Presentation Attack Detection and Unknown Attacks

Marta Gomez-Barrero
Hochschule Ansbach
IFPC‘20, 28/10/2020
Introduction
Unknown PAD
Train on PA and BF
Anomaly detection
Conclusions
What are Presentation Attacks? [ISO/IEC IS 30107-1 on PAD]

- Presentation attack: presentation to the biometric capture subsystem with the goal of interfering with the operation of the biometric system

  - **Impostor**: the attacker attempts to being matched to someone else's biometric reference
  - **Identity concealer**: the attacker attempts to avoid being matched to their own biometric reference (i.e., to avoid a black-list)
Current situation

- There is a clear need to develop Presentation Attack Detection (PAD) methods, especially for non-supervised access control scenarios.
- Good news, there is a lot of work done, reporting really good results and error rates close to 0%.

R. Ramachandra, C. Busch, “Presentation attack detection methods for face recognition systems: A comprehensive survey”, ACM CSUR, 2017
E. Marasco, A. Ross, “A survey on antispooﬁng schemes for ﬁngerprint recognition systems” ACM CSUR, 2014
J. Galbally, M. Gomez-Barrero, , “Presentation attack detection in iris recognition”, in Iris and Periocular Biometrics, IET, 2017
A. Czajka, K. Bowyer, “Presentation attack detection for iris recognition: An assessment of the state-of-the-art”, ACM CSUR, 2018
But what about unknown attacks? Attacks not seen during training

State of the art PAD methods, based on image quality measures and Support Vector Machines (SVMs), present poor generalisation capabilities to detect unknown attacks

- Error rates are multiplied by up to 6!
- Even worse, the performance of the system drops significantly for lower APCER values – probably the most interesting operating points

T. de Freitas Pereira, A. Anjos, J. D. Martino, S. Marcel, “Can face anti-spoofing countermeasures work in a real world scenario?”, in Proc. ICB, 2013

O. Nikisins, A. Mohammadi, A. Anjos, S. Marcel, “On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing”, in Proc. ICB, 2018
In compliance with the ISO/IEC IS 30107-3 on biometric PAD – Part 3: Testing and reporting, the following metrics should be used:

- **Attack Presentation Classification Error Rate (APCER):** percentage of attack presentations wrongly classified as bona fide presentations.

- **Bona Fide Presentation Classification Error Rate (BPCER):** percentage of bona fide presentations wrongly classified as attack presentation.

- **Detection Equal Error Rate (D-EER):** operation point where APCER = BPCER.
What can we do?

- Profit from ALL the data we have and simulate an unknown attack scenario, using different Presentation Attack Instrument (PAI) species for train and test
  - Leave-One-Out (LOO) protocol
  - Deep learning! Generative Adversarial Networks (GANs)

- Employ anomaly detection techniques:
  - Bona fides = “normal“ data
  - Presentation attacks = outliers, noise
Fisher vector representation

- Add generalisation capabilities to existent PAD approaches

- For instance: project the features used by the classifiers into a new feature space, where new attacks will look similar to known attacks

- Good results with Fisher vectors for face, fingerprints, and speech!
  - Deep learning is not the only way to good results and generalisation 😊

L. J. Gonzalez-Soler, M. Gomez-Barrero, C. Busch, “Fisher Vector Encoding of Dense-BSIF Features for Unknown Face Presentation Attack Detection”, in Proc. BIOSIG, 2020


Fisher vector representation

Training on PA and BF

1. Feature extraction
   - Training samples: Known PAIs, Known BP
   - BSIF, LBP, SIFT, SURF, ...

2. GMM feature space
   - PAI features, BP features

3. FV projection
   - $f_{11} \ f_{12} \ ... \ f_{1d}$
   - $f_{21} \ f_{22} \ ... \ f_{2d}$
   - $f_{n1} \ f_{n2} \ ... \ f_{nd}$

4. Prediction
   - BP/AP

Probe unknown PAIs

L. J. Gonzalez-Soler, M. Gomez-Barrero, C. Busch, “Fisher Vector Encoding of Dense-BSIF Features for Unknown Face Presentation Attack Detection”, in Proc. BIOSIG, 2020

Marta Gomez-Barrero

IFPC 2020, PAD and Unknown Attacks, 28/10/2020
Fisher vector representation

- For known attacks (CASIA database):
  - BSIF + SVM → D-ERR = 10.21%
  - State of the art (no deep learning) → D-ERR = 4.60%
  - BSIF + Fisher Vectors + SVM → D-EER = 1.79%

- For unknown attacks:

<table>
<thead>
<tr>
<th></th>
<th>CASIA</th>
<th>REPLAY-ATTACK</th>
<th>REPLAY-MOBILE</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Cut</td>
<td>Warped</td>
<td>Video</td>
</tr>
<tr>
<td>AUC</td>
<td>99.6</td>
<td>97.9</td>
<td>99.9</td>
</tr>
<tr>
<td>D-EER</td>
<td>4.11</td>
<td>6.15</td>
<td>1.37</td>
</tr>
</tbody>
</table>

L. J. Gonzalez-Soler, M. Gomez-Barrero, C. Busch, “Fisher Vector Encoding of Dense-BSIF Features for Unknown Face Presentation Attack Detection”, in Proc. BIOSIG, 2020
Fisher vector representation

- Depending on the PAI species / training set, very different results!
- And this is not exclusive to face, but has also been reported for other characteristics (e.g., fingerprint)


T. Chugh, A. K. Jain, “Fingerprint presentation attack detection: Generalization and efficiency”, in Proc. ICB, 2019
Train on PA and BF

Fisher vector representation

t-SNE visualisation of the FV common feature space for CASIA

- green: bona fide
- blue: cut photo attack
- red: video replay attack
- yellow: warped photo attack
Anomaly detection

One-Class Classifiers

- Adapt classifiers such as SVMs or Gaussian Mixture Models (GMMs) to be trained only on bona fide data

O. Nikisins, A. Mohammadi, A. Anjos, S. Marcel, “On effectiveness of anomaly detection approaches against unseen presentation attacks in face anti-spoofing”, in Proc. ICB, 2018
Generative Adversarial Networks (GANs)

- A GAN is made of two parts:
  - Generator network: takes as input a random vector and decodes it into a synthetic image
  - Discriminator network (adversary): takes as input an image (real or synthetic) and predicts its type
Generative Adversarial Networks (GANs)

- We can use a trained discriminator for PAD:
  - Synthetic = presentation attack
  - Real = bona fide

- For fingerprint PAD, APCER @ BPCER = 0.2%:

<table>
<thead>
<tr>
<th>Material</th>
<th>Gelatin</th>
<th>Pigmented</th>
<th>Playdoh</th>
<th>Woodglue</th>
<th>Transparency</th>
<th>Gold finger</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Binary CNN</strong></td>
<td>32.3%</td>
<td>70.6%</td>
<td>99.4%</td>
<td>44.3%</td>
<td>66.0%</td>
<td>88.2%</td>
</tr>
<tr>
<td><strong>1-class GAN</strong></td>
<td>26.4%</td>
<td>77.7%</td>
<td>3.7%</td>
<td>14.8%</td>
<td>6.0%</td>
<td>60.8%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Material</th>
<th>Dragonskin</th>
<th>Ecoflex</th>
<th>Monster Latex</th>
<th>Crayola Magic</th>
<th>Body Latex</th>
<th>2D paper</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Binary CNN</strong></td>
<td>51.0%</td>
<td>60.7%</td>
<td>45.7%</td>
<td>21.9%</td>
<td>87.9%</td>
<td>53.9%</td>
</tr>
<tr>
<td><strong>1-class GAN</strong></td>
<td>97.9%</td>
<td>95.2%</td>
<td>61.5%</td>
<td>16.4%</td>
<td>99.7%</td>
<td>43.2%</td>
</tr>
</tbody>
</table>

J. J. Engelsma, A. K. Jain, “Generalizing fingerprint spoof detector: Learning a one-class classifier” in Proc. ICB, 2019

Marta Gomez-Barrero

IFPC 2020, PAD and Unknown Attacks, 28/10/2020
Autoencoders (AEs)

- A classical image autoencoder takes an image, maps it to a latent vector space, and decodes it back.
- The autoencoder learns to reconstruct the original inputs.
- If we only train it on bona fide data, it will produce “weird” results for presentation attacks on the test set.
Anomaly detection

Autoencoders

Autoencoders

- 19,711 bona fide and 4,339 PA images, 45 different PAI species
Autoencoders vs. Two-class CNNs

What happens when we benchmark Autoencoders with two-class classifiers evaluated on a leave-one-out (LOO) protocol?

Anomaly Detection

Autoencoders vs. Two-class CNNs

- Baseline training on BFs + PAs (APCER @ BPCER = 0.2%)

![Graph showing comparisons between different models]

- Laser AE (24.33%)
- Laser CNN ResNet (8.70%)
- Laser CNN VGG16 (16.29%)
- Laser CNN VGGFace (4.85%)
- Laser LRCN VGG16 (5.22%)
- SWIR AE (6.59%)
- SWIR CNN MobileNetV2 (3.91%)
- SWIR CNN VGG16 (3.38%)
- SWIR CNN VGG19 (6.25%)
- SWIR CNN VGGFace (10.23%)
- Fusion (2.74%)
Anomaly detection

Autoencoders vs. Two-class CNNs

- LOO per groups: fake fingers
Anomaly detection

Autoencoders vs. Two-class CNNs

LoO per groups: opaque overlays
Anomaly detection

Autoencoders vs. Two-class CNNs

- LOO per groups: transparent overlays

---

Marta Gomez-Barrero

IFPC 2020, PAD and Unknown Attacks, 28/10/2020
Train on PA and BF

Style transfer with deep learning

- Goal: generate synthetic PA images of unknown materials, by transferring the style (texture) characteristics from known materials

Train on PA and BF

Style transfer with GANs and AEs

- Style transfer is done with an AE construction
- A GAN construction enhances the appearance of the synthetic samples

Train on PA and BF

Style transfer with deep learning

- For fingerprint PAD, APCER @ BPCER = 0.2% on a LOO partition:

<table>
<thead>
<tr>
<th>Material</th>
<th>Gelatin</th>
<th>Silicone</th>
<th>Playdoh</th>
<th>Woodglue</th>
<th>Transparency</th>
<th>Gold Finger</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>45.05%</td>
<td>32.28%</td>
<td>41.58%</td>
<td>13.62%</td>
<td>4.17%</td>
<td>11.78%</td>
</tr>
<tr>
<td>CNN + UMG</td>
<td>2.04%</td>
<td>1.36%</td>
<td>27.64%</td>
<td>1.03%</td>
<td>0.00%</td>
<td>11.41%</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Material</th>
<th>Dragonskin</th>
<th>Conductive Ink + Paper</th>
<th>Monster Latex</th>
<th>3D Univ. Targets</th>
<th>Body Latex</th>
<th>2D paper</th>
</tr>
</thead>
<tbody>
<tr>
<td>CNN</td>
<td>2.52%</td>
<td>10.00%</td>
<td>5.23%</td>
<td>5.00%</td>
<td>23.65%</td>
<td>44.56%</td>
</tr>
<tr>
<td>CNN + UMG</td>
<td>0.00%</td>
<td>0.00%</td>
<td>3.76%</td>
<td>0.00%</td>
<td>10.28%</td>
<td>19.78%</td>
</tr>
</tbody>
</table>

Good performing PAD methods are still not robust to unknown attacks

Two approaches to tackle the problem:
- Use a limited number of PAI species for training in order to simulate an unknown attacks scenario
- Use anomaly detection approaches

Main findings:
- Traditional one-class classifiers cannot reach a good performance in the state of the art (so far)
- Alternatives based on feature projection, autoencoders, or style transfer do offer good results, but highly dependent on the PAI species

There is still work to be done!
Marta Gomez-Barrero
(marta.gomez-barrero@hs-ansbach.de)