Face Anti-Spoofing Using Spatial Pyramid Pooling

*Lei Shi, Zhuo Zhou, Zhenhua Guo**

Tsinghua University

Abstract

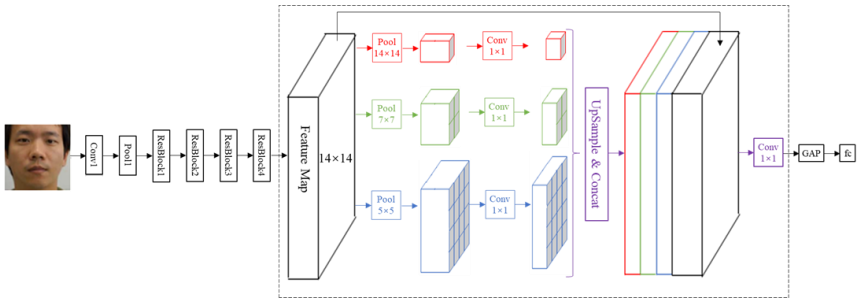
Face recognition system is vulnerable to many kinds of presentation attacks, so how to effectively detect whether the image is from the real face is particularly important. At present, Many anti-spoofing approaches have some limitations: global average pooling (GAP) loses local information of faces, single-scale features ignore information differences in different scales, a complex network is prone to be overfitting. In this paper, we propose a face anti-spoofing approach using spatial pyramid pooling (SPP). Firstly, we use ResNet-18 with a small amount of parameter as the basic model to avoid overfitting. Further, we use spatial pyramid pooling module in the single model to enhance local features while fusing multi-scale information. The effectiveness of the proposed method is evaluated on three databases, CASIA-FASD, Replay-Attack and CASIA-SURF. The experimental results show that the proposed approach can achieve state-of-the-art performance.

Results

First, baseline's performance shows that an appropriate pre-trained network model with fine-tuning is effective. Second, proposed method achieves the best result on CASIA-FASD and competitive 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.

Methods	CASIA-FASD	Replay-Attack	
	EER(%)	*EER(%)*	*HTER(%)*
VGG-Face	5.2	8.4	4.3
DPCNN	4.5	2.9	6.1
COLOR LBP	6.2	0.4	2.9
LBP&LPQ	3.2	0.0	3.5
SURF+Fisher	2.8	0.1	2.2
RI-LBP&SURF	1.5	1.2	4.2
LBP in CNN	2.3	0.1	0.9
Multi-CNNs	2.2	0.5	1.6
Patch+Depth	2.67	0.79	0.72
Quality&Motion	5.83	0.83	**0.00**
Faster R-CNN +Retinex	2.359	0.062	0.183
FaceBagNet	5.56	3.33	1.92
Baseline (Ours)	1.86	0.00	0.63
Proposed method	**0.37**	**0.00**	0.50

Method



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	$\begin{bmatrix} 3\times3, & 512 \\ 3\times3, & 512 \end{bmatrix}\times 2$
SPP	14×14×512	-
GAP	1×1×512	-
fc	1×1×2	512×2

Here, pre-activated ResNet-18 pre-trained on ImageNet is adopted, and the last fully connected layer is replaced with a new fully connected layer with 2 neurons. SPP extracts and fuses information of different positions and scales. Firstly, the feature layers are pooled at different scales to get features of different scales. Secondly, dimensions of the features are reduced by 1×1 convolution kernel, and then are up-sampled to the size of 14×14. After that, the features of different scales are concatenated with the original feature, and convolution of 1×1 is adopted to fuse the information of each scale. Finally, GAP is used.

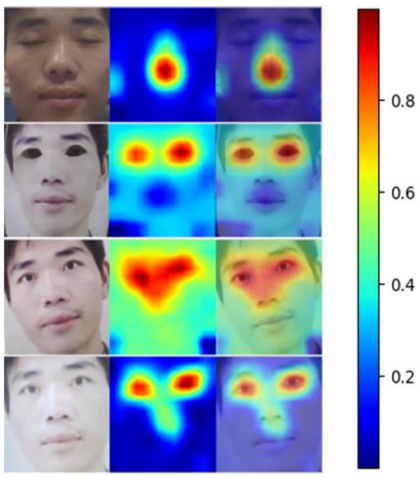
Ablation Study

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 single-scale networks. That means, different pooling operations 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.

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 mainly 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 seriously loses such information after being imaged twice. Therefore, enhancing these local features is necessary.



NSFC National Natural Science Foundation of China

ICPR 2020 25th International Conference on Pattern Recognition ONLINE 10 | 15 January 2021

IAPR Technically Co-Sponsored by IEEE IEEE COMPUTER SOCIETY