Domain Adaptation For Generalization Of Face Presentation Attack Detection In Mobile Settings With Minimal Information

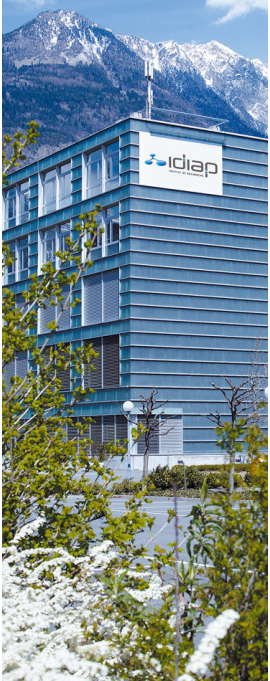
ICASSP 2020

Amir Mohammadi, Sushil Bhattacharjee, Sébastien Marcel

May 6, 2020

idiap RESEARCH INSTITUTE

EPFL



Outline

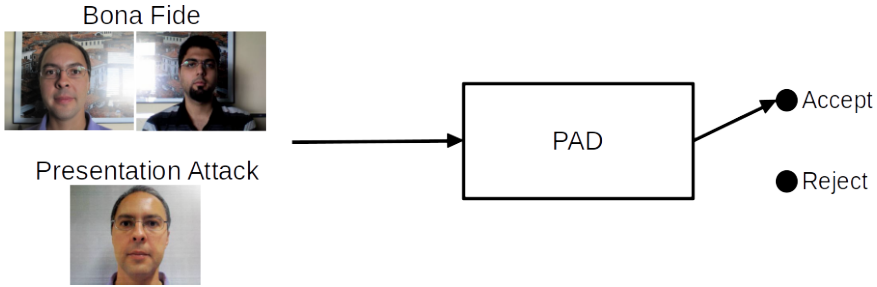
1. Introduction
2. Proposed Method
3. Experiments
4. Conclusions

Face Recognition (FR)

Face Recognition are vulnerable to presentation attacks.

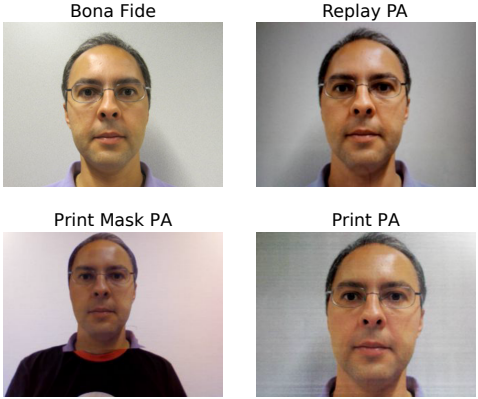
Examples of presentation attacks (PA): print attack (left) and 3D rigid mask attack (right).

Presentation Attack Detection (PAD)



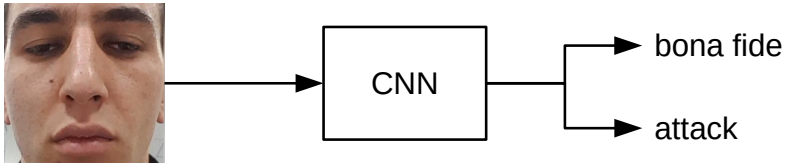
PAD systems are binary classification systems.

Presentation Attack Detection (PAD)



PAD systems rely on artifacts present in PAs to detect them.

PAD Using Deep Learning



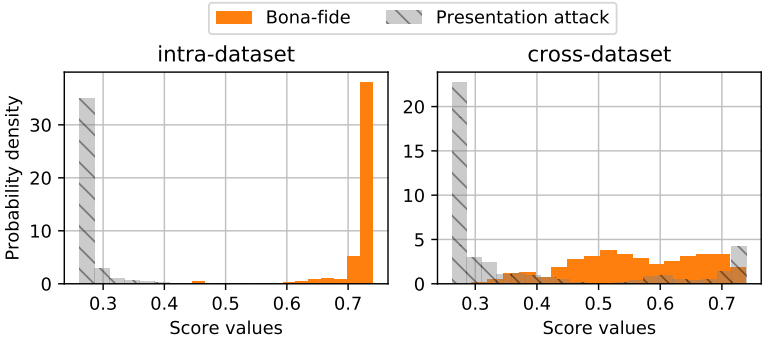
CNN based PAD approaches outperform previous methods which use hand-crafted features[1][2][3].

[1] J. Yang, Z. Lei, and S. Z. Li. "Learn Convolutional Neural Network for Face Anti-Spoofing". In: *arXiv:1408.5601 [cs]* (Aug. 2014).

[2] K. Patel, H. Han, and A. Jain. "Cross-Database Face Antispoofing with Robust Feature Representation". In: *Chinese Conference on Biometric Recognition*. 2016.

[3] Z. Boulkenafet, J. Komulainen, Z. Akhtar, et al. "A Competition on Generalized Software-Based Face Presentation Attack Detection in Mobile Scenarios". In: *Proceedings of the International Joint Conference on Biometrics, 2017*. Oct. 2017.

Problem: Generalization in PAD



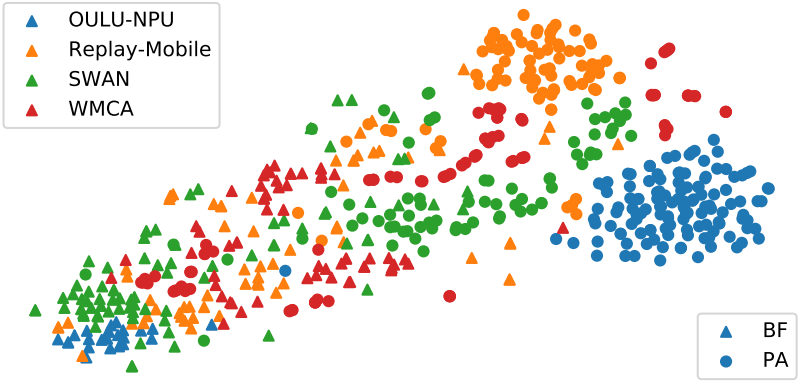
Intra-dataset vs cross-dataset PAD evaluation.

- Cross-dataset evaluations represent real-world scenarios.

Domain Shift

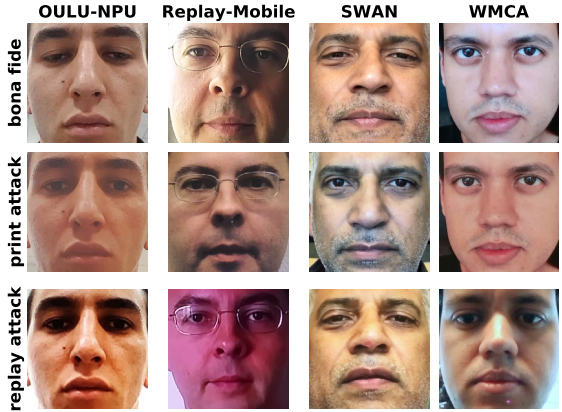
Visualization of learned features of a PAD CNN using t-SNE.
The PAD CNN is trained on the OULU-NPU dataset.

Domain Shift



Visualization of learned features of a PAD CNN using t-SNE.
The PAD CNN is trained on the OULU-NPU dataset.

Domain Shift



The face images can change drastically between datasets.

Solution to the problem: Domain Adaptation

- Each dataset can be seen as a domain.
- Domain adaptation/generalization methods can be used to improve performance.
- A more significant problem is that of data collection in the target domain.
- Specifically, whereas BF samples may be collected in the target domain at reasonable cost, collecting PAs in the target domain is usually much more expensive, if not impossible.
- Also, in real-world scenarios, a PAD system may be presented with attacks of previously unseen classes of PA.

Proposed Method: Domain Guided Pruning

- We propose a novel domain adaptation method relying on minimal information – only BF samples from the target domain.
- We hypothesize that, in a CNN trained for PAD using a source dataset, some learned filters in a layer are domain specific and others are domain invariant.
- We assume that by pruning domain specific layers, which do not generalize to the target dataset, we can improve the performance of the model on the target dataset.

Feature Divergence Measure (FDM)

Feature divergence measure (FDM)[4] is a way of quantifying domain shift at a given layer in a CNN.

[4] X. Pan et al. "Two at Once: Enhancing Learning and Generalization Capacities via Ibn-Net". In: *Proceedings of the European Conference on Computer Vision (ECCV)*. 2018.

Feature Divergence Measure (FDM)

- Given two datasets representing different domains, A and B, we want to determine, how often, on average, a specific filter in layer, L, is activated in each domain.
- Denote the average value of a filter over the spatial dimensions as f and assume a Gaussian distribution for f with mean μ and variance σ^2.
- The symmetric Kullback-Leibler (KL) divergence of this filter between domains A and B is:

$$D(f_A||f_B) = KL(f_A||f_B) + KL(f_B||f_A) \tag{1}$$

where $KL(f_A||f_B)$ is the KL divergence.

Feature Divergence Measure (FDM)

- Let us denote $D(f_{iA}||f_{iB})$ as the symmetric KL divergence of the i^{th} filter in layer L.
- Then, the average feature divergence of layer L is given by

$$D(L_A||L_B) = \frac{1}{C}\sum_{i=1}^{C} D(f_{iA}||f_{iB}) \tag{2}$$

  where C is the total number of filters in layer L.
- Higher values in Eqn. 1 indicate that the given filter is activated differently between datasets.
- Thus, the FDM for a given filter indicates whether it sensitive to the domain shift.

Proposed Method

1. Compute FDM (Eqn. 1) for each filter F at the layer L using only *bona fide* samples of the *training set* of datasets A and B.
2. Prune N percent of the filters[5] of layer L which contribute to the most feature divergence values at layer L.
3. Re-train the layers $L+1$ and after on the *training set* of the *source* dataset again (not the *target* dataset since it is assumed that no PAs are available for training in the target dataset) using the same classification loss-function to account for the pruned filters.

The pruned CNN is evaluated on the *evaluation set* of the target dataset.

[5] Pruning can be implemented either by multiplying the output of a filter by zero, or by removing the filter entirely from calculations to reduce the computational cost. Both methods result in the same behavior.

Proposed Method

- Intuitively, this method works like a feature selection method.
- The first L layers following the input layer of the CNN may be seen as a **feature extractor**.
- Layers $L+1$ and after may be seen as a **classifier**.
- Then, by pruning **features** at layer L and retraining the classifier, the classifier is limited to use only robust features for prediction.

Face PAD Datasets

The following four recent PAD datasets have been used in our experiments:

- OULU-NPU[6]
- Replay-Mobile[7]
- SWAN[8]
- WMCA[9]

OULU-NPU is chosen as the source dataset.

[6] Z. Boulkenafet, J. Komulainen, L. Li, et al. "OULU-NPU: A Mobile Face Presentation Attack Database with Real-World Variations". In: *Automatic Face & Gesture Recognition (FG 2017), 2017 12th IEEE International Conference On*. 2017

[7] A. Costa-Pazo et al. "The REPLAY-MOBILE Face Presentation-Attack Database". In: *Biometrics Special Interest Group (BIOSIG), 2016 International Conference of The*. 2016

[8] R. Ramachandra et al. "Smartphone Multi-Modal Biometric Authentication: Database and Evaluation". In: *arXiv:1912.02487 [cs]* (Dec. 2019)

[9] A. George et al. "Biometric Face Presentation Attack Detection with Multi-Channel Convolutional Neural Network". In: *IEEE Transactions on Information Forensics and Security* (2019)

Pruning Using Bona fide from Face Recognition Datasets

- The proposed method uses only *bona fide* samples from the target dataset.
- Instead of using *bona fide* samples from a target dataset, we can use *bona fide* from a face recognition dataset.
- This allows us to generalize to unknown domains.
- We used 3000 high quality images from IJB-C dataset[10] in the experiments.

[10] https://www.nist.gov/programs-projects/face-challenges

Experiments

PAD systems:

- DeepPixBiS[11] as a baseline PAD CNN.

- DeepPixBiS is pruned using *bona fide* data from a target dataset at layer L and layers $L+1$ and above are re-trained on the source dataset.

[11] A. George and S. Marcel. "Deep Pixel-Wise Binary Supervision for Face Presentation Attack Detection". In: *International Conference on Biometrics*. 2019.

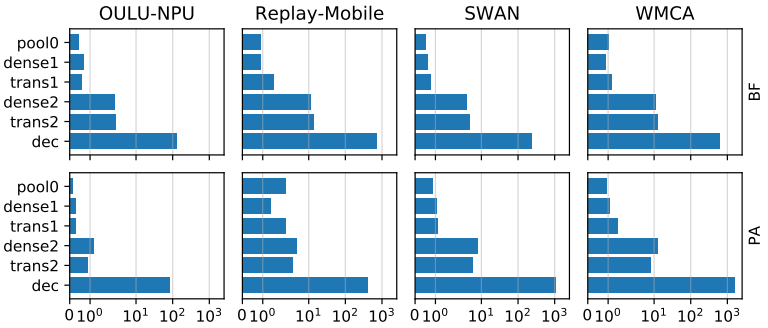
Details of DeepPixBiS

#	Layer	Details	Output Shape	Number of Parameters
1	conv0	Conv2D F=7 S=2	112 x 112 x 96	14,496
	pool0	MaxPool2D F=3 S=2	56 x 56 x 96	0
2	dense1	Dense Block	56 x 56 x 384	756,288
3	trans1	Transition Block	28 x 28 x 192	75,264
4	dense2	Dense Block	28 x 28 x 768	2,077,056
5	trans2	Transition Block	14 x 14 x 384	297,984
6	dec	Conv2D F=1 S=1	14 x 14 x 1	385

F is the number of filters and S is the stride. Layers 1 to 5 are identical with DenseNet-161[12]. The input to the network is a 224×224 pixel color face image.

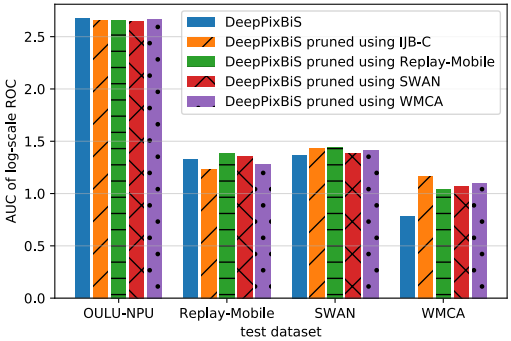
[12] G. Huang et al. "Densely Connected Convolutional Networks". In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2017.

Feature Divergences



Feature divergence at different layers computed between the *training set* of OULU-NPU and the *evaluation set* of OULU-NPU, Replay-Mobile, SWAN, and WMCA. FDM values are computed per layer (y-axis) and per class (top row for BF and bottom row for PA).

Performance Evaluation of the Proposed Method



The models are compared to the baseline when no pruning is performed. The evaluation-dataset is mentioned on the x axis. The higher the value the better is the performance of the system.

Conclusions

- In this work we have formulated the problem of generalization in PAD systems as a domain adaptation (DA) problem.
- DA methods usually rely on having sufficient data in target domain
- In biometrics, collecting *bona fide* samples in target domain is usually affordable, but not PA data.
- Pruning using the target dataset, increased the performance of the model on the target dataset.
- Pruning did not degrade the performance of the model on the source dataset.

Code and models available at: `https://gitlab.idiap.ch/bob/bob.paper.icassp2020_domain_guided_pruning`