DLCV Final Project: Face Anti-Spoofing

Xuan-Rui Chen [1] Wei-Hsu Lee [2]

[1]b06901135 [2]b06901075

Introduction

Face anti-spoofing is an important secure topic in face recognition system. In this work, we implemented **three different approaches** as an attempt to solve the task.

The dataset used in the work is OULU-NPU [1] and SiW [4]. Each sample from OULU-NPU has 11 frames and a label in

$$\{\text{real}, \text{print1}, \text{print2}, \text{replay1}, \text{print2}\},$$

and each sample from SiW has 10 frames without label.

Dataset is setup as follow,

- **training set** is consist of 1200 samples from OULU-NPU,
- **validation set** is consist of 900 samples from OULU-NPU,
- **testing set** is consist of 600 samples from OULU-NPU, and 2053 samples from SiW.

Methods

The general idea behind three different approaches is anomaly detection. As the training set is labeled with different anomaly type, we simply trained a classifier and use the probability score of real category as anomaly score.

The three approaches are respectively video-based, texture-based, and spectrogram-based method.

Video-based approach

In this approach, we view each sample as multiple continuous video frames, and apply video action-recognition model to classify different type of anomaly.

Sense we view each sample as continuous frames, we done the same argumentation to each frame in one sample during training.

The models used are

- ResNet 3D [6],
- ResNet Mixed Convolution [6],
- ResNet (2+1)D [6].

Texture-based approach

In this approach, we expect model to detect anomaly by looking the texture of each image frame.

In training stage, we random crop the image to 224×224 pixels, then train the classifier at image level. In inference stage, we crop each frame to four corners and the central crop, then average the score of all crops in one sample.

The models used are

- ResNet 18, ResNet 50 [2],
- VGG 11, VGG 16, VGG 19 [5],

Spectrogram-based approach

In this approach, we got the idea from *Face De-Spoofing: Anti-Spoofing via Noise Modeling* [3]. The idea is that attacking with replaying or printing will cause some artifact, which can be revealed in spectrogram.

In training and inference stage, we fist convert each frame to spectrogram, then train the classifier.

The models used are the same as texture-based approach.

Results

The final result comparison of three different approaches.

Approach	Model	OULU (auc)	SiW (auc)	SiW (acc)
Video	ResNet 3d	0.96434	0.72271	0.56260
	ResNet mc3	0.99234	0.80554	0.62113
	ResNet r21d	0.96897	0.67417	0.58211
Texture	ResNet 18	0.97602	0.96823	**0.68943**
	ResNet 50	0.97723	0.95986	0.66991
	VGG 11	0.99556	0.98109	0.66178
	VGG 16	**0.99979**	0.98145	0.68455
	VGG 19	0.97965	0.95943	0.65528
	Blend	0.99657	**0.98812**	0.67642
Spectrogram	ResNet 18	0.85425	0.54876	-

Table 1:Performance comparison.

Experinments

TSNE visualization

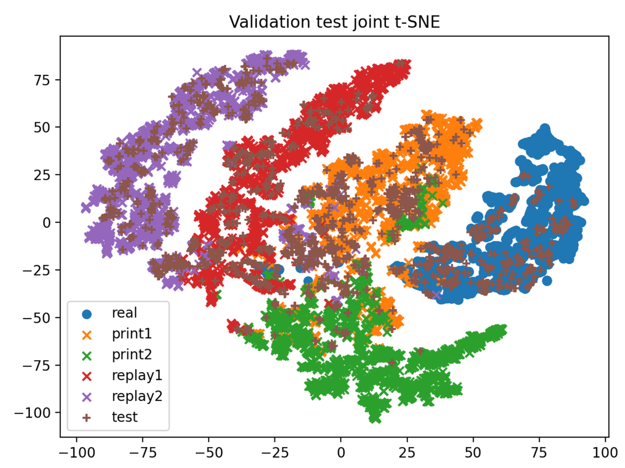

Figure 1:TSNE visualization of the last layer before output of texture-base approach (VGG16) on validation set and test set of OULU-NPU.

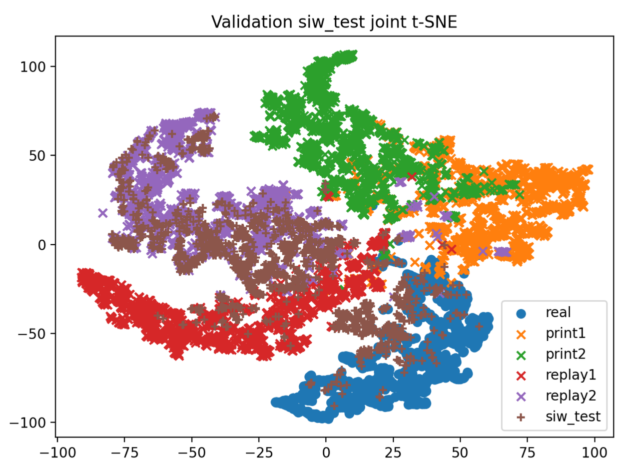

Figure 2:TSNE visualization of the last layer before output of texture-base approach (VGG16) on validation set of OULU-NPU and test set of SiW.

ROC curve

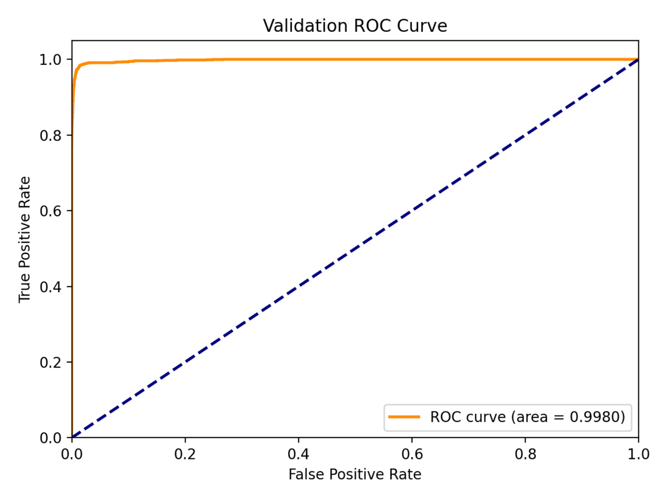

Figure 3:ROC curve of texture-base approach (VGG16) on validation set of OULU-NPU.

Conclusion

Single Model (VGG16) can perfome well on test set. With blending, we can reduce dataset biasing to get better score on out-domain testing SiW data. We found SiW is an unbalance dataset via TSNE visualization.

Also, we found our model struggled classifying "print1" and "real".

References

[1] Z. Boulkenafet, J. Komulainen, L. Li, X. Feng, and A. Hadid. Oulu-npu: A mobile face presentation attack database with real-world variations. In *2017 12th IEEE International Conference on Automatic Face Gesture Recognition (FG 2017)*, pages 612--618, 2017.

[2] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition, 2015.

[3] Amin Jourabloo, Yaojie Liu, and Xiaoming Liu. Face de-spoofing: Anti-spoofing via noise modeling, 2018.

[4] Yaojie Liu*, Amin Jourabloo*, and Xiaoming Liu. Learning deep models for face anti-spoofing: Binary or auxiliary supervision. In *In Proceeding of IEEE Computer Vision and Pattern Recognition*, Salt Lake City, UT, June 2018.

[5] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition, 2015.

[6] Du Tran, Heng Wang, Lorenzo Torresani, Jamie Ray, Yann LeCun, and Manohar Paluri. A closer look at spatiotemporal convolutions for action recognition, 2018.