Tabula Rasa: Model Transfer for Object Category Detection

Yusuf Aytar and Andrew Zisserman

Presenter: Tinghui Zhou

March 19, 2012
Motivation

- Traditional machine learning methods start from scratch when presented a new category.
- Many categories have few training data available. Distribution follows the power law (Salakhutdinov et al.).
Overview

**Training Target** (Bicycle)
- A Few Positive (Bicycle) Samples
- Many Negative Samples

**Training Source** (Motorbike)
- Positive Samples
- Negative Samples

**Transfer Training**
Learning Model Transfer SVMs using a source classifier for transfer regularization.

**Target Detector** (Bicycle)

**Motorbike Detector**
Adaptive SVM (A-SVM)

Objective of classic SVM (baseline):

\[
L = \min_{w, b} \|w\|^2 + C \sum_{i}^{N} loss(x_i, y_i; w, b)
\]

where \(loss(x_i, y_i; w, b) = \max(0, 1 - y_i(w^T x_i + b))\).

Objective of Adaptive SVM (A-SVM):

\[
L_A = \min_{w, b} \|w - \Gamma w^s\|^2 + C \sum_{i}^{N} loss(x_i, y_i; w, b)
\]
Role of $\Gamma$ in A-SVM

Objective of Adaptive SVM (A-SVM):

$$L_A = \min_{w, b} \| w - \Gamma w^s \|^2 + C \sum_{i}^{N} \text{loss}(x_i, y_i; w, b)$$

Assuming $w^s$ is normalized to 1, then

$$\| w - \Gamma w^s \|^2 = \| w \|^2 - 2\Gamma \| w \| \cos \theta + \Gamma^2$$

- $\Gamma$ defines tradeoff between maximizing the margin $\| w \|$ and knowledge transfer (minimizing $\theta$).
Better transfer regularization

- Increase the amount of transfer without penalizing margin maximization.
- Transfer by minimizing the projection of $w$ onto the separating hyperplane orthogonal to $w^s$. 
Projective Model Transfer SVM (PMT-SVM)

\[ L_{PMT} = \min_{w, b} \|w\|^2 + \Gamma \|Pw\|^2 + C \sum_{i=1}^{N} \text{loss}(x_i, y_i; w, b) \]

s.t. \( w^T w^s \geq 0 \)

where \( P = I - \frac{w^s w^s^T}{w^s^T w^s} \), and \( \|Pw\|^2 = \|w\|^2 \sin^2 \theta \) is the squared norm of the projection of \( w \) onto the source hyperplane.
Deformable Adaptive SVM (DA-SVM)

\[
\min_{f, w, b} \| w - \Gamma \tau(w^s) \|^2 + C \sum_{i}^{N} l(x_i, y_i; w, b) \\
+ \lambda \left( \sum_{i \neq j}^{M,M} f_{ij}^2 d_{ij} + \sum_{i}^{M} (1 - f_{ii})^2 d \right)
\]

**Deformation**

Deformed template \( \tau(w^s) \) is obtained as a flow from the source template \( w^s \).

\[
\tau(w^s)_i = \sum_{j}^{M} f_{ij} w^s_j
\]
Dataset

Evaluations are performed on PASCAL VOC 2007.

▶ Test sets:

1. PASCAL-COMPLETE - Complete PASCAL VOC 2007
2. PASCAL-500 - A subset of PASCAL VOC 2007 including all positives and up to a total of 500 negatives.

▶ Test procedures:

2. pascal-side-only: PASCAL VOC 2007 evaluation using only side view annotations as the ground truth.
One Shot Learning

- Learning using a single positive sample of the target class.
- Investigates the *higher start* benefit of transfer learning and how each method responds to varying the quality of the samples.

Examples of target (bicycle) training samples which are ranked using the source (motorbike) classifier.
One Shot Learning

- Higher AP and smaller variance than baseline SVM.
- PMT-SVM is highly sensitive to bad (low ranked) samples.
- Source ‘motorbike’ detector (44.7%)

<table>
<thead>
<tr>
<th>Ranks</th>
<th>Base. SVM</th>
<th>A-SVM</th>
<th>DA-SVM</th>
<th>PMT-SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>01-15</td>
<td>40.5 ± 07.2</td>
<td><strong>53.9 ± 04.2</strong></td>
<td>53.7 ± 04.3</td>
<td>53.5 ± 05.7</td>
</tr>
<tr>
<td>16-30</td>
<td>33.0 ± 13.5</td>
<td>52.5 ± 08.3</td>
<td>51.9 ± 08.8</td>
<td><strong>54.7 ± 05.7</strong></td>
</tr>
<tr>
<td>31-45</td>
<td>26.4 ± 13.3</td>
<td>47.1 ± 07.3</td>
<td>47.1 ± 07.6</td>
<td><strong>48.5 ± 08.7</strong></td>
</tr>
<tr>
<td>46-60</td>
<td>14.0 ± 09.3</td>
<td>42.4 ± 03.7</td>
<td><strong>42.5 ± 04.2</strong></td>
<td>27.8 ± 11.3</td>
</tr>
</tbody>
</table>

Source: motorbike(44.7%), Target: bicycle(70.1%), Test-set: PASCAL-500, Test-procedure: pascal-side-only

<table>
<thead>
<tr>
<th>Ranks</th>
<th>Base. SVM</th>
<th>A-SVM</th>
<th>DA-SVM</th>
<th>PMT-SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>01-15</td>
<td>15.0 ± 08.0</td>
<td>30.2 ± 04.0</td>
<td><strong>30.3 ± 03.5</strong></td>
<td>30.1 ± 05.4</td>
</tr>
<tr>
<td>16-30</td>
<td>07.2 ± 04.1</td>
<td>27.0 ± 04.2</td>
<td>27.0 ± 04.3</td>
<td><strong>27.7 ± 07.1</strong></td>
</tr>
<tr>
<td>31-45</td>
<td>07.2 ± 08.2</td>
<td><strong>24.1 ± 06.0</strong></td>
<td>23.8 ± 05.9</td>
<td>11.9 ± 10.3</td>
</tr>
</tbody>
</table>

Source: cow(26.1%), Target: horse(60.2%), Test-set: PASCAL-500, Test-procedure: pascal-side-only
Multiple Shot Learning

- Learning using multiple positive samples of the target class.
- The improvement of transfer learning decreases as the number of target samples increases.

Source: motorbike(16.9%), Target: bicycle(59.0%), Test-set: PASCAL-COMPLETE, Test-procedure: pascal-side-only
HOG templates

Learned bicycle detector

Source motorbike detector

Difference between target and source detector

A-SVM

PMT-SVM

DA-SVM
Multiple Shot for Ranked Samples

- Target samples are ranked according to the source detector.
Effect of changing $\Gamma$
Negative Transfer

- Learn a bicycle classifier using a horse classifier as the source.
- A-SVM and DA-SVM perform worse than the baseline.
- PMT-SVM still outperforms the baseline.

<table>
<thead>
<tr>
<th>#</th>
<th>Base. SVM</th>
<th>A-SVM</th>
<th>DA-SVM</th>
<th>PMT-SVM</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>26.9 ± 11.2</td>
<td>06.0 ± 09.4</td>
<td>06.2 ± 09.3</td>
<td>27.8 ± 08.1</td>
</tr>
<tr>
<td>2</td>
<td>48.4 ± 05.0</td>
<td>26.4 ± 05.0</td>
<td>27.8 ± 05.4</td>
<td>50.0 ± 06.0</td>
</tr>
<tr>
<td>3</td>
<td>46.9 ± 11.0</td>
<td>33.3 ± 12.7</td>
<td>33.5 ± 12.5</td>
<td>51.4 ± 12.9</td>
</tr>
<tr>
<td>4</td>
<td>48.2 ± 09.5</td>
<td>36.3 ± 14.7</td>
<td>36.0 ± 14.3</td>
<td>51.1 ± 12.7</td>
</tr>
<tr>
<td>5</td>
<td>52.5 ± 09.1</td>
<td>45.4 ± 13.0</td>
<td>45.5 ± 13.4</td>
<td>56.0 ± 09.3</td>
</tr>
</tbody>
</table>

Source: horse(0.9%), Target: bicycle(70.1%), Test-set: PASCAL-500, Test-procedure: pascal-side-only
Specialization

- Superior to subordinate category transfer
- Learn a ‘quadruped’ category detector from horse, cow and sheep categories as the source detector
- Specialize (transfer) to horse detector from the ‘quadruped’ detector

<table>
<thead>
<tr>
<th></th>
<th>1</th>
<th>3</th>
<th>5</th>
<th>7</th>
<th>10</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Number of Samples</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Test-procedure: pascal-side-only</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base. SVM</td>
<td>03.6 ± 03.8</td>
<td>14.3 ± 07.6</td>
<td>20.0 ± 09.0</td>
<td>25.0 ± 07.3</td>
<td>29.9 ± 04.3</td>
</tr>
<tr>
<td>A-SVM</td>
<td>21.2 ± 05.5</td>
<td>29.7 ± 06.0</td>
<td>30.9 ± 04.3</td>
<td>32.6 ± 04.7</td>
<td>35.3 ± 03.0</td>
</tr>
<tr>
<td>DA-SVM</td>
<td>20.9 ± 05.6</td>
<td>29.2 ± 06.0</td>
<td>31.5 ± 03.9</td>
<td>32.1 ± 04.4</td>
<td>36.6 ± 02.8</td>
</tr>
<tr>
<td><strong>Test-procedure: pascal-default</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Base. SVM</td>
<td>03.6 ± 03.6</td>
<td>10.3 ± 02.6</td>
<td>10.6 ± 01.8</td>
<td>12.7 ± 02.0</td>
<td>13.8 ± 03.3</td>
</tr>
<tr>
<td>A-SVM</td>
<td>11.5 ± 04.0</td>
<td>14.5 ± 03.2</td>
<td>13.8 ± 03.3</td>
<td>15.2 ± 03.4</td>
<td>16.0 ± 01.8</td>
</tr>
<tr>
<td>DA-SVM</td>
<td>11.3 ± 04.5</td>
<td>14.2 ± 03.4</td>
<td>13.6 ± 03.0</td>
<td>15.3 ± 03.0</td>
<td>16.2 ± 01.7</td>
</tr>
</tbody>
</table>
Dataset Transfer

- Transfer the Motorbike detector in PASCAL to Caltech101.
- Again, the boost of transfer learning is not significant.

<table>
<thead>
<tr>
<th># Samples</th>
<th>1</th>
<th>6</th>
<th>11</th>
<th>16</th>
<th>21</th>
<th>26</th>
</tr>
</thead>
<tbody>
<tr>
<td>Base. SVM</td>
<td>88.4 ± 0.56</td>
<td>89.3 ± 0.33</td>
<td>88.9 ± 0.45</td>
<td>89.2 ± 0.22</td>
<td>89.0 ± 0.10</td>
<td>89.1 ± 0.11</td>
</tr>
<tr>
<td>A-SVM</td>
<td>86.8 ± 1.10</td>
<td>89.4 ± 0.49</td>
<td><strong>89.4 ± 0.27</strong></td>
<td><strong>89.5 ± 0.16</strong></td>
<td>89.1 ± 0.09</td>
<td>89.2 ± 0.08</td>
</tr>
<tr>
<td>PMT-SVM</td>
<td><strong>88.5 ± 0.69</strong></td>
<td>89.3 ± 0.37</td>
<td>89.0 ± 0.33</td>
<td>89.3 ± 0.23</td>
<td>89.0 ± 0.11</td>
<td>89.1 ± 0.10</td>
</tr>
<tr>
<td>DA-SVM</td>
<td>86.9 ± 1.15</td>
<td><strong>89.4 ± 0.47</strong></td>
<td>89.3 ± 0.32</td>
<td>89.5 ± 0.17</td>
<td><strong>89.1 ± 0.08</strong></td>
<td><strong>89.2 ± 0.07</strong></td>
</tr>
</tbody>
</table>
Dataset Transfer

- **Source detector**: PASCAL Motorbike
- **Target detector**: Caltech101 Motorbike
Dataset Transfer

- **Source detector**: PASCAL Bicycle
- **Target detector**: Caltech101 Motorbike