Visual Domain Adaptation Challenge 2017: Semantic Segmentation

Yiheng Zhang, Zhaofan Qiu, Ting Yao, Tao Mei

Multimedia Search and Mining Group @ Microsoft Research Asia
Task:
Domain Adaptation for Semantic Segmentation

Source Domain: labeled data
Target Domain: unlabeled data

Training
Adaptation
Outline

• Frame level domain adaptation
  • Explore visual similarity between domains

• Feature level domain adaptation
  • Explore domain invariant representation

• Detailed Implementation for semantic segmentation
  • ResNet + PSP + Multi-scale Test + Ensemble
Outline

• Frame level domain adaptation
  • Explore visual similarity between domains

• Feature level domain adaptation
  • Explore domain invariant representation

• Detailed Implementation for semantic segmentation
  • ResNet + PSP + Multi-scale Test + Ensemble
Frame level domain adaptation

• Observation:
  • Visually difference between synthetic frames and real frames

• Solution:
  • Explicitly adapt frames across domains
  • Draw frames with domain and semantic constraints
Outline

• Frame level domain adaptation
  • Explore visual similarity between domains

• Feature level domain adaptation
  • Explore domain invariant representation

• Detailed Implementation for semantic segmentation
  • ResNet + PSP + Multi-scale Test + Ensemble
Feature level domain adaptation

Adapted Source-domain Frame

Adapted Target-domain Frame

Fully Convolutional Network

Supervised Training

Pixel-wise Classifier

Domain Discriminator

Adversarial Training

Source Domain

Target Domain
Outline

• Frame level domain adaptation
  • Explore visual similarity between domains

• Feature level domain adaptation
  • Explore domain invariant representation

• Detailed Implementation for semantic segmentation
  • ResNet + PSP + Multi-scale Test + Ensemble
Detailed Implementation

• ImageNet pre-trained ResNet:
  • ResNet-101 [K. He et al., CVPR’16]
  • ResNet-152 [K. He et al., CVPR’16]
  • SE-ResNeXt-101 [J. Hu et al., arXiv’17]

• Pyramid Spatial Pooling (PSP) [H. Zhao et al., CVPR’17]

• Multi-scale Test:
  • Scales: [0.50, 0.75, 1.00, 1.25, 1.50, 1.75]

• Ensemble
Domain Adaptation Results

Original Frame  Before Adaptation  After Adaptation
Domain Adaptation Results

Original Frame | Before Adaptation | After Adaptation
Evaluations

• Dataset
  • Training Domain: GTA5, 24,966 labeled synthetic frames
  • Validation Domain: Cityscapes, 500 labeled real frames
  • Test Domain: 1,500 unlabeled real frames

Evaluation Results on Test Domain

<table>
<thead>
<tr>
<th>Model</th>
<th>road</th>
<th>sdwlk</th>
<th>blng</th>
<th>wall</th>
<th>fence</th>
<th>pole</th>
<th>light</th>
<th>sign</th>
<th>vgttn</th>
<th>trrn</th>
<th>sky</th>
<th>person</th>
<th>rider</th>
<th>car</th>
<th>truck</th>
<th>bus</th>
<th>train</th>
<th>mcycl</th>
<th>bcycl</th>
<th>mIoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>ResNet-101</td>
<td>84.5</td>
<td>36.9</td>
<td>72.9</td>
<td>15.8</td>
<td>23.3</td>
<td>39.4</td>
<td>41.8</td>
<td>36.8</td>
<td>67.1</td>
<td>25.2</td>
<td>89.1</td>
<td>50.5</td>
<td>20.6</td>
<td>77.8</td>
<td>22.1</td>
<td>24.3</td>
<td>22.8</td>
<td>28.5</td>
<td>37.9</td>
<td>43.0</td>
</tr>
<tr>
<td>ResNet-152</td>
<td><strong>87.4</strong></td>
<td>38.3</td>
<td>73.4</td>
<td>20.0</td>
<td>29.7</td>
<td>39.9</td>
<td>42.2</td>
<td><strong>36.9</strong></td>
<td>70.9</td>
<td>30.7</td>
<td>89.4</td>
<td>53.3</td>
<td>22.4</td>
<td>79.0</td>
<td><strong>23.5</strong></td>
<td><strong>25.6</strong></td>
<td>11.6</td>
<td>30.2</td>
<td>41.1</td>
<td>44.5</td>
</tr>
<tr>
<td>SE-ResNeXt-101</td>
<td>83.4</td>
<td>34.9</td>
<td>72.3</td>
<td>20.8</td>
<td>27.7</td>
<td>39.3</td>
<td>39.3</td>
<td>34.6</td>
<td>67.5</td>
<td>27.1</td>
<td>89.4</td>
<td>49.6</td>
<td>21.7</td>
<td>77.5</td>
<td>21.5</td>
<td>25.1</td>
<td>12.0</td>
<td>25.0</td>
<td>42.1</td>
<td>42.7</td>
</tr>
<tr>
<td>Ensemble</td>
<td>87.0</td>
<td><strong>38.5</strong></td>
<td><strong>74.7</strong></td>
<td><strong>23.7</strong></td>
<td><strong>30.5</strong></td>
<td><strong>41.1</strong></td>
<td><strong>45.2</strong></td>
<td><strong>36.9</strong></td>
<td><strong>72.1</strong></td>
<td><strong>32.6</strong></td>
<td><strong>90.4</strong></td>
<td><strong>55.9</strong></td>
<td><strong>26.8</strong></td>
<td><strong>80.0</strong></td>
<td>23.4</td>
<td>25.1</td>
<td><strong>28.7</strong></td>
<td><strong>44.6</strong></td>
<td><strong>46.0</strong></td>
<td><strong>47.5</strong></td>
</tr>
</tbody>
</table>
Thanks !