SSPP-DAN: Deep Domain Adaptation Network for Face Recognition with Single Sample Per Person

Sungeun Hong, Woobin Im, Jongbin Ryu, Hyun S. Yang

AIM Lab, KAIST, South Korea
Motivation & Problem Definition
- Single sample per person (SSPP)
- Challenges in real-world face recognition

Proposed method
- Domain adaptation
- Face synthesis

Experiments
- New heterogeneous dataset
- LFW for SSPP
Motivation & Problem Definition
SSPP face recognition

- Face recognition using Single Sample Per Person (SSPP)
  - Identify or verify identities using **only one single gallery image**
  - Related to the recently attracted one-shot learning

Train (on single gallery)

Test (on probe images)

Illumination
(b)

Expression
(c)

Disguise
(d)

Multiple
(e)

Sample images of the AR database
SSPP face recognition

- Limitations of existing SSPP datasets
  - Lab controlled environment
  - Consistent shooting environment

Real-world SSPP Face recognition

Registration (Gallery)

A stable image like clear frontal mugshot
e.g., ID card or e-passport

Identification (Probe)

Unstable images including non-trivial variations
e.g., surveillance camera, web images

Variations:
camera sensor, blur, noise, pose, illumination
Challenges

1. **Heterogeneity of the shooting environments**
   - Gallery: *stable environment*
   - Probe: *highly unstable environment*

2. **Shortage of training samples**
   - Only one training sample per person is available
Proposed method
Real-world SSPP Face recognition

Challenges

1. Heterogeneity of the shooting environments
   - Gallery: stable environment
   - Probe: highly unstable environment
Real-world SSPP Face recognition

Challenges

1. Heterogeneity of the shooting environments
   - Gallery: stable environment
   - Probe: highly unstable environment
Domain Adaptation (DA)

- Adjust a model to a different target domain distribution starting from the source domain knowledge.

**Source domain**

**Target domain**

*With labels*

*Without labels*
Domain Adaptation (DA)

- Adjust a model to a different target domain distribution starting from the source domain knowledge.

Source domain w/ label
Domain Adaptation (DA)

- Adjust a model to a different target domain distribution starting from the source domain knowledge
Domain Adaptation (DA)

- Adjust a model to a different target domain distribution starting from the source domain knowledge.
Adjust a model to a different target domain distribution starting from the source domain knowledge.
Domain Adaptation (DA)

- Adjust a model to a different **target domain distribution** starting from the **source domain knowledge**

```
Source domain w/ label
\downarrow
Model
\downarrow
Fine classification
```
Domain Adaptation (DA)

- Adjust a model to a different target domain distribution starting from the source domain knowledge.
Domain Adaptation (DA)

- Adjust a model to a different **target domain distribution** starting from the **source domain knowledge**

![Diagram showing Domain Adaptation process]

- **Source domain w/ label**
- **Model**
- **Domain Adaptation**
- **Target domain w/o label**
- **Better result**
- **Fine classification**
Domain Adaptation (DA)

<table>
<thead>
<tr>
<th>Purpose</th>
<th>Train</th>
<th>Test</th>
</tr>
</thead>
<tbody>
<tr>
<td>Domain</td>
<td>Source</td>
<td>Target</td>
</tr>
<tr>
<td>Image condition</td>
<td>Stable</td>
<td>Unstable</td>
</tr>
<tr>
<td>Label</td>
<td>O</td>
<td>X</td>
</tr>
</tbody>
</table>

Source domain w/ label

Domain Adaptation

Fine classification

Target domain w/o label

Better result
Domain Adaptation (DA)

- Basic assumptions of DA
  - samples are **abundant** in each domain
  - sample distribution of each domain is **related but different**

**Lack of samples**

Registration (Gallery) → Domain Adaptation → Identification (Probe)
Generate virtual samples for lack of samples.
SSPP-DAN

- Image synthesis

Without Synthesis

```
Class1_s  Class1_s (virtual)  Class1_t
Class2_s  Class2_s (virtual)  Class2_t
```

Condition

Pose
- Image synthesis
  >> distribution of samples

Without Synthesis

With Synthesis
- Image synthesis
  >> distribution of samples >> Success of DA

Without Synthesis

With Synthesis

DA fails

DA succeeds
Real-world SSPP Face recognition

Challenges

1. **Heterogeneity of the shooting environments**
   - Gallery: *stable environment*
   - Probe: *highly unstable environment*

2. **Shortage of training samples**
   - Only one training sample per person is available
1. **Domain adaptation network**
   - From stable face domain (source) to unstable face domain (target)

2. **Face synthesis**
   - Generate virtual samples
1. Domain adaptation network with domain-adversarial training
   - Feature learning
   - Domain adaptation
   - Classifier learning

\[
L_C = \sum_{i \in S} L_C^i \\
L_D = \sum_{i \in S \cup T} L_D^i \\
L_F = \sum_{i \in S} L_C^i - \lambda \sum_{i \in S \cup T} L_D^i
\]

when update \( \theta_C \)
when update \( \theta_D \)
when update \( \theta_F \)

1. Domain adaptation network
   - From stable face domain (source) to unstable face domain (target)

2. Face synthesis
   - Generate virtual samples

1) Landmark detection
   - Supervised descent method
2) 2D → 3D mapping
3) Pose estimation
4) Image synthesis
   - (yaw: -80°~+80°, pitch: -10°~40°)
Domain Adaptation

- Data Feed
  - Source (with label): frontal images + synthesized images
  - Target (without label): surveillance camera images
Experiments
Heterogeneous dataset

(a) Shooting condition for the source (left) and target (center and right)

(b) Face regions from the source (leftmost) and target (the others)

Table 1: Dataset specification

<table>
<thead>
<tr>
<th>Domain</th>
<th>Source</th>
<th>Target</th>
</tr>
</thead>
<tbody>
<tr>
<td>Set</td>
<td>webcam</td>
<td>surveillance</td>
</tr>
<tr>
<td>Subjects</td>
<td>30</td>
<td>30</td>
</tr>
<tr>
<td>Samples</td>
<td>30</td>
<td>15,900</td>
</tr>
<tr>
<td>Pose</td>
<td>frontal</td>
<td>various</td>
</tr>
<tr>
<td>Condition</td>
<td>stable</td>
<td>unstable (blur, noise, illumination)</td>
</tr>
</tbody>
</table>
Table 2: Recognition rates (%) for different models and different training sets of the EK-LFH

<table>
<thead>
<tr>
<th>Model</th>
<th>Training set</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source only</td>
<td>S</td>
<td>39.22</td>
</tr>
<tr>
<td></td>
<td>S + S_v</td>
<td>37.15</td>
</tr>
<tr>
<td>DAN</td>
<td>S + T</td>
<td>31.11</td>
</tr>
<tr>
<td><strong>SSPP-DAN</strong></td>
<td><strong>S + S_v + T</strong></td>
<td><strong>58.53</strong></td>
</tr>
</tbody>
</table>

**Lower bound**

**Upper bound**

- Train on target: T_1
- S: Labeled webcam
- T: Unlabeled surveillance
- S_v: Virtual set from S
- T_1: Labeled surveillance

only using face synthesis

only using domain adaptation
### Table 2: Recognition rates (%) for different models and different training sets of the EK-LFH

<table>
<thead>
<tr>
<th>Model</th>
<th>Training set</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source only</td>
<td>S</td>
<td>39.22</td>
</tr>
<tr>
<td></td>
<td>S + Sᵥ</td>
<td>37.15</td>
</tr>
<tr>
<td>DAN</td>
<td>S + T</td>
<td>31.11</td>
</tr>
<tr>
<td>SSPP-DAN</td>
<td>S + Sᵥ + T</td>
<td>58.53</td>
</tr>
<tr>
<td>Semi DAN</td>
<td>S + T + T₁</td>
<td>67.28</td>
</tr>
<tr>
<td><strong>Semi SSPP-DAN</strong></td>
<td>S + Sᵥ + T + T₁</td>
<td><strong>72.08</strong></td>
</tr>
<tr>
<td>Train on target</td>
<td>T₁</td>
<td>88.31</td>
</tr>
</tbody>
</table>

- **S**: Labeled webcam
- **T**: Unlabeled surveillance
- **Sᵥ**: Virtual set from **S**
- **T₁**: Labeled surveillance

*Semi: 3 samples per person from target domain are revealed*
Labeled Faces in the Wild (LFW)

- **Dataset summary**
  - Rearranged LFW-A for SSPP face recognition
  - Gallery: 50 images for 50 people
  - Generic set: 108 subjects
Labeled Faces in the Wild (LFW)

- LFW for SSPP protocol
  - Gallery: 50 images for 50 people
  - Generic set: 108 subjects

<table>
<thead>
<tr>
<th>Method</th>
<th>Accuracy</th>
<th>Method</th>
<th>Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>DMMA [1]</td>
<td>17.8</td>
<td>RPR [20]</td>
<td>33.1</td>
</tr>
<tr>
<td>ESRC [7]</td>
<td>27.3</td>
<td>VGG-Face [8]</td>
<td>96.43</td>
</tr>
<tr>
<td>LGR [22]</td>
<td>30.4</td>
<td>Ours</td>
<td>97.91</td>
</tr>
</tbody>
</table>

Deep learning
1. Heterogeneity of the shooting environments

2. Shortage of training samples
1. Heterogeneity of the shooting environments
   ➢ Domain adaptation network

2. Shortage of training samples
   ➢ Face synthesis
Questions?
Appendix
Domain Adversarial Network

(Ganin, Yaroslav, and Victor Lempitsky. "Unsupervised domain adaptation by backpropagation." ICML 2015.)

- Unified framework using adversarial training
Domain Adversarial Network
(Ganin, Yaroslav, and Victor Lempitsky. "Unsupervised domain adaptation by backpropagation." ICML 2015.)

- Unified framework using adversarial training
Domain Adaptation

- **Adversarial Training**
  - \( F \) is trained to fool \( D \) so that \( D \) cannot determine domain of data.
  - Gradient Reversal Layer (GRL)
    - forward: identity operation \( R_{\lambda}(x) = x \)
    - backward: multiply by \(-\lambda\) \( \frac{dR_{\lambda}}{dx} = -\lambda I \)

- **Loss for training**

\[
\begin{align*}
L_C &= \sum_{i \in S} L^i_C & \text{when update } \theta_C \\
L_D &= \sum_{i \in S \cup T} L^i_D & \text{when update } \theta_D \\
L_F &= \sum_{i \in S} L^i_C - \lambda \sum_{i \in S \cup T} L^i_D & \text{when update } \theta_F
\end{align*}
\]

\( L^i_C \) and \( L^i_D \): loss of \( C \) and \( D \), and 
\( \theta_D, \theta_F, \theta_C \): parameters of \( D, F, C \)