FaceBagNet: Bag-of-local-Features models for Multi-modal Face Anti-spoofing

March 12, 2019

1 Team details

- Team name
  ReadSense
- Team leader name
  Tao Shen (codalab username: SeuTao)
- Team leader address, phone number and email
  1st Fl., 28 East Yuyuan Rd., JingAn District;
  phone number: 15651667172;
  email: taoshen.seu@gmail.com;
- Rest of the team members
  Yuyu Huang
- Affiliation
  ReadSense

2 Contribution details

- Title of the contribution
  FaceBagNet: Bag-of-local-Features models for Multi-modal Face Anti-spoofing
- Validation/Final score (if any)
  Validation TPR@FPR=10E-4 1.00 /Final score TPR@FPR=10E-4 0.9991
- General method description
  In this challenge, we used a shallow seresnext network to classify the multi-modal face images based on image patches in different sizes. For further improve the performance, a multi-stream fusion network with three-modal images is utilized. The fusion network is trained from scratch in which RGB, Depth and IR data are feed into it at the same time. What's

more, heavy image augmentation is applied and modalities are randomly dropped during training. For optimization, we follow a cyclic cosine annealing learning rate schedule which yields better performance.

3 Face Anti-spoofing Analysis

3.0.1 Fusing methods

Feature fusing methods, i.e. multi-modality feature fusion, others (if any)
We use a multi-stream architecture with three subnetworks to perform multimodal feature fusion. The fusion network is trained from scratch in which RGB, Depth and IR data are feed into it at the same time.

3.0.2 Learning strategy

Learning strategy applied FOR face anti-spoofing (OR/AND SPOTTING) STAGE (if any)
We follow a cyclic cosine annealing learning rate schedule during training which yields better performance.

3.0.3 Method complexity

Method complexity FOR face anti-spoofing
For one image input, our model infers 36 times (9 position image patches with 4 flips). Considering the small input sizes (32x32 or 48x48) of our model, the complexity for face anti-spoofing is acceptable. The patch number per image can be reduced without losing too much precision if the efficiency is the primary concern.

3.1 Data Fusion Strategies

List data fusion strategies (how different feature descriptions are combined) for learning the model / network: RGB, depth, ir. (if any)
The fusion network is trained from scratch in which RGB, Depth and IR data are feed into it at the same time. Heavy image augmentation is applied and some of the three modal images are randomly dropped during training.

3.2 Global Method Description

- Qualitative advantages of the proposed solution
  Local features is important in this task and models using patch images with appropriate learning strategy is robust and has strong performance compared to more complex architectures. Our solution is simple but effective and also easy to use in practical application scenarios.

- Novelty degree of the solution and if is has been previously published
  Training CNN models with multi-modal image patches;

Random dropout input modalities;
Cyclic learning rate strategy for better performance;

4 Other details

- Language and implementation details (including platform, memory, parallelization requirements)
  Hardware: Intel(R) Xeon(R) CPU E5-2620 ,8 Titan X (Pascal);
  Software: Python 2, Pytorch 0.4.1;

- Training/testing expended time?
  Single fusion model training with patch size 32x32 takes about 5 hours on a NVIDIA TITAN X (Pascal). The test phase only takes a few minutes;

- General comments and impressions of the challenge? what do you expect from a new face anti-spoofing challenge?
  Great to have such a large scale data in the field of face anti-spoofing. We expect to see a new face anti-spoofing attack detection challenge with large scale data in the wild.