Abstract

The semantic understanding of urban scenes is one of the key components for autonomous driving systems. Deep neural networks require huge amount of labeled data, which is difficult and expensive to acquire. A recent workaround is to exploit synthetic data but differences between real and synthetic scenes limit the performance. We propose an unsupervised domain adaptation strategy from a synthetic supervised training to real data. Experimental results demonstrate that the proposed approach is able to adapt a network trained on synthetic datasets to a real one.

Proposed Approach

\[ \mathcal{L}_{G,1} = \text{standard cross-entropy loss (on source dataset)} \]

\[ \text{Trained end-to-end minimizing } \mathcal{L}_{\text{full}} = \mathcal{L}_{G,1} + w_{s,t}^\alpha \mathcal{L}_{G,2}^s + w_{t} \mathcal{L}_{G,3} \]

Dataset

**SOURCE (SYNTHETIC)**

- GTA
  - ~25k images
  - High quality
  - Car viewpoints

- SYNTHIA
  - ~9k images
  - Medium quality
  - Different viewpoints

**TARGET (REAL WORLD)**

- CITYSCAPES
  - ~3k images
  - Car viewpoints

Adversarial Training

\[ \mathcal{L}_{s,t}^s = -\log(D(G(X^{s,t}_n))) \]

\[ \mathcal{L}_D = -\log(1 - D(G(X^{s,t}_n))) + \log(D(Y^s_n)) \]

- \( s \): source dataset
- \( t \): target dataset

Self-Taught Loss

Predictions of \( G \) are more reliable where \( D \) marks them as GT with high accuracy

\[ \mathcal{L}_{G,3} = \frac{1}{n} \sum_{c=1}^{C} \mathbb{Y}_c \mathbb{c} \cdot \log \left( \mathbb{G}(X^{s}_n) \mathbb{c} \right) \]

- \( c \): classes
- Threshold weighting on confidence maps from \( D \)

Results

<table>
<thead>
<tr>
<th>From GTA</th>
<th>road</th>
<th>sidewalk</th>
<th>building</th>
<th>wall</th>
<th>fence</th>
<th>pole</th>
<th>light</th>
<th>sign</th>
<th>terrain</th>
<th>sky</th>
<th>person</th>
<th>rider</th>
<th>car</th>
<th>truck</th>
<th>bus</th>
<th>train</th>
<th>motorcycle</th>
<th>bike</th>
<th>mIoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours (( \mathcal{L}_{G,1} ) only)</td>
<td>45.3</td>
<td>20.6</td>
<td>50.1</td>
<td>9.3</td>
<td>12.7</td>
<td>19.5</td>
<td>4.3</td>
<td>0.7</td>
<td>81.9</td>
<td>21.1</td>
<td>63.3</td>
<td>52.0</td>
<td>1.7</td>
<td>77.9</td>
<td>26.0</td>
<td>39.8</td>
<td>0.1</td>
<td>4.7</td>
<td>0.0</td>
</tr>
<tr>
<td>Ours (( \mathcal{L}_{\text{full}} )) [1]</td>
<td>54.9</td>
<td>23.8</td>
<td>50.9</td>
<td>16.2</td>
<td>11.2</td>
<td>20.0</td>
<td>3.2</td>
<td>0.0</td>
<td>79.7</td>
<td>31.6</td>
<td>64.9</td>
<td>52.5</td>
<td>7.9</td>
<td>79.5</td>
<td>27.2</td>
<td>41.8</td>
<td>0.5</td>
<td>10.7</td>
<td>1.3</td>
</tr>
<tr>
<td>Hung et al. [2]</td>
<td>81.7</td>
<td>0.3</td>
<td>68.4</td>
<td>4.5</td>
<td>2.7</td>
<td>8.5</td>
<td>0.6</td>
<td>0.0</td>
<td>82.7</td>
<td>21.5</td>
<td>67.9</td>
<td>40.0</td>
<td>3.3</td>
<td>80.7</td>
<td>34.2</td>
<td>45.9</td>
<td>0.2</td>
<td>8.7</td>
<td>0.0</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>From SYNTHIA</th>
<th>road</th>
<th>sidewalk</th>
<th>building</th>
<th>wall</th>
<th>fence</th>
<th>pole</th>
<th>light</th>
<th>sign</th>
<th>terrain</th>
<th>sky</th>
<th>person</th>
<th>rider</th>
<th>car</th>
<th>truck</th>
<th>bus</th>
<th>train</th>
<th>motorcycle</th>
<th>bike</th>
<th>mIoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ours (( \mathcal{L}_{G,1} ) only)</td>
<td>10.3</td>
<td>20.5</td>
<td>35.5</td>
<td>1.5</td>
<td>0.0</td>
<td>28.9</td>
<td>0.0</td>
<td>1.2</td>
<td>83.1</td>
<td>74.8</td>
<td>53.5</td>
<td>7.5</td>
<td>65.8</td>
<td>18.1</td>
<td>4.7</td>
<td>1.0</td>
<td>25.4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ours (( \mathcal{L}_{\text{full}} )) [1]</td>
<td>78.4</td>
<td>0.1</td>
<td>73.2</td>
<td>0.0</td>
<td>0.0</td>
<td>16.9</td>
<td>0.0</td>
<td>0.2</td>
<td>84.3</td>
<td>78.8</td>
<td>46.0</td>
<td>0.3</td>
<td>74.9</td>
<td>30.8</td>
<td>0.0</td>
<td>0.1</td>
<td>30.2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hung et al. [2]</td>
<td>72.5</td>
<td>0.0</td>
<td>63.8</td>
<td>0.0</td>
<td>0.0</td>
<td>16.3</td>
<td>0.0</td>
<td>0.5</td>
<td>84.7</td>
<td>76.9</td>
<td>45.3</td>
<td>1.5</td>
<td>77.6</td>
<td>31.3</td>
<td>0.0</td>
<td>0.1</td>
<td>29.4</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

References
