Unsupervised Domain Adaptation in Semantic Segmentation via Orthogonal and Clustered Embeddings

Marco Toldo, Umberto Michieli, Pietro Zanuttigh

January 5-9th, 2021
Semantic Segmentation

**Task** → Assign each pixel of an image with a semantic label

- Deep learning as enabling factor
- Fully convolutional auto-encoders [1]

Unsupervised Domain Adaptation

Issues of FCNs:
1. Tons of training samples to avoid overfitting
2. Low generalization capability

Domain Adaptation → from a *label-abundant source* to a *label-scarce target* domain
- Unsupervised ⇒ Source supervision only
- Distribution alignment across domains (e.g. adversarial learning)
- Multiple adaptation levels [2]

Model Architecture - Overview

→ *Class-conditional* domain alignment of feature distribution

**Clustering:** group together features of similar semantics

**Orthogonality:** different active channels for distinct semantic classes

**Sparsity:** suppress weak activations

**Entropy minimization (EM):** push away features from decision boundaries
Model Architecture - Clustering

Clustering loss: $L_{cl} = \frac{1}{|F_n^{s,t}|} \sum_{f_i \in F_n^{s,t}} d(f_i, c_{\hat{y}_i}) - \frac{1}{|C|(|C|-1)} \sum_{j \in C} \sum_{k \in C, k \neq j} d(c_j, c_k)$

Class centroid: $c_j = \frac{\sum_{\hat{y}_i} \delta_{j,\hat{y}_i} f_i}{\sum_{\hat{y}_i} \delta_{j,\hat{y}_i}}$, $j \in C$

$\rightarrow$ Group together features of same class from both domains

Intra-class contraction

Inter-class spacing

Class centroids

Attractive force

Repulsive force

Source features

Target features

Source features

Target features

$\rightarrow$ Group together features of same class from both domains
Model Architecture - Orthogonality & Sparsity

- **Orthogonality loss:** \( \mathcal{L}_{or} = - \sum_{f_i \in F(X_n^{x,t})} \sum_{j \in \mathcal{C}} p_j(f_i) \log p_j(f_i) \)

  Similarity based distribution:
  \[ p_j(f_i) = \frac{e^{f_i, c_j}}{\sum_{k \in \mathcal{C}} e^{f_i, c_k}}, \quad j \in \mathcal{C} \]

  → Orthogonality on **distinct** classes, similarity within **same** class

- **Sparsity loss:** \( \mathcal{L}_{sp} = - \sum_{i \in \mathcal{C}} ||\bar{c}_i - \rho||_2^2 \)

  Indirect action over centroids → **class-wise** influence

  → Lower volume of active feature channels

\[ \mathcal{L}_{cl} + \mathcal{L}_{sp} + \mathcal{L}_{or} \]
Experiments: GTA → Cityscapes

Source domain: *synthetic* GTA dataset

Target domain: *real-world* Cityscapes dataset

[Image showing comparison between source and target images, with annotations and model predictions.]

<table>
<thead>
<tr>
<th>Method</th>
<th>mIoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Only</td>
<td>37.0</td>
</tr>
<tr>
<td>Tsai et al. [3] (feat)</td>
<td>39.3</td>
</tr>
<tr>
<td>MinEnt [4]</td>
<td>42.3</td>
</tr>
<tr>
<td>MaxSquare IW [5]</td>
<td>45.2</td>
</tr>
<tr>
<td>( \mathcal{L}<em>{cl} + \mathcal{L}</em>{or} + \mathcal{L}_{sp} )</td>
<td>45.3</td>
</tr>
<tr>
<td>( \mathcal{L}<em>{cl} + \mathcal{L}</em>{or} + \mathcal{L}_{sp} + \text{EM} )</td>
<td>45.9</td>
</tr>
</tbody>
</table>

Results for 19 classes, DeepLab-v2 with ResNet-101 backbone

Experiments: SYNTTHIA → Cityscapes

Source domain: synthetic SYNTTHIA dataset
Target domain: real-world Cityscapes dataset

Results for 13 classes, DeepLab-v2 with ResNet-101 backbone

<table>
<thead>
<tr>
<th>Method</th>
<th>mIoU</th>
</tr>
</thead>
<tbody>
<tr>
<td>Source Only</td>
<td>40.5</td>
</tr>
<tr>
<td>Tsai et al. [3] (feat)</td>
<td>40.8</td>
</tr>
<tr>
<td>MinEnt [4]</td>
<td>44.2</td>
</tr>
<tr>
<td>MaxSquare IW [5]</td>
<td>46.9</td>
</tr>
<tr>
<td>$\mathcal{L}<em>{cl} + \mathcal{L}</em>{or} + \mathcal{L}_{sp}$</td>
<td>44.2</td>
</tr>
<tr>
<td>$\mathcal{L}<em>{cl} + \mathcal{L}</em>{or} + \mathcal{L}_{sp} + \text{EM}$</td>
<td>48.2</td>
</tr>
</tbody>
</table>

References:
Conclusion

Feature space regularization in UDA for Semantic Segmentation

Our main contributions:

- **Feature clustering** for semantic segmentation
- **Orthogonality** and **sparsity** objectives to force a regular structure of the embedding space
- State-of-the-art results on feature-level non adversarial adaptation on two widely used benchmarks
Thank you for the attention!

Paper website: https://lttm.dei.unipd.it/paper_data/UDAclustering/
Code: https://github.com/LTTM/UDAclustering
Contact: toldomarco@dei.unipd.it