Session III: Unknown Attacks, Additional Sensors and Practical Tips

Host: Xiaoming Liu
Outline

• Training-Testing Difference
• Alternative/Additional Sensors
• Practical Tips
• Future
Outline

• Training-Testing Difference
• Alternative/Additional Sensors
• Practical Tips
• Future
Training-Testing Difference

The testing scenarios are different with the training phase.

- Environment (Lighting, Indoor/outdoor, etc.)
- Camera/Image quality
- Subjects (Age, Race, etc.)
- Spoof types
Training-Testing Difference

The testing scenarios are different with the training phase.

- Environment (Lighting, Indoor/outdoor, etc.)
- Camera/Image quality
- Subjects (Age, Race, etc.)
- Spoof types
Cross-database Domain Adaption

Learning Generalizable and Identity-Discriminative Representations for Face Anti-Spoofing, TIFS, 2018

Multi-adversarial Discriminative Deep Domain Generalization, CVPR, 2019

Domain Adaptation in Multi-Channel Autoencoder based Features for Robust Face Anti-Spoofing, ICB 2019

Improving Cross-database Face Presentation Attack Detection via Adversarial Domain Adaptation, ICB 2019
Learning Generalizable and Identity-Discriminative Representations for Face Anti-Spoofing

- Learn face anti-spoofing and face recognition at the same time
- Apply a Fast Domain Adaption (FDA) to remove the bias of different domain
- Share the weights of face anti-spoofing and face recognition
Learning Generalizable and Identity-Discriminative Representations for Face Anti-Spoofing

• Total Pairwise Confusion (TPC) loss

$$L_{tpc}(x_i, x_j) = \sum_{i \neq j} ||\psi(x_i) - \psi(x_j)||^2_2$$

$$\psi(x)$$ is the second fully connected layer of the face anti-spoofing branch

• Anti-loss: cross entropy losses for face anti-spoofing

• Recognition loss: cross entropy losses for face recognition
Feature w/ and w/o TPC loss

• Remove person id information from anti-spoofing feature
  • Irrelevant to face anti-spoofing
  • May lead to a more generalized feature
Feature w/ and w/o TPC loss

- Remove person id information from anti-spoofing feature
  - Irrelevant to face anti-spoofing
  - May lead to a more generalized feature

<table>
<thead>
<tr>
<th>TPC/FDA</th>
<th>Intra-Test</th>
<th>Cross-Test</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>MFSD</td>
<td>Replay</td>
</tr>
<tr>
<td>− −</td>
<td>10.5</td>
<td>0.6</td>
</tr>
<tr>
<td>− +</td>
<td>11.2</td>
<td>0.6</td>
</tr>
<tr>
<td>+ −</td>
<td><strong>6.4</strong></td>
<td><strong>0</strong></td>
</tr>
<tr>
<td>+ +</td>
<td>8.3</td>
<td>0.3</td>
</tr>
</tbody>
</table>
Learning Generalizable and Identity-Discriminative Representations for Face Anti-Spoofing

- **Fast Domain Adaption (FDA)**
  - Style transfer network
  - Content loss + Style (domain) loss

\[
\mathcal{L}_{\text{content}} = \frac{1}{C_jH_jW_j} \frac{1}{\| \varphi_j(y) - \varphi_j(x) \|_2^2}
\]

\[
\mathcal{L}_{\text{domain}} = \frac{1}{C_jH_jW_j} \frac{1}{\| G_j(y) - G_j(y_d) \|_F^2}
\]

\[
\hat{y} = \arg \min_{P} (\lambda_c \mathcal{L}_{\text{content}}(y, x) + \lambda_s \mathcal{L}_{\text{domain}}(y, y_d))
\]
## Testing on Oulu

<table>
<thead>
<tr>
<th>Protocol</th>
<th>Method</th>
<th>APCER</th>
<th>BPCER</th>
<th>ACER</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>P1</strong></td>
<td>GRADIANT</td>
<td>1.3%</td>
<td>12.5%</td>
<td>6.9%</td>
</tr>
<tr>
<td></td>
<td>Auxiliary</td>
<td>1.6%</td>
<td><strong>1.6%</strong></td>
<td>1.6%</td>
</tr>
<tr>
<td></td>
<td>DS Net</td>
<td><strong>1.2%</strong></td>
<td>1.7%</td>
<td><strong>1.5%</strong></td>
</tr>
<tr>
<td></td>
<td>GFA-CNN</td>
<td>2.5%</td>
<td>8.9%</td>
<td>5.7%</td>
</tr>
<tr>
<td><strong>P2</strong></td>
<td>Auxiliary</td>
<td>2.7%</td>
<td>2.7%</td>
<td>2.7%</td>
</tr>
<tr>
<td></td>
<td>GRADIANT</td>
<td>3.1%</td>
<td>1.9%</td>
<td>2.5%</td>
</tr>
<tr>
<td></td>
<td>DS Net</td>
<td>4.2%</td>
<td>4.4%</td>
<td>4.3%</td>
</tr>
<tr>
<td></td>
<td>GFA-CNN</td>
<td><strong>2.5%</strong></td>
<td>1.3%</td>
<td><strong>1.9%</strong></td>
</tr>
<tr>
<td><strong>P3</strong></td>
<td>GRADIANT</td>
<td><strong>2.6±3.9%</strong></td>
<td>5.0±5.3%</td>
<td>3.8±2.4%</td>
</tr>
<tr>
<td></td>
<td>Auxiliary</td>
<td>2.7±1.3%</td>
<td><strong>3.1±1.7%</strong></td>
<td><strong>2.9±1.5%</strong></td>
</tr>
<tr>
<td></td>
<td>DS Net</td>
<td>4.0±1.8%</td>
<td>3.8±1.2%</td>
<td>3.6±1.6%</td>
</tr>
<tr>
<td></td>
<td>GFA-CNN</td>
<td>4.3%</td>
<td>7.1%</td>
<td>5.7%</td>
</tr>
<tr>
<td><strong>P4</strong></td>
<td>GRADIANT</td>
<td><strong>5.0±4.5%</strong></td>
<td>15.0±7.1%</td>
<td>10.0±5.0%</td>
</tr>
<tr>
<td></td>
<td>Auxiliary</td>
<td>9.3±5.6%</td>
<td>10.4±6.0%</td>
<td>9.5±6.0%</td>
</tr>
<tr>
<td></td>
<td>DS Net</td>
<td>5.1±6.3%</td>
<td><strong>6.1±5.0%</strong></td>
<td><strong>5.6±5.7%</strong></td>
</tr>
<tr>
<td></td>
<td>GFA-CNN</td>
<td>7.4%</td>
<td>10.4%</td>
<td>8.9%</td>
</tr>
</tbody>
</table>

Li et. al., Learning Generalizable and Identity-Discriminative Representations for Face Anti-Spoofing, TIFS, 2018
Improving Cross-database Face Presentation Attack Detection via Adversarial Domain Adaptation

- Pretrain a source encoder/decoder
- Learn a target encoder such that discriminator cannot correctly predict the domain
- Classify with k-NN classifier
Improving Cross-database Face Presentation Attack Detection via Adversarial Domain Adaptation

• Encoder:
  • 4 convolution blocks
  • 1 pooling layer

• Decoder:
  • 2 fully connected layers

Wang et. al., Improving Cross-database Face Presentation Attack Detection via Adversarial Domain Adaptation, 2019
Improving Cross-database Face Presentation Attack Detection via Adversarial Domain Adaptation

<table>
<thead>
<tr>
<th>Method</th>
<th>C → I</th>
<th>C → M</th>
<th>I → C</th>
<th>I → M</th>
<th>M → C</th>
<th>M → I</th>
<th>Average</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed w/o ML&amp;ADA</td>
<td>43.8</td>
<td>33.8</td>
<td>49.5</td>
<td>41.3</td>
<td>45.4</td>
<td>39.6</td>
<td>42.2</td>
</tr>
<tr>
<td>Proposed w/o ML</td>
<td>43.7</td>
<td>29.6</td>
<td>50.0</td>
<td>35.4</td>
<td>46.5</td>
<td>38.7</td>
<td>40.7</td>
</tr>
<tr>
<td>Proposed w/o ADA</td>
<td>43.3</td>
<td>14.0</td>
<td>45.4</td>
<td>35.3</td>
<td>37.8</td>
<td>11.5</td>
<td>31.2</td>
</tr>
<tr>
<td>Proposed (full method)</td>
<td>17.5</td>
<td>9.3</td>
<td>41.6</td>
<td>30.5</td>
<td>17.7</td>
<td>5.1</td>
<td>20.3</td>
</tr>
</tbody>
</table>
Multi-adversarial Deep Domain Generalization for Face Presentation Attack Detection

• Learn a feature space that is discriminative and shared by multiple source domains

Shao et. al., Multi-adversarial Discriminative Deep Domain Generalization for Face Presentation Attack Detection, CVPR, 2019
Multi-adversarial Deep Domain Generalization for Face Presentation Attack Detection

- Feature generator
  - extract features for face anti-spoofing
  - adversarial-trained to remove domain information
- Depth estimation
  - improve the discriminativeness
- Dual-force triplet mining
  - enforce a smaller intra-class distance
  - enforce a larger inter-class distance
  - cross domain
Multi-adversarial Deep Domain Generalization for Face Presentation Attack Detection

- Learn features extractors for N domains
- Learn a feature generator for all domains
- Adversarial train N discriminators to make the feature generator more generalized.

Shao et. al., Multi-adversarial Discriminative Deep Domain Generalization for Face Presentation Attack Detection, CVPR, 2019
Dual-force Triplet Mining

• In one domain
  • Minimize live-to-live / spoof-to-spoof distance between different subjects
  • Maximize live-to-spoof distance between different subjects

• Cross domains
  • Minimize live-to-live / spoof-to-spoof distance between different subjects
  • Maximize live-to-spoof distance between different subjects

Shao et. al., Multi-adversarial Discriminative Deep Domain Generalization for Face Presentation Attack Detection, CVPR, 2019
### Multi-adversarial Discriminative Deep Domain Generalization

<table>
<thead>
<tr>
<th>Method</th>
<th>O&amp;C&amp;I to M</th>
<th>O&amp;M&amp;I to C</th>
<th>O&amp;C&amp;M to I</th>
<th>I&amp;C&amp;M to O</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HTER(%)</td>
<td>AUC(%)</td>
<td>HTER(%)</td>
<td>AUC(%)</td>
</tr>
<tr>
<td>MS_LBP</td>
<td>29.76</td>
<td>78.50</td>
<td>54.28</td>
<td>44.98</td>
</tr>
<tr>
<td>Binary CNN</td>
<td>29.25</td>
<td>82.87</td>
<td>34.88</td>
<td>71.94</td>
</tr>
<tr>
<td>IDA</td>
<td>66.67</td>
<td>27.86</td>
<td>55.17</td>
<td>39.05</td>
</tr>
<tr>
<td>Color Texture</td>
<td>28.09</td>
<td>78.47</td>
<td>30.58</td>
<td>76.89</td>
</tr>
<tr>
<td>LBPTOP</td>
<td>36.90</td>
<td>70.80</td>
<td>42.60</td>
<td>61.05</td>
</tr>
<tr>
<td>Auxiliary(Depth Only)</td>
<td>22.72</td>
<td>85.88</td>
<td>33.52</td>
<td>73.15</td>
</tr>
<tr>
<td>Auxiliary(All)</td>
<td>–</td>
<td>–</td>
<td>28.4</td>
<td>–</td>
</tr>
<tr>
<td>Ours (MADDG)</td>
<td><strong>17.69</strong></td>
<td><strong>88.06</strong></td>
<td><strong>24.5</strong></td>
<td><strong>84.51</strong></td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Method</th>
<th>O&amp;C&amp;I to M</th>
<th>O&amp;M&amp;I to C</th>
<th>O&amp;C&amp;M to I</th>
<th>I&amp;C&amp;M to O</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>HTER(%)</td>
<td>AUC(%)</td>
<td>HTER(%)</td>
<td>AUC(%)</td>
</tr>
<tr>
<td>MMD-AAE</td>
<td>27.08</td>
<td>83.19</td>
<td>44.59</td>
<td>58.29</td>
</tr>
<tr>
<td>Ours (MADDG)</td>
<td><strong>17.69</strong></td>
<td><strong>88.06</strong></td>
<td><strong>24.5</strong></td>
<td><strong>84.51</strong></td>
</tr>
</tbody>
</table>

Shao et. al., Multi-adversarial Discriminative Deep Domain Generalization, CVPR, 2019
Domain Adaptation in Multi-Channel Autoencoder based Features for Robust Face Anti-Spoofing

- Use multi-modality data (RGB, NIR, and Depth) instead of RGB only
- Domain Adaption: fine-tuning (RGB $\rightarrow$ NIR-Depth)

George et. al., Biometric Face Presentation Attack Detection with Multi-Channel Convolutional Neural Network, TIFS 2019
Domain Adaptation in Multi-Channel Autoencoder based Features for Robust Face Anti-Spoofing

<table>
<thead>
<tr>
<th>Method</th>
<th>dev (%)</th>
<th>test (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>APCER</td>
<td>ACER</td>
</tr>
<tr>
<td>Color (IQM-LR)</td>
<td>76.58</td>
<td>38.79</td>
</tr>
<tr>
<td>Depth (LBP-LR)</td>
<td>57.71</td>
<td>29.35</td>
</tr>
<tr>
<td>Infrared (LBP-LR)</td>
<td>32.79</td>
<td>16.9</td>
</tr>
<tr>
<td>Thermal (LBP-LR)</td>
<td>11.79</td>
<td>6.4</td>
</tr>
<tr>
<td>Score fusion (IQM-LBP-LR Mean fusion)</td>
<td>10.52</td>
<td>5.76</td>
</tr>
<tr>
<td>Color (RDWT-Haralick-SVM)</td>
<td>36.02</td>
<td>18.51</td>
</tr>
<tr>
<td>Depth (RDWT-Haralick-SVM)</td>
<td>34.71</td>
<td>17.85</td>
</tr>
<tr>
<td>Infrared (RDWT-Haralick-SVM)</td>
<td>14.03</td>
<td>7.51</td>
</tr>
<tr>
<td>Thermal (RDWT-Haralick-SVM)</td>
<td>21.51</td>
<td>11.26</td>
</tr>
<tr>
<td>Score fusion (RDWT-Haralick-SVM Mean fusion)</td>
<td>6.2</td>
<td>3.6</td>
</tr>
<tr>
<td>FASNet</td>
<td>18.89</td>
<td>9.94</td>
</tr>
</tbody>
</table>
Unknown Attack Detection

• One-class SVM
• Gaussian Mixture Model
• AutoEncoder
Unknown Attack Detection


Unknown Presentation Attack Detection with Face RGB Images, ICB, 2018

Deep Anomaly Detection for Generalized Face Anti-Spoofing, CVPRW, 2019

Deep Tree Learning for Zero-shot Face Anti-Spoofing, CVPR 2019

A very comprehensive study on various hand-crafted feature and classifiers.

- Feature: LBP-TOP, LPQ-TOP, BSIF-TOP, Image quality measures
- Classifier: SVM1, SVM2, LDA2, Sparse representation classifier (SRC)1, SRC 2
- Dataset: CASIA-FASD, Replay-attack, MSU-MFSD

A very comprehensive study on various hand-crafted feature and classifiers.

- Feature: LBP-TOP, LPQ-TOP, BSIF-TOP, Image quality measures
- Classifier: SVM1, SVM2, LDA2, Sparse representation classifier (SRC)1, SRC 2
- Dataset: CASIA-FASD, Replay-attack, MSU-MFSD

Conclusion: neither the two-class systems nor the one-class approaches perform well enough
Unknown Presentation Attack Detection with Face RGB Images

A very comprehensive study on various hand-crafted feature and classifiers.

- Feature: Color LBP
- Classifier: SVM1, Auto Encoder, GMM
- Dataset: CASIA-FASD, Replay-attack, MSU-MFSD
Unknown Presentation Attack Detection with Face RGB Images

- **Dataset:** CASIA-FASD, Replay-attack, MSU-MFSD

- **Conclusion:** improve the performance
  - NN+LBP works best on C+R+M protocols
  - AE+LBP works best on Oulu protocols
Deep Anomaly Detection for Generalized Face Anti-Spoofing

- Deep metric learning
- Triplet Focal loss
  - Focus on the harder cases
Literature and Issues

• Limited Spoof Types\textsuperscript{1,2}

• Only model the live distribution\textsuperscript{1,2}

[2] F. Xiong and W. Abdalmageed. Unknown presentation attack detection with face RGB images. BTAS 2018
What if More Spoof Types?

Live
- Half Mask
- Silicone
- Transparent
- Papercraft
- Mannequin

Print
- Obfuscation
- Imperson.
- Cosmetic
- Funny Eye
- Paperglasses
- Partial Paper

3D Mask Attacks
- Makeup Attacks

Replay
Deep Tree Learning for Zero-shot Face Anti-Spoofing

- Previous methods only model the live
- Learning semantic spoof attributes
Deep Tree Networks (DTN)

Tree Nodes

Leaf Nodes

Convolutional Residual Unit

Tree Routing Unit

Supervised Feature Learning

256×256×6 (RGB+HSV)
Deep Tree Networks (DTN)

256×256×6 (RGB+HSV)
Deep Tree Networks (DTN)

\[ \varphi(x) < 0 \]
Deep Tree Networks (DTN)

256×256×6
(RGB+HSV)

φ(x) ≥ 0
Deep Tree Networks (DTN)

256×256×6 (RGB+HSV)

φ(x) ≥ 0
Deep Tree Networks (DTN)

256×256×6
(RGB+HSV)
Supervised Feature Learning

Classification

Binary Mask Regression

Leaf Node 32×32×40

Mask Map 32×32×1

Conv 40

Conv 40

vectorize

fc 500

fc 2

0/1

SFL 256×256×6 (RGB+HSV)
Supervised Feature Learning

Binary Mask Regression
Training TRU

256×256×6 (RGB+HSV)

Feature Space
Training TRU
Tree Routing Unit (TRU)

- Routing Function
  \[ \varphi(x) = (x - \mu)^T \cdot v, \quad \|v\| = 1 \]

- Based on eigen-analysis of visiting set \( \bar{X}_S = X_S - \mu \)
  \[ \bar{X}_S^T \bar{X}_S v = \lambda v \]

- We optimize:
  \[ \arg \max_{v, \theta} \lambda = \arg \max_{v, \theta} v^T \bar{X}_S^T \bar{X}_S v \]
## Results

- **Evaluation Metrics:** ACER (the lower the better)

<table>
<thead>
<tr>
<th>Methods</th>
<th>Replay</th>
<th>Print</th>
<th>Mask Attacks</th>
<th>Makeup Attacks</th>
<th>Partial Attacks</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM+LBP¹</td>
<td>20.6</td>
<td>18.4</td>
<td>31.3</td>
<td>21.4</td>
<td>45.5</td>
<td>11.6</td>
</tr>
<tr>
<td>Auxiliary²</td>
<td>16.8</td>
<td>6.9</td>
<td>19.3</td>
<td><strong>14.9</strong></td>
<td>52.1</td>
<td>8.0</td>
</tr>
<tr>
<td>Ours</td>
<td><strong>9.8</strong></td>
<td><strong>6.0</strong></td>
<td><strong>15.0</strong></td>
<td>18.7</td>
<td><strong>36.0</strong></td>
<td><strong>4.5</strong></td>
</tr>
</tbody>
</table>

ACER = (Spoof Error Rate (APCER) + Live Error Rate (BPCER))/2


### Results

- **Evaluation Metrics:** EER (the lower the better)

<table>
<thead>
<tr>
<th>Methods</th>
<th>Replay</th>
<th>Print</th>
<th>Mask Attacks</th>
<th>Makeup Attacks</th>
<th>Partial Attacks</th>
<th>Avg.</th>
</tr>
</thead>
<tbody>
<tr>
<td>SVM+LBP</td>
<td>20.8</td>
<td>18.6</td>
<td>36.3</td>
<td>21.4</td>
<td>37.2</td>
<td>7.5</td>
</tr>
<tr>
<td>Auxiliary</td>
<td>14.0</td>
<td>4.3</td>
<td>11.6</td>
<td>12.9</td>
<td>24.6</td>
<td>7.8</td>
</tr>
<tr>
<td>Ours</td>
<td>10.0</td>
<td>2.1</td>
<td>14.4</td>
<td>18.6</td>
<td>26.5</td>
<td>5.7</td>
</tr>
</tbody>
</table>


t-SNE Results
Outline

• Training-Testing difference
• Alternative/Additional Sensors
• Practical Tips
• Future
Light Reflection

• Skin and spoof material have different reflection properties:
  • Reflectance
  • 3D shape

Liu et. al., Aurora guard: real-time face anti-spoofing via light reflection, arXiv 2019
Additional Sensors

• NIR
  • Human skin has different reflectance compared with spoof material

• Depth

• Thermal

• Multi-modality
Others

- Light field
- Polarized camera
- Structured Light
  - NIR with specific pattern (iPhone X)
- ToF (Time of flight)
  - Multi-point distance measurement
Question for Additional Sensors

- Data << RGB Data
Outline

• Training/Testing difference
• Alternative/Additional Sensors
• Practical Tips
• Future
Data are Your Friend

• More data → better performance

• Data augmentation (session II)

• (Efficient, effective) data collection
Updating Systems

- Use current model to collect failure cases
- Add failure cases to training set to fine-tune the model
- Update the current model
- Repeat several times
Updating Systems

• Manage the training data, not just mix everything
  • Eg. Base data 80%, New data 20%
  • Add subclasses based on lighting, walking and etc
Image Quality is the Devil

- Image resolution
- JPEG compression
  - Check the image bitrate
- Dark environment → ISO noise
Image Quality is the Devil

- Image resolution
- JPEG compression
  - Check the image bitrate
- Dark environment → ISO noise
Outline

• Training/Testing difference
• Alternative/Additional Sensors
• Practical Tips
• Summary and Future
Unsolved Problems

- Training/Testing difference
- Explainablity
- New attacks
- Unknown attack
- Data and evaluation
Problem 1: Training-Testing Difference

• Cross-database testing performances are still poor
  • EER for intra-testing: ~ 0% – 5%
  • EER for inter-testing: ~ 15% - 50%

• Can we use few-shot learning to improve the cross-database testing?
Problem 2: Explainablity

• Spatial explainablity
• Temporal explainablity
• Spoofing process explainablity
• Research on camera and imaging
Problem 3: New Attacks

• Makeup attacks

• Counter attacks to current methods
  • 3D mask attacks with flashing light → rPPG methods
  • Adversarial attacks → Texture based methods
Problem 4: Unknown Attacks

• Similar situation to cross-database testing

• Can we leverage the knowledge from other unknown object detection tasks?

• Identity variations > anti-spoofing variation
Problem 5: Data and Evaluation

• Intra-testing protocols too easy
• Inter-testing protocols too hard
• Represent previous problems as the testing protocols
Summary

• What and why face anti-spoofing?

• Traditional methods

• Deep learning methods

• Unknown attacks

• Additional sensors

• Practical tips