Face Anti-Spoofing via Disentangled Representation Learning (Supplementary Material)

Anonymous ECCV submission

Paper ID 3315

1 Overview

The contents of the supplementary material are as follows:

- Distributions of content feature and liveness feature.
- The role of low level supervision.
- Details of our network structure.

2 Distributions of Content Feature and Liveness Feature

We address face anti-spoofing via disentangled representation learning, which separates image representation into content feature and liveness feature, in which only the liveness feature is used for classification. In the experimental part Sec.4.2 of our submission, we demonstrate that liveness feature is not related to ID, background and light information. For further exploration, we show distributions of content feature and liveness feature by t-SNE in Fig. 1, where 500 live face images and $2,000$ spoof images from Oulu prototol1 are chosen for extracting features. In content space, the liveness samples are mixed together with attack samples. However they are separated and clustered into two groups in liveness space, which proves that the image features are indeed divided into two parts, one related to face anti-spoofing and the other irrelevant.

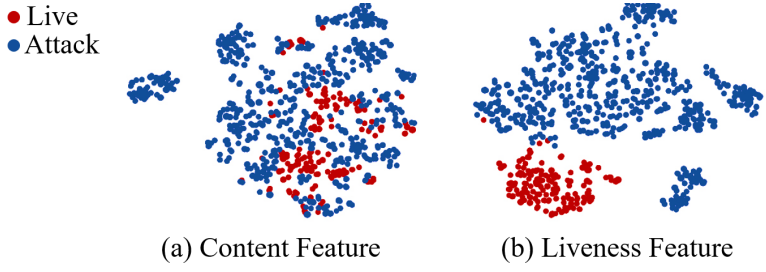
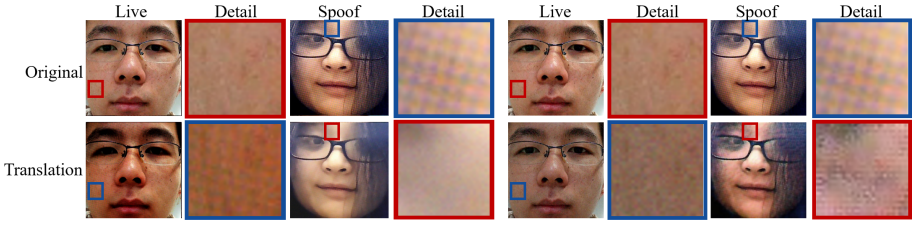


Fig. 1. Distributions of content feature and liveness feature. Content feature is mixed together, while liveness feature is separated.



(a) Translation details of our method (b) Translation details of method w/o LBP map

Fig. 2. Translation details of different methods. With translation between the same live image and spoof image, our method exchanges the key details while method without LBP doesn't capture the meaningful details.

Table 1. Comparison of two low level supervision.

Method	LBP Map	HoG Map	LBP-Depth		HoG-Depth	
			Maximum	Average	Maximum	Average
APCER	**1.25**	1.88	2.92	1.67	2.92	2.08
BPCER	1.67	1.67	0.83	**0.83**	0.83	0.83
ACER	1.56	1.78	1.88	**1.25**	1.88	1.46

3 Role of Low Level Supervision

3.1 Translation Details with and without Low Level Supervision

In order to guide the network in extracting liveness feature, we put forward the combination of low level and high level supervision, where the LBP map is utilized as low level supervision, and the depth as high level supervision. In the Ablation Study Sec.4.3 of our submission, experimental results have demonstrated the effectiveness of such combination. To show the effect of the low level supervision more intuitively, we compare the translation details of methods with and without low level supervision. As illustrated in Fig. 2, if the method uses LBP supervision, when swapping the liveness feature, the spoof pattern in image level, such as moire, is also swapped. On the contrary, the spoof pattern of the reconstructed images is inconsistent. This indicates that low level supervision can help the liveness feature focuses on the discriminative places in the texture, which is in line with human perception.

3.2 Results of Different Low Level Supervision

In addition to LBP map, there are other kinds of low level supervision, such as HoG operator [1]. We substitute HoG constraint for LBP map as low level supervision and record the performance on Oulu-NPU protocol1. As shown in Table 1, results with LBP supervision are slightly better than with HoG supervision.

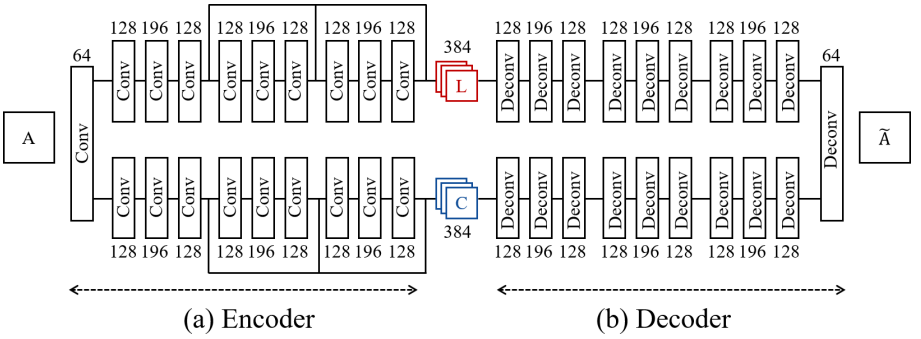


Fig. 3. Structure of Disentanglement Net

Table 2. The details of Auxiliary nets of our method.

LBP Net			Depth Net			Discriminator		
Layer	chan./Stri.	Out.Size	Layer	chan./Stri.	Out.Size	Layer	chan./Stri.	Out.Size
Input:liveness feature			**Input**:image			**Input**:image		
			conv2-0	64/1	256			
conv1-0	384/1	32	conv2-1	128/1	256	conv3-1	64/1	256
			conv2-2	196/1	256	pool3-1	-/2	128
			conv2-3	128/1	256			
			pool2-1	-/2	128			
conv1-1	128/1	32	conv2-4	128/1	128	conv3-2	128/1	128
			conv2-5	196/1	128	pool3-2	-/2	64
			conv2-6	128/1	128			
			pool2-2	-/2	64			
conv1-2	64/1	32	conv2-7	128/1	64	conv3-3	256/1	64
			conv2-8	196/1	64	pool3-3	-/2	32
			conv2-9	128/1	64			
			pool2-3	-/2	32			
conv1-2			pool2-1+pool2-2+pool2-3			pool3-3		
			conv2-10	128/1	32	conv3-4	512/1	16
			conv2-11	64/1	32	pool3-4	-/2	256
conv1-3	1/1	32	conv2-12	1/1	32	conv3-5	512/1	16

4 Network Structure

We demonstrate the structure of disentanglement net in Fig. 3. The detailed structure of three auxiliary nets are illustrated in Table 2. Each convolutional layer in Fig.3 is followed by an instance normalization layer and a rectified linear unit (ReLU) activation function with kernel size to be 3×3. While Each convolutional layer in Table.2 is followed by a batch normalization layer and a rectified linear unit (ReLU) activation function

References

1. Dalal, N., Triggs, B.: Histograms of oriented gradients for human detection. In: 2005 IEEE computer society conference on computer vision and pattern recognition (CVPR'05). vol. 1, pp. 886–893. IEEE (2005)