Adversarial Domain Adaptation for Semantic Segmentation

Wei-Chih Hung¹, Yi-Hsuan Tsai², Ming-Hsuan Yang¹
¹UC Merced, ²NEC Labs America
VisDA 2017 Challenge

Source Dataset: GTA5

Target Dataset (Validation): Cityscapes

Target Dataset (Challenge)
Motivation

Input images are different

Idea: make the predictions on both datasets look similar

Ground truth labels’ layout distributions are similar
Adversarial Learning for Domain Adaptation

Source Prediction

Map

{Source, Target}

Source

Target

Segmentation Network

Softmax Outputs

L_adv (Target)

L_seg (Source)
Implementation Details

• We use PyTorch

• Baseline model: DeepLab-v2 without multi-scale
  • ResNet-101
  • Pretrained only with ImageNet
  • ~65% mean IOU on Cityscapes

• It is essential to balance:
  • Segmentation network and discriminator
  • L_seg and L_adv
Experimental Results

• GTA -> Cityscapes

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Adapt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean IOU</td>
<td>32.33</td>
<td>42.44</td>
</tr>
</tbody>
</table>

• GTA -> Test Set

<table>
<thead>
<tr>
<th></th>
<th>Baseline</th>
<th>Adapt</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean IOU</td>
<td>30.3</td>
<td>42.4</td>
</tr>
</tbody>
</table>
## Detailed Class Performance on Test Set

<table>
<thead>
<tr>
<th>#</th>
<th>User</th>
<th>Team Name</th>
<th>road</th>
<th>sdwlk</th>
<th>bldng</th>
<th>wall</th>
<th>fence</th>
<th>pole</th>
<th>light</th>
<th>sign</th>
<th>vgttn</th>
<th>trrn</th>
</tr>
</thead>
<tbody>
<tr>
<td>3</td>
<td>whung</td>
<td>VLLAB</td>
<td>src</td>
<td>adapt</td>
<td>s</td>
<td>a</td>
<td>s</td>
<td>a</td>
<td>s</td>
<td>a</td>
<td>s</td>
<td>a</td>
</tr>
<tr>
<td>3</td>
<td>whung</td>
<td>VLLAB</td>
<td>sky</td>
<td>person</td>
<td>rider</td>
<td>car</td>
<td>truck</td>
<td>bus</td>
<td>train</td>
<td>mcycl</td>
<td>bcycl</td>
<td>MeanIoU</td>
</tr>
<tr>
<td>3</td>
<td>whung</td>
<td>VLLAB</td>
<td>70.0</td>
<td>90.3</td>
<td>43.2</td>
<td>47.0</td>
<td>23.0</td>
<td>35.7</td>
<td>63.5</td>
<td>78.0</td>
<td>10.6</td>
<td>24.8</td>
</tr>
</tbody>
</table>

- Improve 17 classes
- 11 classes have improvement over 10%
- 2 classes (turn light, motorcycle) perform a bit worse
Visualization
Visualization

baseline

adapt
Visualization

baseline

adapt
Conclusions

• **Adversarial learning** can help domain adaptation without any hand-crafted criterions

• Our designed model is **end-to-end, one-stage** training, and can be adapted to other segmentation networks

• During inference, there is **no extra computation**