UNSUPERVISED DOMAIN ADAPTATION FOR URBAN SCENES SEGMENTATION
Biasetton M., Michieli U., Agresti G., Zanuttigh P. - University of Padova
(biasetto, michieli, agrestig, zanuttigh)@dei.unipd.it

Abstract
The semantic understanding of urban scenes is one of the key components for an autonomous driving system. Deep neural networks require to be trained with a huge amount of labeled data, which is difficult and expensive to acquire. A recently proposed workaround is the usage of synthetic data, however the differences between real world and synthetic scenes limit the performance. We propose an unsupervised domain adaptation strategy from a synthetic supervised training to real data exploiting three components: supervised learning on synthetic data, adversarial learning strategy and finally self-teaching strategy working on unlabeled data. Experimental results prove that the proposed approach is able to adapt a network trained on synthetic dataset to a real one.

Cross-Entropy Loss
\[ \mathcal{L}_{G,1} = - \sum_{c \in C} Y_n^s[c] \cdot \log \left( G(X_n^s)[c] \right) \]
s: source dataset
t: target dataset

Adversarial Training
\[ \mathcal{L}_{G,2}^{s,t} = - \log(D(G(X_n^{s,t}))) \]
\[ L_D = -\log(1 - D(G(X_n^{s,t}))) + \log(D(Y_n^t)) \]
\[ \mathcal{L}_D = -\log(1 - D(G(X_n^{s,t}))) + \log(D(Y_n^t)) \]
\( c \): classes

Self-Taught Loss
Predictions of \( G \) are more reliable where \( D \) marks them as GT with high accuracy
\[ \mathcal{L}_{G,3} = -T_s \cdot \sum_{c \in C} Y_n^s[c] \cdot \log \left( G(X_n^s)[c] \right) \]

Threshold on confidence maps from \( D \)

Qualitative Results
From GTA5
From SYNTHIA
RGB
annotation
baseline (\( \mathcal{L}_{G,1} \))
Hung et al. [2]
\( \mathcal{L}_{full} \) [1]

Quantitative Results
From GTA
<table>
<thead>
<tr>
<th>Class</th>
<th>mIoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>road</td>
<td>81.7</td>
</tr>
</tbody>
</table>

From SYNTHIA
<table>
<thead>
<tr>
<th>Class</th>
<th>mIoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>road</td>
<td>10.3</td>
</tr>
</tbody>
</table>

Hung et al. [2]
<table>
<thead>
<tr>
<th>Class</th>
<th>mIoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>road</td>
<td>10.3</td>
</tr>
</tbody>
</table>


Proposed Approach

Dataset

SOURCE (SYNTHETIC)
TARGET (REAL)

\( \mathcal{L}_{G,1} \)
\( \mathcal{L}_{G,2}^{s,t} \)
\( \mathcal{L}_D \)
\( \mathcal{L}_D \)

\( \mathcal{L}_{G,3} \)

\( \mathcal{L}_{full} \) [1]