00

Title

Face Anti-Spoofing Using Spatial Pyramid Pooling

*Lei Shi, Zhuo Zhou, Zhenhua Guo**

Tsinghua University

ICPR2020

Reporter: Lei Shi

January 2021

CONTENT

01 Motivation

As an important application of computer vision, face recognition has been widely used in many fields. However, the system is also vulnerable to many kinds of presentation attacks. So, how to effectively detect whether the image is from the real face is very important. At present, many deep learning anti-spoofing methods have been proposed, But they have some limitations. For example, global average pooling (GAP) easily loses local information, single-scale features easily ignore information differences in different scales, while a complex network is prone to be overfitting.

Fig. 1. Kids break through the express cabinet with one photo.

Pretrained-Resnet18 + SPP

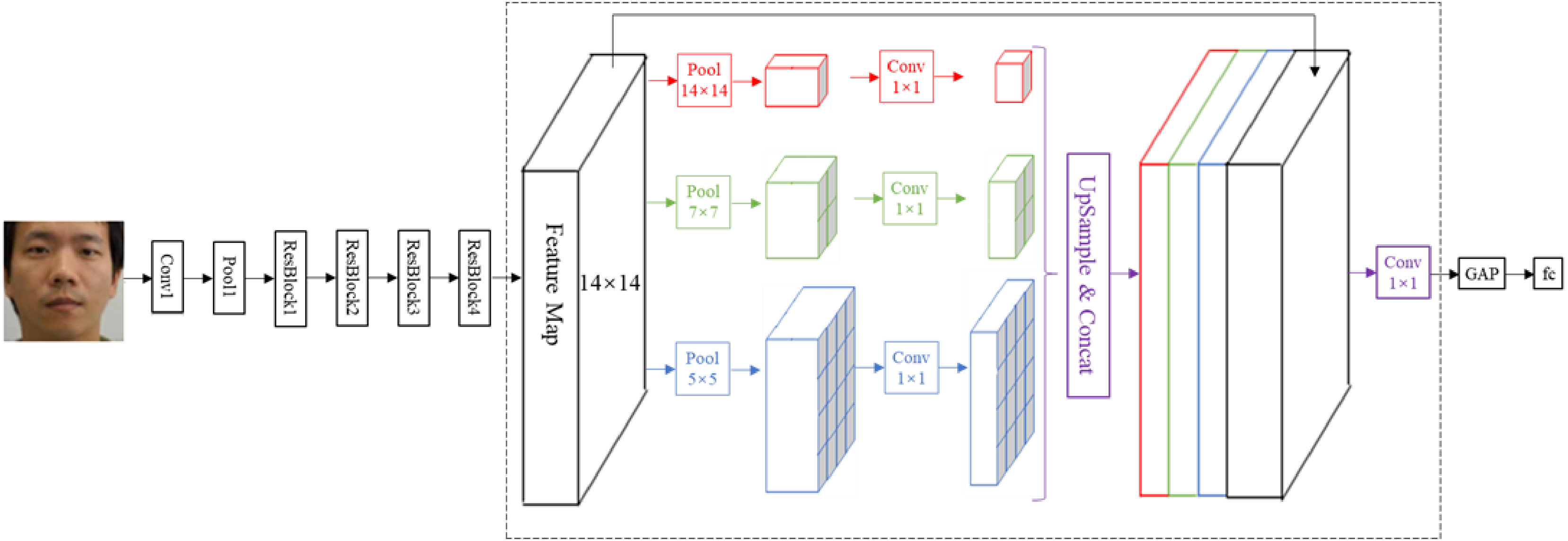

Fig. 2. The proposed network with SPP module. The dotted box is the utilized SPP module.

Resnet-18

Here, ResNet-18 pre-trained on ImageNet is adopted, and the last fully connected layer has 2 neurons. By this setting, the relationship between the feature map and output categories is established, so that the feature map can be regarded as confidence map. For the fully connected layer parameters, the *Kaiming* initialize method is used. Then we fine-tune this network and use it as the baseline model.

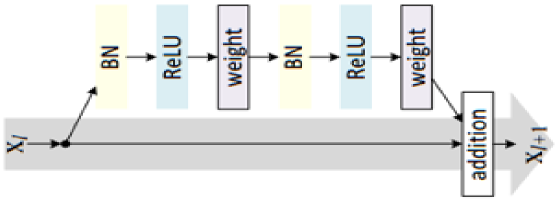

Fig. 2. Improved ResNet residual unit [19].

Table I. Network Structure of ResNet-18

Feature layer	**Output size of feature layer**	**Specific parameters**
Input	224×224×3	-
Conv1	112×112×64	7×7, 64, stride 2
Pool1	56×56×64	3×3, max pool, stride 2
ResBlock1	56×56×64	$\begin{bmatrix} 3\times3, & 64 \\ 3\times3, & 64 \end{bmatrix} \times 2$
ResBlock2	28×28×128	$\begin{bmatrix} 3\times3, & 128 \\ 3\times3, & 128 \end{bmatrix} \times 2$
ResBlock3	14×14×256	$\begin{bmatrix} 3\times3, & 256 \\ 3\times3, & 256 \end{bmatrix} \times 2$
ResBlock4	14×14×512	$\begin{bmatrix} 3\times3, & 512 \\ 3\times3, & 512 \end{bmatrix} \times 2$
GAP	1×1×512	-
fc	1×1×2	512×2

Spatial Pyramid Pooling

Further, this figure shows the structure of the SPP, where feature map is output of the last convolution layer of ResBlock4, and the size is 14×14 if the input is 224×224. The workflow of SPP is described below.

First, the feature layers are pooled at different scales. Second, feature dimensions are reduced by 1×1 convolution kernel, and then are upsampled to the size of 14×14. After that, the features of different scales are concatenated with the original feature. Last, convolution of 1×1 is adopted to fuse the information of each scale.

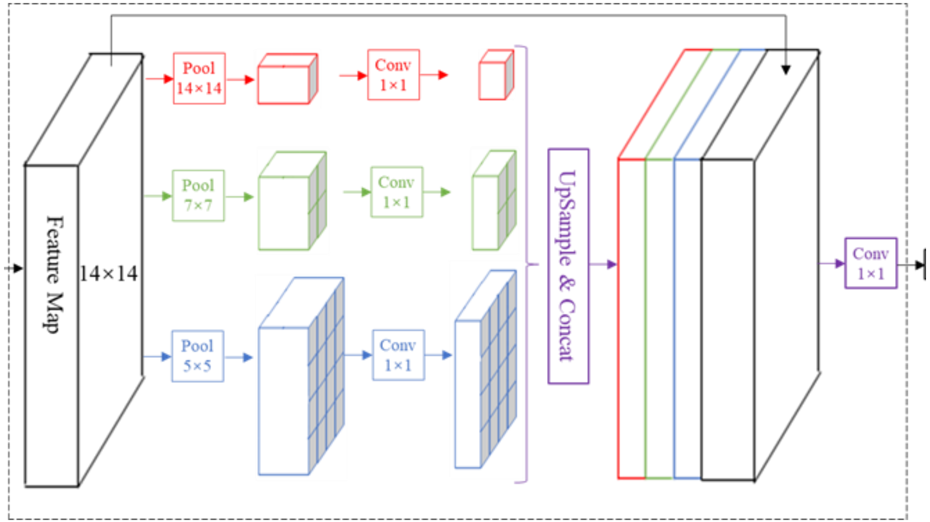

Fig. 3. Structure of Spatial Pyramid Pooling

Results

First, baseline's performance shows that fine-tuning an proper pre-trained model is effective. Second, proposed method achieves the best result on CASIA-FASD and good results on Replay-Attack and CASIA-SURF. Compared with other methods, the proposed one is single-frame based, which is simple, easy to train, fast in application and robust.

Table II&III. Test Results on CASIA-FASD, Replay-Attack and CASIA-SURF (RGB Only)

Methods	CASIA-FASD	Replay-Attack	
	EER(%)	*EER(%)*	*HTER(%)*
Fine-tune VGG-Face [11][a]	5.2	8.4	4.3
DPCNN [11][a]	4.5	2.9	6.1
COLOR LBP [26]	6.2	0.4	2.9
Boulkenafet et al. [5]	3.2	0.0	3.5
Boulkenafet et al. [6]	2.8	0.1	2.2
Boulkenafet et al. [31]	1.5	1.2	4.2
Li et al. [13][a]	2.3	0.1	0.9
Li et al. [15][a]	2.2	0.5	1.6
Atoum et al. [14]	2.67	0.79	0.72
Feng et al. [32]	5.83	0.83	**0.00**
Chen et al. [29][b]	2.359	0.062	0.183
FaceBagNet [33, 28][c]	5.56	3.33	1.92
Baseline (Ours)[b]	1.86	0.00	0.63
Proposed method (SPP$_{5\text{-}7\text{-}14}$)[b]	**0.37**	**0.00**	0.50

a. Fine-tuning pre-trained VGG-face.
b. Pre-trained on ImageNet.
c. Results achieved according to [28].

Methods	TPR(%)@FPR			ACER(%)	Test Time(ms)
	10^{-2}	*10^{-3}*	*10^{-4}*		
Zhang et al. [24][e]	49.3	16.6	6.8	11.3	-
Kuang et al. [34][d]	49.3	18.3	6.83	11.8	-
Wang et al. [35][ae]	89.5	69.5	39.8	3.9	-
Parkin et al. [36][bd]	71.74	22.34	7.85	-	-
FaceBagNet [33, 28][ef]	88.4	64.7	39.3	4.1	12.2
Baseline(Ours)[ce]	68.3	33.5	15.8	8.1	**2.9**
Proposed method (SPP$_{4\text{-}8}$)[ce]	72.8	34.7	11.6	6.4	5.5
Fusion[ceg]	**93.0**	**75.9**	**45.3**	**3.0**	17.7

a. Pre-trained on three face anti-spoofing databases (CASIA-FASD, MSU-MFSD, Replay-Attack).
b. Pre-trained on CASIA-Web face [38].
c. Pre-trained on ImageNet.
d. Results on validation set.
e. Results on test set.
f. Results achieved according to [28].
g. Results computed by (0.4*SPP$_{4\text{-}8}$+0.6*FaceBagNet).

Ablation Study

We conducted experiments using each single-scale pooling, which is a simple ablation study. Baseline and three single-scale pooling models have their own performances, and the results of $SPP_{5\text{-}7\text{-}14}$ is better than all of them. That means, different pooling extract different scale features, and $SPP_{5\text{-}7\text{-}14}$ fuses them effectively to get better results. We draw a conclusion from the experiment that multi-scale information is crucial for face anti-spoofing, and the proposed method is effective.

Table IV. Ablation Study of Scales

Pooling Scales	**CASIA-FASD**	**Replay-Attack**	
	EER(%)	*EER(%)*	*HTER(%)*
Baseline	1.86	0.00	0.63
5×5 Single pooling	2.23	0.00	1.40
7×7 Single pooling	1.49	0.00	0.76
14×14 Single pooling	0.74	0.00	0.63
$SPP_{5\text{-}7\text{-}14}$	**0.37**	**0.00**	**0.50**

Feature Visualization

The network pays much attention on the areas of eyes, nose and mouth. Possible reason is that the real face has strong depth information and detailed information in these areas, while the fake face loses such information after being imaged twice. Therefore, enhancing these local features is necessary.

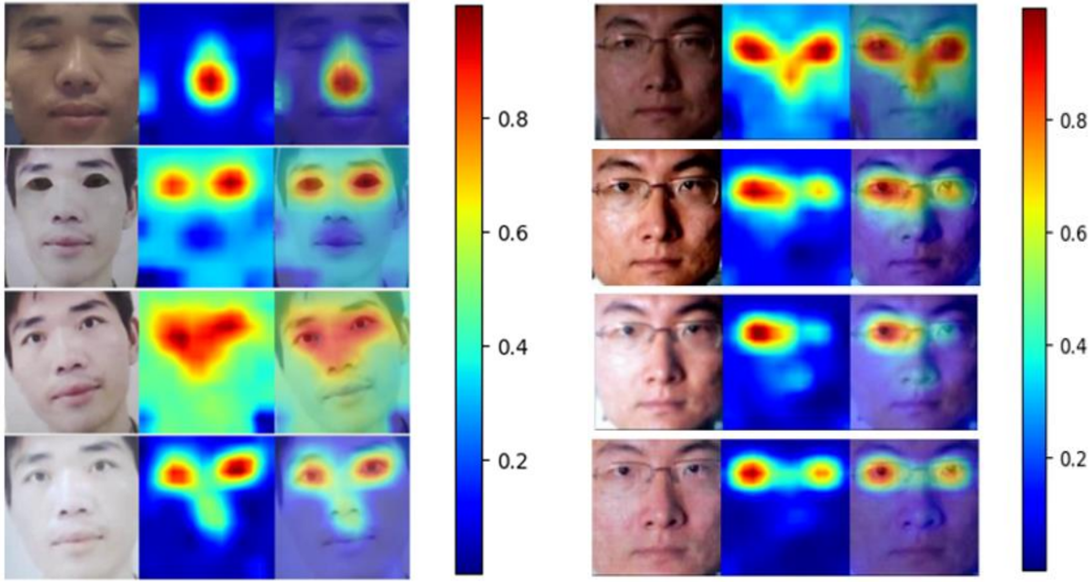

Fig. 4. Network Classification Visualization.

Reference

1. S. Chakraborty and D. Dhrubajyoti, "An overview of face liveness detection," arXiv preprint arXiv:1405.2227, 2014.
2. Y. Atoum, Y. Liu, A. Jourabloo and X. Liu, "Face anti-spoofing using patch and depth-based CNNs," in *IEEE International Joint Conference on Biometrics*, 2017, pp. 319-328.
3. J. Li, Y. Wang, T. Tan and A. K. Jain, "Live face detection based on the analysis of Fourier spectra," *Biometric Technology for Human Identification*, vol. 5404, pp. 296-303, International Society for Optics and Photonics, 2004.
4. J. Määttä, A. Hadid and M. Pietikäinen, "Face spoofing detection from single images using micro-texture analysis," in *IEEE International Joint Conference on Biometrics*, 2011, pp. 1-7.
5. Z. Boulkenafet, J. Komulainen and A. Hadid, "Face spoofing detection using colour texture analysis," *IEEE Transactions on Information Forensics and Security*, vol. 11, no. 8, pp. 1818-1830, 2016.
6. Z. Boulkenafet, J. Komulainen and A. Hadid, "Face antispoofing using speeded-up robust features and fisher vector encoding," *IEEE Signal Processing Letters*, vol. 24, no. 2, pp. 141-145, 2017.
7. F. Perronnin, J. Sánchez, and T. Mensink, "Improving the fisher kernel for large-scale image classification," in *European conference on computer vision*, 2010, pp. 143-156.
8. D. Wen, H. Han and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 746-761, 2015.
9. J. Yang, Z. Lei and S. Z. Li, "Learn convolutional neural network for face anti-spoofing," *arXiv preprint arXiv:1408.5601*, 2014.
10. A. Krizhevsky, I. Sutskever and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Advances in neural information processing systems*, 2012, pp. 1097-1105.
11. 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 *Sixth International Conference on Image Processing Theory, Tools and Applications*, 2016, pp. 1-6.
12. O. M. Parkhi, A. Vedaldi and A. Zisserman, "Deep face recognition," 2015.
13. L. Li and X. Feng, "Face anti-spoofing via deep local binary pattern," in *Deep Learning in Object Detection and Recognition*, 2019, pp. 91-111.
14. Y. Atoum, Y. Liu, A. Jourabloo and X. Liu, "Face anti-spoofing using patch and depth-based CNNs," in *IEEE International Joint Conference on Biometrics*, 2017, pp. 319-328.
15. L. Li, Z. Xia, L. Li, X. Jiang, X. Feng and F. Roli, "Face anti-spoofing via hybrid convolutional neural network," in *International Conference on the Frontiers and Advances in Data Science*, 2017, pp. 120-124.
16. A. Jourabloo, Y. Liu and X. Liu, "Face de-spoofing: Anti-spoofing via noise modeling," in *Proceedings of the European Conference on Computer Vision*, 2018, pp. 290-306.
17. K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.
18. K. He, X. Zhang, S. Ren and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770-778.
19. K. He, X. Zhang, S. Ren and J. Sun, "Identity mappings in deep residual networks," in *European conference on computer vision*, 2016, pp. 630-645.
20. K. He, X. Zhang, S. Ren and J. Sun, "Delving deep into rectifiers: Surpassing human-level performance on imagenet classification," in *Proceedings of the IEEE international conference on computer vision*, 2015, pp. 1026-1034

Reference

21. K. He, X. Zhang, S. Ren and J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, vol. 37, no. 9, pp. 1904-1916, 2015.
22. Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi and S. Z. Li, "A face antispoofing database with diverse attacks," in *5th IAPR international conference on Biometrics*, 2012, pp. 26-31.
23. I. Chingovska, A. Anjos and S. Marcel, "On the effectiveness of local binary patterns in face anti-spoofing," in *BIOSIG-proceedings of the international conference of biometrics special interest group*, 2012, pp. 1-7.
24. S. Zhang, X. Wang, A. Liu, C. Zhao, J. Wan, S. Escalera et al.,"A dataset and benchmark for large-scale multi-modal face anti-spoofing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 919-928.
25. I. Loshchilov and F. Hutter, "Sgdr: Stochastic gradient descent with warm restarts," *arXiv preprint arXiv:1608.03983*, 2016.
26. 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.
27. B. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, "Learning deep features for discriminative localization," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 2921-2929.
28. Project *webpages, https://github.com/SeuTao/CVPR19-Face-Anti-spoofing.*
29. H. Chen, Y. Chen, X. Tian, and R. Jiang, "A cascade face spoofing detector based on face anti-spoofing R-CNN and improved Retinex LBP," *IEEE Access*, vol. 7, pp. 170116-170133, 2019.
30. T. Kim, Y. Kim, I. Kim and D. Kim, "BASN: Enriching feature representation using bipartite auxiliary supervisions for face anti-spoofing," in *Proceedings of the IEEE International Conference on Computer Vision Workshops*, 2019, pp. 0-0.
31. Z. Boulkenafet, J. Komulainen and A. Hadid, "On the generalization of color texture-based face anti-spoofing," *Image and Vision Computing*, vol. 77, pp. 1-9, 2018.
32. L. Feng, L. Po et al., "Integration of image quality and motion cues for face anti-spoofing: A neural network approach," *Journal of Visual Communication and Image Representation* vol. 38, pp. 451-460, 2016.
33. T. Shen, Y. Huang and Z. Tong, "FaceBagNet: Bag-of-local-features model for multi-modal face anti-spoofing," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0-0.
34. H. Kuang, R. Ji et al., "Multi-modal multi-layer fusion network with average binary center loss for face anti-spoofing," in *Proceedings of the 27th ACM International Conference on Multimedia*, 2019, pp. 48-56.
35. G. Wang, C. Lan, H. Han, S. Shan and X. Chen, "Multi-modal face presentation attack detection via spatial and channel attentions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0-0.
36. A. Parkin and O. Grinchuk, "Recognizing multi-modal face spoofing with face recognition networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, 2019, pp. 0-0.
37. D. Wen, H. Han and A. K. Jain, "Face spoof detection with image distortion analysis," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 4, pp. 746-761, 2015.
38. D. Yi, Z. Lei, S. Liao and S. Z. Li, "Learning face representation from scratch," *arXiv preprint arXiv:1411.7923*, 2014.
39. J. Deng, W. Dong et al., "Imagenet: A large-scale hierarchical image database," *in IEEE conference on computer vision and pattern recognition*, 2009, pp. 248-255.

THANKS

Face Anti-Spoofing Using Spatial Pyramid Pooling

*Lei Shi, Zhuo Zhou, Zhenhua Guo**

Tsinghua University