CrDoCo: Pixel-level Domain Transfer with Cross-Domain Consistency

Yun-Chun Chen¹,²  Yen-Yu Lin¹  Ming-Hsuan Yang³,⁴  Jia-Bin Huang⁵

¹Academia Sinica  ²National Taiwan University  ³UC Merced  ⁴Google  ⁵Virginia Tech

http://bit.ly/CrDoCo

Unsupervised domain adaptation

Input: A source dataset (labeled) and a target dataset (unlabeled)
Goal: Transfer knowledge from source to target domains

Training

Our approach

Experimental results

### A) Synthetic-to-real adaptation

<table>
<thead>
<tr>
<th>Method</th>
<th>Abs. Rel.</th>
<th>Sq. Rel.</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours</td>
<td>0.386</td>
<td>0.364</td>
<td>1.024</td>
</tr>
</tbody>
</table>

### B) Cross-city adaptation

<table>
<thead>
<tr>
<th>Method</th>
<th>Cityscapes</th>
<th>Cross-city</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours</td>
<td>50.4</td>
<td>914</td>
</tr>
</tbody>
</table>

### C) Visual results

- Target domain is unsupervised
- Domain gap between the source and target datasets
- Images of different styles should have the same task predictions

Challenges

Cross-domain consistency

- Images of different styles should have the same task predictions

Download our source code at: https://yunchunchen.github.io/CrDoCo

The domain-specific task network takes images of the target domain as input and produces consistent predictions. We show the applicability of our approach to multiple different tasks in the unsupervised domain adaptation setting. The image translation network learns to translate images of different styles should have the same task predictions. Challenges often do not capture pixel-level domain shifts that are not predictable. We exploit this property and introduce a cross-domain consistency loss that explicitly penalizes inconsistencies (e.g., semantic segmentation as shown in this example). We achieve state-of-the-art performance against existing unsupervised domain adaptation techniques.

Optical flow estimation

Cross-City

RMSE

<table>
<thead>
<tr>
<th>Method</th>
<th>Cityscapes</th>
<th>Cross-city</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours</td>
<td>0.377</td>
<td>0.272</td>
</tr>
</tbody>
</table>

Semantic segmentation

Single-view depth prediction

<table>
<thead>
<tr>
<th>Method</th>
<th>Abs. Rel.</th>
<th>Sq. Rel.</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours</td>
<td>0.254</td>
<td>0.257</td>
<td>0.866</td>
</tr>
</tbody>
</table>

**Consistency loss:**

The consistency loss encourages the source and target domain predictions to be similar.

**Unsupervised domain adaptation:**

The domain gap between the source and target datasets is utilized.

**Cross-domain adaptation:**

Models are trained on the source domain and adapted to the target domain without paired data.