily access control and authentication with faces. As the human face is a type of easily acquired biometrics feature, anti-spoofing with faces is an effective and efficient technique to improve reliability and safety of an access or authentication system. There are several types of face spoofing attacks, including photo printing attack, video replay attack, 3D mask attack, etc. Thereinto, the 3D mask attack is costly, while photo printing attack and video replay attack are low-cost which are commonly utilized to attack a system. Therefore, this work mainly focus on photo printing attack and video attack which is a great demand in practice.

Most early face anti-spoofing works are designed with hand-crafted feature such as LBP [1], [2], [3], [4], [5], color-texture [6], movement of lips [7], etc. These methods heavily depend on the human experience. Recently, the learning based methods, especially those deep learning based ones [8], [9], [10], [11], [12], are proposed to learn better feature to distinguish the real faces from the spoofing ones. Although great process has been made, face anti-spoofing from single image is still challenging as only limited information can be obtained from single image. It is observed that most prior face anti-spoofing works only use single-scale information of an image or video, mainly exploiting the characteristic lying in the facial region to determine whether an spoofing appears. Besides the facial region, the fitness between the facial region and background is also beneficial for the anti-spoofing, as showed in Fig. 1. There are also some studies showing that selecting suitable background can improve the performance of anti-spoofing [9], [12]. They both used face images of different scales as input, and found that the performance will increase as the scale increases within a certain range, but when the scale is too large, the performance will be degraded. It is still an open problem to determine an optimal scale to include just right amount of the background.

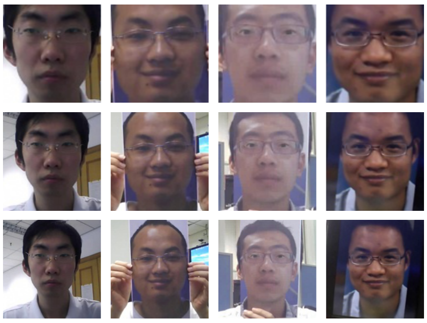


Fig. 1. Face crops in different scales from CASIA-FASD dataset [13]. This example is used to intuitively show the effect of multi-scale information. The images on the same row are in the same scale. The images in the first column are real faces and those in second, third, fourth columns respectively warped photo attack, cut photo attack and video attack. As shown, it is hard to distinguish attacks from the real face with the images in the first row, while it is easier to distinguish them in the second and third rows by including more background.

Differently, in this work, we propose to use multiple scales as input and learn to adaptively integrate facial and background information from multiple scales by using deep neural network (DNN), as showed in Fig. 2. In early works [1], [3], [14] multi-scale feature is indeed shown to be effective. However, they are mainly for hand-crafted feature, and how to apply the same principle in DNN framework is unknown. Unlike traditional multi-scale LBP [1], [3], [14], we use the single-scale convolution kernels but input images of different scales to achieve multi-scale effects. Specifically, multiple crops of different scales from an image are formed as a sequence, then the deep feature of each crop extracted from a CNN is fed into a LSTM to generate the integration weight for each scale, and finally the weighted sum of the feature from each scale, as the

978-1-5386-3788-3/18/$31.00 ©2018 IEEE

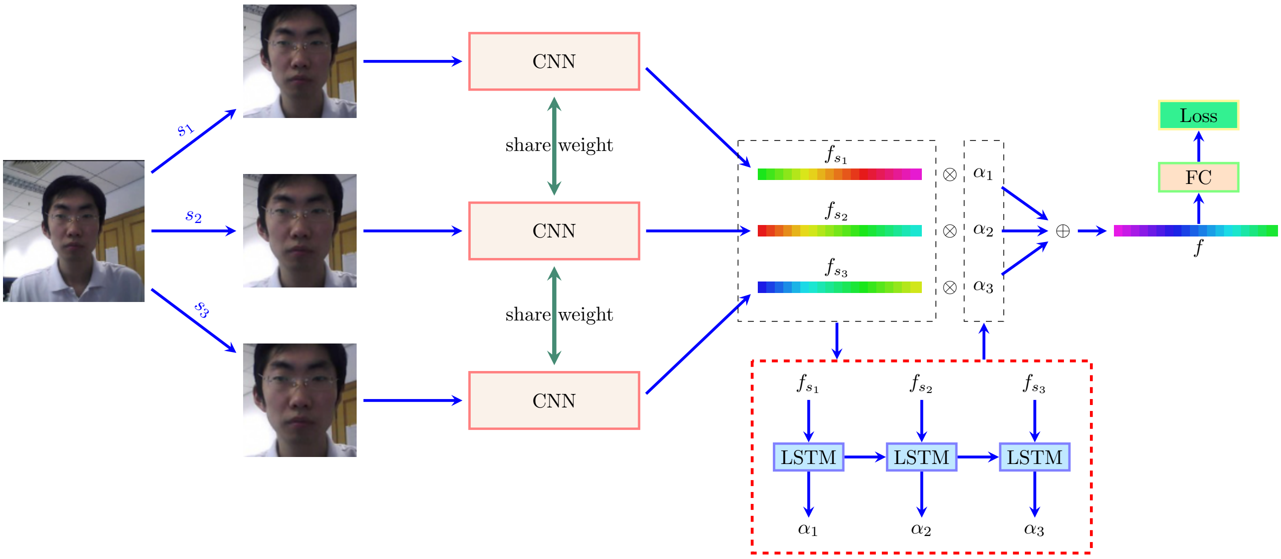


Fig. 2. An overview of the proposed Multi-Scale Face ANti-Spoofing method, referred to as MS-FANS. MS-FANS is designed in an end-to-end manner. Specifically, multiple face crops from different scales are fed into a convolutional neural network (CNN) to extract features. Then the features from different scales are formed as a sequence which is sequentially fed into a Long Short-Term Memory (LSTM) network to generate the weight α for fusing the multi-scale feature. Finally, the representation which adaptively fuses multi-scale information is used for the real vs. spoofing classification.

integrate multi-scale information, is used to determine whether the input face image is a real face or attack one. The overall architecture is optimized in an end-to-end manner.

Briefly, the contributions of this work can be summarized as below:

1) This work integrates multi-scale information in the deep neural network, which is optimized by data-driven feature learning and is superior to traditional hand-crafted features.
2) The integration weights for fusing multiple scale information is adaptively predicted from a LSTM for each input image.
3) The proposed method achieves state-of-the-art performance on several datasets.

This rest of this work is organized as follows: Section II summaries and analyzes the existing face anti-spoofing works. Section III presents a detailed description of our method, MS-FANS. Section IV describes and discusses the experimental evaluation. Finally, the work is concluded in Section V.

II. Related work

The face anti-spoofing methods can be roughly grouped into hand-crafted feature based methods and learning based methods especially those deep learning based methods.

A. *Hand-crafted Feature Based Methods*

A great number of methods put forward different types of hand-crafted features to capture the texture difference between the real faces and attacking ones, including LBP [1], [2], [3], [4], [5], HOG [15], and SURF [16]. Boulkenafet et al. [6] pay attention to HSV and YC_bC_r color space. Bao et al. proposed to employ Optical Flow Maps (OFM) [17] for face anti-spoofing. Unlike normal hand-crafted feature of focusing on the details of the face, Galbally et al. [18] proposed a biometric liveness detection method for iris, fingerprint and face images by using 25 image quality measures, and in [19] Image Distortion Analysis (IDA) has been used for robust face spoof algorithm. Using image quality measures can enhance the generalization ability of the model, because these methods capture the quality difference of face images instead of capturing the facial details [19]. These hand-crafted features generally perform very fast which is favorable for real-time application. However, the discriminative ability of them is usually deficient.

Motion-based methods distinguish real faces from attacking faces by exploiting facial organs or local movements such as eye-blinking [20], [21], and movement of lips [7]. These Motion-based methods usually perform better than the above mentioned appearance-based methods. However, they need collaboration from user which may be inapplicable for those noninvasive scenarios. In addition, movement detection depends on accuracy of landmark detection in the face, which may affect the robustness of movement detection when the landmark detection is inaccurate in some challenging conditions.

B. *Deep Learning Based Methods*

Researchers have explored several ways to use convolutional neural network (CNN) for face anti-spoofing. Yang et al. [9] use CNN as feature extractor and support vector machine (SVM) as classifier to distinguish genuine and spoofing faces. In [10] , fine-tuned VGG-face model and PCA are respectively used to extract deep part features and reduce the dimension. Xu et al. [12] propose an LSTM-CNN architecture to learn the temporal structure from videos, and show that temporal information is helpful for face anti-spoofing. In [11], Atoum et al. firstly use patches to learn local features and then employ the fully convolutional network (FCN) to learn holistic deep features for anti-spoofing.

Deep learning based approaches can learn more discriminative features, and have achieved promising performance for anti-spoofing. Although [11] has some exploration of local and global information, the above methods only exploit single-scale information, leaving room for multi-scale information fusion.

Face anti-spoofing may also be implemented with a variety of other cues, including context [15], depth [22] and etc. Komulainen et al. [15] introduce upper-body detection to help face anti-spoofing. In [22], the depth images captured by Kine are used and achieve remarkable results.

III. Method

The proposed multi-scale face anti-spoofing (MS-FANS) method is designed as an end-to-end stack of a CNN and an LSTM, of which the CNN aims for learning discriminative feature and the LSTM for learning adaptive weights for fusing features from multiple scales. Specifically, a CNN is firstly used to extract features for face crops from multiple scales. Secondly, an LSTM is exploited to learn the fusion weights for features from different scales. Finally, all features are aggregated according to the weights predicted from the LSTM for final classification. The overall architecture is trained in end-to-end manner with softmax loss for real face vs. spoofing face classification like [12]. An overview of our MS-FANS is shown in Fig. 2.

A. Feature Extraction with CNN

As mentioned above, the convolutional neural networks (CNN) is used for feature extraction. Compared with hand-crafted features which are designed empirically, CNN can automatically learn feature representation from data and thus has better modeling ability for more complex variations. Conventionally, the structure of a CNN consists of several convolutional layers for local feature representation, batch normalization layer after each convolutional layer for convergence acceleration, several fully connected layer for global representation, and dropout layer to increase the generalization of the network. Two types of exemplar CNN structure used in this work are shown in the Table I.

For an input image I, the faces are cropped at multiple scales $s_i (i = 1, 2, \ldots, n)$, denoted as $x_{s_i} (i = 1, 2, \ldots, n)$. Formally, for each face crop the feature extracted from the CNN is described as below:

$$f_{s_i} = CNN(x_{s_i}), \qquad i = 1, 2, \ldots, n \tag{1}$$

where n is the number of scales, i is the index of scale, s_i is the i^{th} scale, x_{s_i} is face region cropped at i^{th} scale, and f_{s_i} is the CNN feature for x_{s_i}. As exemplar with multiple scales is shown in Fig. 3.

TABLE I

Details of two networks structure used in this paper, a shallow one CNN_S and a deep one CNN_D. conv(w, s, N) denotes a convolution layer which has N filters of size $w \times w$ with stride s; pool(w, s) is a $w \times w$ max-pooling layer with stride s; fc(N) is a fully connected layer with N neurons. All the convolution layers followed by a batch normalization layer. ReLU is used as the non-linear activation function.

	shallow network(CNN_S)	deep network(CNN_D)
Input	112×112×3	
Conv-1	conv(3,2,48)	conv(3,1,32)
Pooling-1	pool(3,2)	pool(3,2)
Conv-2	conv(3,2,96)	conv(3,1,64)
Pooling-2	pool(3,2)	pool(3,2)
Conv-3	-	conv(3,1,128)
Pooling-3	-	pool(3,2)
Conv-4	-	conv(3,1,256)
Pooling-4	-	pool(3,2)
FC-1	fc(1024)	fc(2048)
FC-2	-	fc(1024)

B. Multi-scale Feature Fusion with LSTM

Features from different scales usually play different roles, and therefore an LSTM is exploited to fuse multi-scale features by adaptively predicting the fusion weights. Another straightforward way of employing multi-scale information is concatenating or averaging them. The strategy of concatenation or average considers all scales equally, which however are usually not true. So we believe that our strategy of adaptive fusion can well take advantage of multi-scale information by revealing the different importance of each scale.

The features from different scales are related to each other. To effective model the relationship, following [12] the LSTM is used to analyze the different importance of each scale, i.e. generate the weights for each scale. Specifically, the features of different scale extracted from CNN are sequentially input into the LSTM, and the LSTM sequentially outputs the weight for each scale as follows (which is also shown in Fig. 2):

$$[\alpha_{s_i}, h_i] = LSTM(f_{s_i}, h_{i-1}). \tag{2}$$

Having the weight for each scale, the multi-scale features are fused according to the following Eq. 3 to obtain the final features for classification:

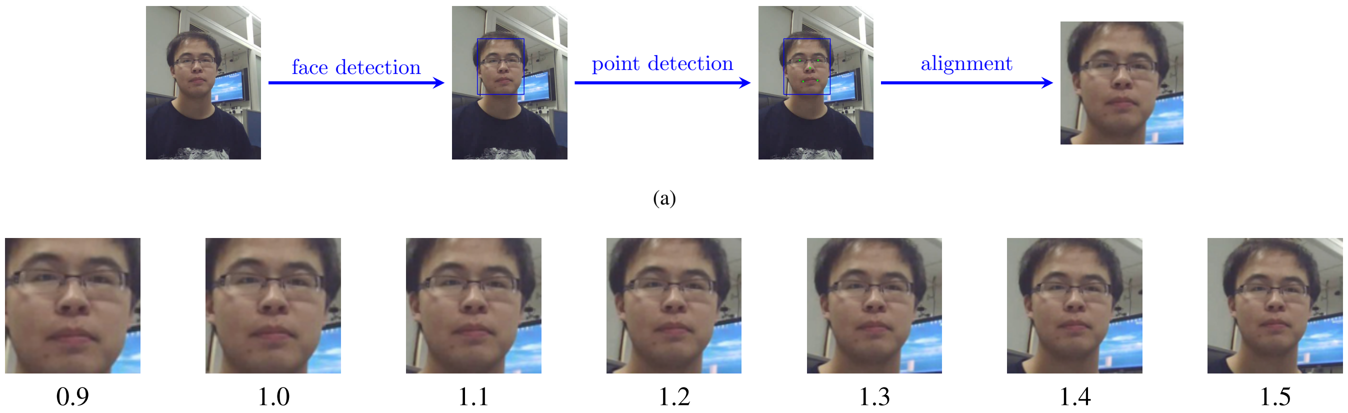
$$f = \sum_{i=1}^{n} \alpha_{s_i} f_{s_i}, \tag{3}$$

where α_{s_i} is the weight of f_{s_i} generated from the LSTM and f is the fused feature for final real vs. spoofing classification, h_i is the hidden state of LSTM model.

C. End-to-End Training

With the fused feature f, the final classification is model as a two-class classification problem by using the softmax loss as below:

$$L_f = softmax(f). \tag{4}$$



(a)

(b)

Fig. 3. Illustration of (a) the process of face cropping of each video frame including face detection, five points detection and face alignment; (b) face crops in different scales.

In order to obtain more distinguishable features from CNN, an auxiliary loss of real vs. spoofing classification is imposed on the CNN features of each scale during training, which is also formulated as softmax loss as follows:

$$L_{f_i} = softmax(f_{si}). \tag{5}$$

As a result, there are n auxiliary losses during the initial training phase. Overall, the loss of the whole network $\mathbb{L}$ is formulated as below:

$$\mathbb{L} = \begin{cases} L_f + 0.5 \sum\limits_{i=1}^{n} L_{f_{s_i}}, & iter < \frac{MAXITER}{2} \\ L_f, & iter > \frac{MAXITER}{2} \end{cases} \tag{6}$$

where $L_{f_{s_i}}$ is the auxiliary loss for the feature f_{s_i} from i^{th} scale, L_f is the loss for the integrated multi-scale feature f, and $MAXITER$ is the maximum number of iterations during training. As shown in Eq. 6, introducing the auxiliary loss in the initial training phase can guide the CNN to learn more distinguishable features and reduce the optimization difficulty, while removing auxiliary loss in the late training phase can ensure the whole network focus on optimizing multi-scale feature fusion. The whole MS-FANS network can be optimized end-to-end by using gradient descent as most deep network do.

IV. Experiment

To assess of the effectiveness of our proposed multi-scale face anti-spoofing method, MS-FANS, a series of experiments are performed on two widely used datasets CASIA-FASD [13] and Idiap REPLAY-ATTACK [1]. First, we analyze the effectiveness of the adaptive multi-scale information fusion by comparing it with the single-scale baseline and straightforward fusion. Then, we compare our method with the state-of-the-art methods.

A. Datasets and Pre-processing

The datasets of CASIA-FASD [13] and Idiap REPLAY-ATTACK [1] are two commonly used datasets for face anti-spoofing, mainly including print photos and video attacks.

a) CASIA-FASD : In CASIA-FASD dataset there are 50 subjects, and each subject contains 12 videos with 3 different image quality: low quality, normal quality and high quality. For attack videos in each quality, there are three different attacks including warped photo attack, cut photo attack and video replay attack. In the standard protocol, the dataset is divided into train set and test sets, consisting of 20 and 30 subjects respectively.

b) Idiap: The Idiap REPLAY-ATTACK dataset contains 50 subjects with a total of 1300 videos. The dataset is divided into train, development and test sets, including 15, 15 and 20 subjects respectively. The videos are in two kinds of constraints and adverse light conditions. The attack videos contain two types with the play device handheld or supported.

For both datasets, a preprocessing is conducted to obtain face crops at different scales. All videos are decoded into frames, and each frame is pre-processed separately. Firstly, face detection [23] and points detection [24] are conducted to get five facial landmarks including two eyes centers, one nose tip, and two mouth corners. Then, according to the five points each face is aligned to a canonical one. Finally, face crops of 128 × 128 at different scales are obtained by cutting out the face region from the aligned frame. The process is shown in Fig. 3a. For each frame, 7 scales are used as the baseline scale, including 0.9, 1.0, 1.1, 1.2, 1.3, 1.4, 1.5 as shown in Fig. 3b. On CASIA-FASD datset, the performance is evaluated in terms of Equal Error Rate (EER), and on Idiap REPLAY-ATTACK dataset the performance is evaluated in terms of Equal Error Rate (EER) and Half Total Error Rate (HTER) following most of the existing methods.

For the optimization of our MS-FANS, Pytorch toolkit [25] is employed which is a platform that uses dynamic graphs for neural network training. The stochastic gradient descent (SGD) optimizer is used, the momentum is set as 0.9, weight decay as 0.0005, and learning rate as 0.0002. The multi-steps strategy is chosen to update the learning rate, that is, reduce the learning rate as 0.1 of the original one after a certain number of iterations. For each input image, 112 × 112 random croped

TABLE II
THE PERFORMANCE OF EACH SCALE ON CASIA-FASD DATASET. FOR ALL SCALES, THE ARCHITECTURE OF THE NETWORK IS THE SAME, I.E. CNN_S DESIGNED IN TABLE I, AND ONLY THE SCALE OF THE INPUT IS DIFFERENT.

Scale	0.9	1.0	1.1	1.2	1.3	1.4	1.5
EER(%)	5.93	6.09	5.26	4.67	4.84	4.74	5.55

TABLE III
PERFORMANCE OF DIFFERENT FUSION STRATEGIES IN TERMS OF EER(%) ON CASIA-FASD DATASET. THE RESULTS FROM BEST SINGLE SCALE, AVERAGING THE MULTI-SCALE INFORMATION, CONCATENATING THE MULTI-SCALE INFORMATION, AND OUR ADAPTIVELY FUSING THE MULTI-SCALE INFORMATION ARE REPORTED RESPECTIVELY.

Methods	Scale				
	[0.9,1.0,1.1]	[1.0,1.1,1.2]	[1.1,1.2,1.3]	[1.2,1.3,1.4]	[1.3,1.4,1.5]
Best Single Scale	5.26	4.67	4.67	4.67	4.74
Average	4.60	**3.79**	3.74	3.44	4.16
Concat	4.81	3.95	3.57	3.68	3.65
MS-FANS	**4.18**	3.89	**3.15**	**3.30**	**3.29**

region is used as input.

B. Ablation Analysis of Multi-scale Fusion

Firstly, we evaluate the performance of each scale to investigate the effect of scale on performance. Here, the architecture of the CNN is designed as the shallow network shown in Table I, i.e. CNN_S, and the softmax is directly used for the classification of real face and spoofing. For all scales, the architecture of the CNN is the same, but only the scale of the input is different. The performance is evaluated on the CASIA-FASD dataset and the results are showed in Table II.

As shown in Table II, the performance of different scales are much different from each other, with the best EER up to 4.67% and worst down to 6.09%. This demonstrates that different scales play different roles for the distinguishment of real faces from the spoofing faces, and it is hard to determine an optimal scale to obtain a best result. Besides, it can be observed that the performance first rises and then fall as the scale becomes larger, which illustrating that the including a certain amount of background can benefit the performance, but too much of them might degrade the performance. The two observations enlighten us naturally that it would be a better solution to fuse the multi-scale information and adaptively determine the importance of each scale.

Furthermore, we investigate the performance of different strategies for fusing the multi-scale information on CASIA-FASD dataset. For all methods, the shallow CNN_S is used as the base architecture, and for our MS-FANS an additional LSTM with only 1 hidden node is stacked to predict the fusion weight with very low cost. For the overall 7 scales, three adjacent scales is taken as a group, ie. $[0.9, 1.0, 1.1]$, $[1.0, 1.1, 1.2]$, $[1.1, 1.2, 1.3]$, $[1.2, 1.3, 1.4]$, $[1.3, 1.4, 1.5]$. For the feature fusion, several commonly used strategies are evaluated in Table III: taking the best performance of all scales denoted as "Best Single Scale", averaging the multi-scale information f_{s_i} denoted as "Average", concatenating the multi-scale information f_{s_i} as a long feature denoted as "Concat", and our adaptively fusing the multi-scale information f_{s_i} denoted as "MS-FANS".

The results are shown in Table III. As can be seen, all of the fusing strategy including average, concatenation, and our adaptive fusion perform better than the best scale, which indicates that fusing multi-scale information can effectively benefit the anti-spoofing. Moreover, our MS-FANS performs the best in most cases and significantly reduces the EER even up to 32.5%(from 4.67% to 3.15%), demonstrating the effectiveness of our adaptive fusion of multi-scale information.

C. Comparison with the Existing Methods

Finally, we compare our MS-FANS with the state-of-the-art methods on CASIA-FASD and Idiap REPLAY-ATTACK datasets. Here, a deeper architecture CNN_D shown in Table I is used for MS-FANS for fair comparison as most of the existing methods employ a larger network than ours. On both datasets, we follow the standard protocol as the existing methods do, and the results of the existing methods are directly copied from the original works. For our MS-FANS, we only report the results of the best scale group, i.e. $[1.1, 1.2, 1.3]$ on CASIA-FASD and $[1.3, 1.4, 1.5]$ on Idiap REPLAY-ATTACK, although the results of different scale group is slightly different.

The results on CASIA-FASD dataset is shown in Table IV. As can be seen, our MS-FANS with deep architecture outperforms all the other methods and even our MS-FANS with shallow architecture performs better than most of the deep CNN based methods. Besides, the computation complexity in terms of FLOPS of our MS-FANS with both architecture is much smaller than the existing methods. The results demonstrate that our MS-FANS is an effective method for adaptively exploring the multi-scale information, leading to better performance even with much smaller computation complexity.

The results on Idiap REPLAY-ATTACK dataset is shown in Table V. From the results, the same conclusion can be obtained that our adaptive fusion of multi-scale information can better distinguish the real faces from the spoofing faces. Another observation is that on this dataset our MS-FANS with deep architecture CNN_D performs worse than that with shallow CNN_S. This probably because that this dataset is simpler and the deep architecture is easily to overfit considering the overall performance on this dataset is better than that on Idiap REPLAY-ATTACK dataset. Our MS-FANS with shallow network architecture achieves 0.002% EER which almost does the anti-spoofing perfectly.

V. CONCLUSION AND FUTURE WORK

This paper introduces an approach of adaptive fusion of multi-scale information for face anti-spoofing, which performs much better than the single-scale methods and most existing deep methods. In our proposed MS-FANS, a CNN used for extracting the features of different scale input and an LSTM

TABLE IV
COMPARISON WITH THE EXISTING METHODS ON CASIA-FASD DATASET IN TERMS OF EER. THE SYMBOL † REFERS TO CNN-BASED METHODS. FLOPS IS THE NUMBER OF MULTIPLY AND ADD IN THE FORWARD PROCESSING. DUE TO THE LOW COMPUTATIONAL COMPLEXITY OF HAND-CRAFTED FEATURES, WE ONLY COMPARE FLOPS OF CNN-BASED METHODS. FOR MS-FANS, WE USE THE SUM FLOPS OF THREE SCALES.

Method	EER(%)	FLOPS
†Fine-tuned VGG-Face [10]	5.20	$30.919G$
†DPCNN [10]	4.50	$30.919G$
†Yang et al. [9]	4.92	$0.349G$
†LSTM-CNN [12]	5.17	$3.045G$
Boulkenafet et al. [6]	6.20	-
Siddiqui et al. [14]	3.14	-
Boulkenafet et al. [16]	2.8	-
†Atoum et al. [11]	2.67	$42.045G$
$[1.1, 1.2, 1.3]_{CNN_S}$(Ours)	3.15	$0.075G$
$[1.1, 1.2, 1.3]_{CNN_D}$(Ours)	**2.40**	$0.768G$

TABLE V
COMPARISON WITH THE EXISTING METHODS ON IDIAP REPLAY-ATTACK DATASET IN TERMS OF EER AND HTER. THE SYMBOL † REFERS TO CNN-BASED METHODS. FLOPS IS THE NUMBER OF MULTIPLY AND ADD IN FORWARD PROCESSING. WE ONLY COMPARE FLOPS OF CNN-BASED METHODS.

Method	EER(%)	HTER(%)	FLOPS
†Fine-tuned VGG-Face [10]	8.40	4.30	$30.919G$
†DPCNN [10]	2.90	6.10	$30.919G$
†Yang et al. [9]	2.14	-	$0.349G$
Boulkenafet et al. [6]	0.40	2.90	-
Boulkenafet et al. [16]	0.10	2.20	-
†Atoum et al. [11]	0.79	0.72	$41.825G$
$[1.3, 1.4, 1.5]_{CNN_S}$(Ours)	**0.002**	**0.24**	$0.075G$
$[1.3, 1.4, 1.5]_{CNN_D}$(Ours)	**0.02**	**0.39**	$0.768G$

used for adaptive fusion of multi-scale features is formulated as an end-to-end pipeline. Experimental results on two datasets show that our method achieves state-of-the-art performance, demonstrating the effectiveness of our proposed method for face anti-spoofing. In this work, the scale group is manually determined, and in future we will explore to predict the scale group automatically. Moreover, how to fuse multi-scale information for video face anti-spoofing is also an interesting future work.

ACKNOWLEDGMENT

This work was partially supported by National Key R&D Program of China under contracts No.2017YFA0700804 and Natural Science Foundation of China under contracts Nos. 61650202, 61772496, 61402443, and 61532018.

REFERENCES

[1] I. Chingovska, A. Anjos, and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *International Conference of Biometrics Special Interest Group (BIOSIG)*, 2012, pp. 1–7.

[2] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "LBP–TOP based countermeasure against face spoofing attacks," in *Asian Conference on Computer Vision Workshop (ACCVW)*, 2013, pp. 121–132.

[3] J. Mtt, A. Hadid, and M. Pietikinen, "Face spoofing detection from single images using micro-texture analysis," in *The IEEE International Joint Conference on Biometrics (IJCB)*, 2011, pp. 1–7.

[4] T. de Freitas Pereira, A. Anjos, J. M. D. Martino, and S. Marcel, "Can face anti-spoofing countermeasures work in a real world scenario?" in *The IEEE International Conference on Biometrics (ICB)*, 2013, pp. 1–8.

[5] T. d. Freitas Pereira, J. Komulainen, A. Anjos, J. M. De Martino, A. Hadid, M. Pietikäinen, and S. Marcel, "Face liveness detection using dynamic texture," *EURASIP Journal on Image and Video Processing (IVP)*, vol. 2014, no. 1, p. 2, 2014.

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

[7] K. Kollreider, H. Fronthaler, M. I. Faraj, and J. Bigun, "Real-time face detection and motion analysis with application in "liveness" assessment," *The IEEE Transactions on Information Forensics and Security (TIFS)*, vol. 2, no. 3, pp. 548–558, 2007.

[8] L. Feng, L.-M. Po, Y. Li, X. Xu, F. Yuan, T. C.-H. Cheung, and K.-W. Cheung, "Integration of image quality and motion cues for face anti-spoofing: A neural network approach," *Journal of Visual Communication and Image Representation (VCIR)*, vol. 38, pp. 451–460, 2016.

[9] J. Yang, Z. Lei, and S. Z. Li, "Learn convolutional neural network for face anti-spoofing," *CoRR*, vol. abs/1408.5601, 2014.

[10] L. Li, X. Feng, Z. Boulkenafet, Z. Xia, M. Li, and A. Hadid, "An original face anti-spoofing approach using partial convolutional neural network," in *International Conference on Image Processing Theory, Tools and Applications (IPTA)*, 2016, pp. 1–6.

[11] Y. Atoum, Y. Liu, A. Jourabloo, and X. Liu, "Face anti-spoofing using patch and depth-based cnns," in *The IEEE International Joint Conference on Biometrics (IJCB)*, 2017, pp. 319–328.

[12] Z. Xu, S. Li, and W. Deng, "Learning temporal features using lstm-cnn architecture for face anti-spoofing," in *Asian Conference on Pattern Recognition (ACPR)*, 2015, pp. 141–145.

[13] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," in *The IEEE International Conference on Biometrics (ICB)*, 2012, pp. 26–31.

[14] T. A. Siddiqui, S. Bharadwaj, T. I. Dhamecha, A. Agarwal, M. Vatsa, R. Singh, and N. Ratha, "Face anti-spoofing with multifeature videolet aggregation," in *International Conference on Pattern Recognition (ICPR)*, 2016, pp. 1035–1040.

[15] J. Komulainen, A. Hadid, and M. Pietikinen, "Context based face anti-spoofing," in *The IEEE International Conference on Biometrics: Theory, Applications and Systems (BTAS)*, 2013, pp. 1–8.

[16] Z. Boulkenafet, J. Komulainen, and A. Hadid, "Face antispoofing using speeded-up robust features and fisher vector encoding," *The IEEE Signal Processing Letters (LSP)*, vol. 24, no. 2, pp. 141–145, 2017.

[17] W. Bao, H. Li, N. Li, and W. Jiang, "A liveness detection method for face recognition based on optical flow field," in *International Conference on Image Analysis and Signal Processing (IASP)*, 2009, pp. 233–236.

[18] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *The IEEE Transactions on Image Processing (TIP)*, vol. 23, no. 2, pp. 710–724, 2014.

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

[20] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," in *The IEEE International Conference on Computer Vision (ICCV)*, 2007, pp. 1–8.

[21] L. Sun, G. Pan, Z. Wu, and S. Lao, "Blinking-based live face detection using conditional random fields," in *The IEEE International Conference on Biometrics (ICB)*, 2007, pp. 252–260.

[22] Y. Wang, F. Nian, T. Li, Z. Meng, and K. Wang, "Robust face anti-spoofing with depth information," *Journal of Visual Communication and Image Representation (VCIR)*, vol. 49, pp. 332–337, 2017.

[23] S. Wu, M. Kan, Z. He, S. Shan, and X. Chen, "Funnel-structured cascade for multi-view face detection with alignment-awareness," *Neurocomputing*, vol. 221, pp. 138–145, 2017.

[24] J. Zhang, S. Shan, M. Kan, and X. Chen, "Coarse-to-fine auto-encoder networks (cfan) for real-time face alignment," in *European Conference on Computer Vision (ECCV)*, 2014, pp. 1–16.

[25] Pytorch. [Online]. Available: http://pytorch.org/