ChaLearn Face Anti-spoofing Attack Detection Challenge fact sheets

March 12, 2019

1 Team details

- Team name

  VisionLabs

- Team leader name

  Parkin, Aleksandr

- Team contacts

  Address: Vijzelstraat 20, 4th Floor,
  1017 HK, Amsterdam,
  the Netherlands
  Email: a.parkin@visionlabs.ai

- Rest of the team members

  Parkin, Aleksandr (corresponding author)
  Grinchuk, Oleg

- Team website URL

  `https://visionlabs.ai`

- Affiliation

  VisionLabs

2 Contribution details

- Title of the contribution

  HardTune: Ensembling and Fine-tuning Face Recognition Networks for Multi-Modal Face Anti-spoofing

- Validation/Final score (if any)

	ACER	tpr@fpr=10e-2	tpr@fpr=10e-3	tpr@fpr=10e-4
Validation	0.0000	1.0000	1.0000	1.0000
Test	0.0008	0.9999	0.9996	0.9988

- General method description

  Our method uses a modified network architecture in [1]. As shown in Figure 1, the RGB, Depth and IR inputs are processed by separate streams followed by the concatenation and fully-connected layers. Differently from [1] we use aggregation blocks (Agg res2, ...) to aggregate outputs from multiple layers of the network. We pre-train network weights on four different tasks for face recognition and gender recognition. We then fine-tune these networks separately on the training set of the CASIA-SURF face anti-spoofing dataset. To increase the robustness to various attacks, we ensemble networks trained on three training folds and with two initial seeds. Results of our models evaluated separately and in combination are illustrated in Table 1.

- References

  [1] Shifeng Zhang, Xiaobo Wang, Ajian Liu, Chenxu Zhao, Jun Wan, Sergio Escalera, Hailin Shi, Zezheng Wang, Stan Z. Li, "CASIA-SURF: A Dataset and Benchmark for Large-scale Multi-modal Face Anti-spoofing", arXiv, 2018.

  [2] Dong Yi, Zhen Lei, Shengcai Liao, Stan Z. Li, "Learning Face Representation from Scratch", arXiv, 2014.

  [3] Zhenxing Niu, Mo Zhou, Le Wang, Xinbo Gao, Gang Hua, "Ordinal Regression With Multiple Output CNN for Age Estimation", The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2016, pp. 4920-4928

  [4] Yandong Guo, Lei Zhang, Yuxiao Hu, Xiaodong He, Jianfeng Gao, "MS-Celeb-1M: Challenge of Recognizing One Million Celebrities in the Real World", 2016

- Representative image / diagram of the method

  See Figure 1.

- Describe data preprocessing techniques applied (if any)

  The input image of each modality and its horizontal flip was resized to 125×125 pixels and cropped to 112×112 pixels around the center.

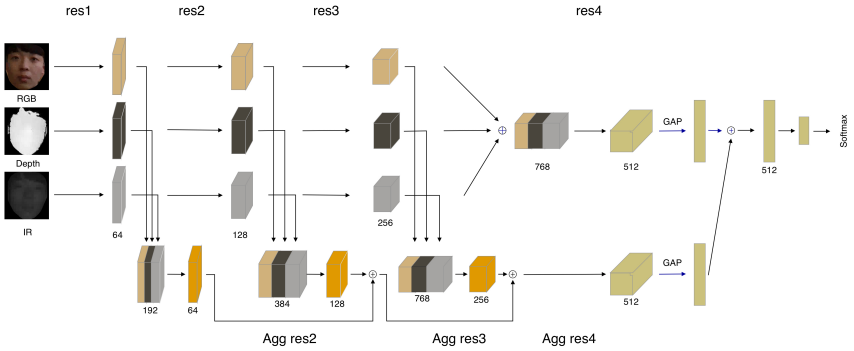

Figure 1: Deep layer aggregation architecture

3 Face Anti-spoofing Analysis

3.1 Features / Data representation

We use neural networks pre-trained on face and gender recognition tasks as a basis for our submission.

3.2 Dimensionality reduction

3.3 Fusing methods

3.4 Learning strategy

To increase robustness to unknown attacks, we split the training set into three folds according to different attacks present in the training subset. The outputs of three networks were ensembled by averaging to produce results on the final Validation and Test sets.

We use several learning strategies to train our models:

1. StepLR. Decay learning rate by 0.5 every 15 epochs.
2. CosineLR. Set learning rate by cosine annealing schedule with the period of 50 epochs

3.5 Other techniques

3.6 Method complexity

Our final model is an ensemble of 24 networks trained on three folds, using two initial seeds and pre-trained on four face and gender recognition tasks.

Table 1: Results on the Validation and Test sets of the challenge. "Seeds" indicates the use of two different initial seeds. See text for the description of different NN models.

NN1	NN1a	NN2	NN3	NN4	Seeds	Val tpr@fpr=10e-4	Test tpr@fpr=10e-4
✓						0.9943	-
	✓					0.9987	-
		✓				0.9870	-
			✓			0.9963	-
				✓		0.9933	-
✓		✓				0.9963	-
✓		✓	✓			0.9983	-
✓		✓	✓		✓	0.9997	-
✓		✓	✓	✓	✓	1.0000	-
	✓	✓	✓	✓	✓	**1.0000**	**0.9988**

3.7 Data Fusion Strategies

We fuse information from different modality streams by aggregating feature maps after each ResNet block and concatenates streams after res3 block, as show in Figure 1. Some of our models also use Squeeze and Excitation module after the res3 block.

3.8 Global Method Description

We opted to use an ensemble of various models to get a stable result on the test. The models were added to the ensemble in a greedy manner if they did not decrease the performance on the validation set. Results of our individual models and their combinations are presented in Table 1. We provide below descriptions for our individual models.

NN1 resnet-34. Pretrain description: facial recognition training on Casia-WebFace[2] with SphereFace loss. Antispoofing training: cosine LR strategy

NN2 resnet-34. Pretrain description: gender recognition training on AFAD-Lite[3], cosine lr strategy. Antispoofing training: cosine LR strategy

NN3 resnet-50. Pretrain description: facial recognition training on MSCeleb-1m with ArcFace loss, cosine LR strategy. Antispoofing training: cosine LR strategy

NN4 resnet-50. Pretrain description: facial recognition training on private asian dataset with ArcFace loss. Antispoofing training: cosine LR strategy

NN1a resnet-34. Pretrain description: facial recognition training on Casia-WebFace[2] with SphereFace loss. Antispoofing training: step LR strategy, last epoch is trained with weak augmentation

- Which pre-trained or external methods have been used (for any stage, if any)

  Models pre-trained on face recognition tasks have shown best results on the validation. We therefore use the pre-trained ResNet-34 on Casia-WebFace [2] from adversarial attacks on black box face recognition challenge (`https://competitions.codalab.org/competitions/19090`), ResNet-50 trained on MSCeleb-1M [4] and private asian dataset from public Jian Zhao repository (`https://github.com/ZhaoJ9014/face.evoLVe.PyTorch`).

- Which additional data has been used in addition to the provided ChaLearn training and validation data (at any stage, if any)

  We use no additional data specifically dedicated to the anti-spoofing task. We have trained ResNet-34 for the age recognition task on the AFAD-Lite [3] dataset to create a pre-trained model.

- Qualitative advantages of the proposed solution

- Results of the comparison to other approaches (if any)

- Novelty degree of the solution and if is has been previously published

4 Other details

- Language and implementation details (including platform, memory, parallelization requirements)

  We use python with the PyTorch deep learning framework

- Human effort required for implementation, training and validation?

- Training/testing expended time?

  Training one model takes about three hours on a 4 1080ti GPU. The inference on 1000 images takes about 8 seconds.

- General comments and impressions of the challenge? what do you expect from a new face anti-spoofing challenge?

  The CASIA-SURF dataset is well-prepared and has well-defined train/test splits for the comparison of defence methods against spoofing attacks. Meanwhile, we have noticed poor quality of some Depth images. We believe better results could be obtained if Depth input would be saved in the original format.