Unsupervised Domain Adaptation in Semantic Segmentation via Orthogonal and Clustered Embeddings

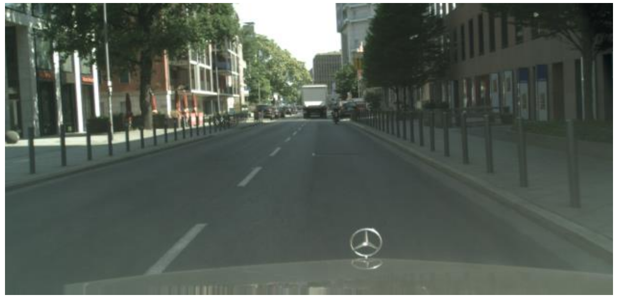
Marco Toldo, Umberto Michieli, Pietro Zanuttigh

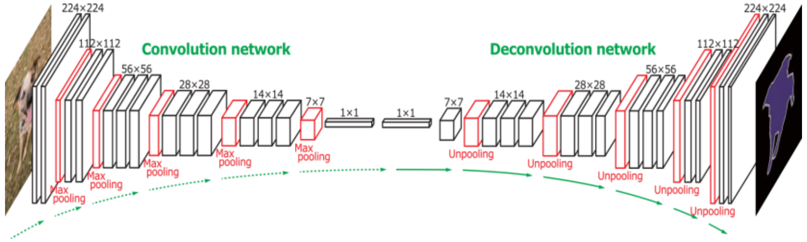
January 5-9th, 2021

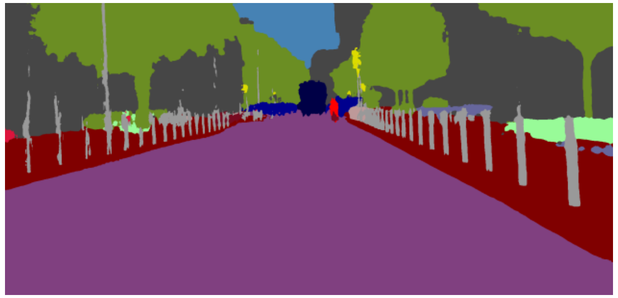
Semantic Segmentation

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

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







[1] J. Long et al. , “Fully convolutional networks for semantic segmentations”, *CVPR*, 2015.

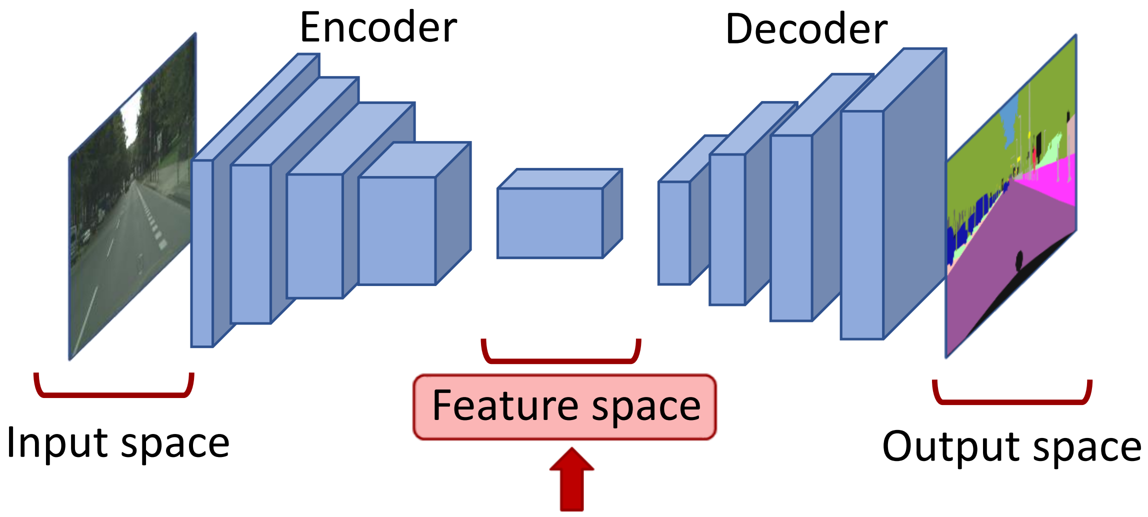
Unsupervised Domain Adaptation

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]



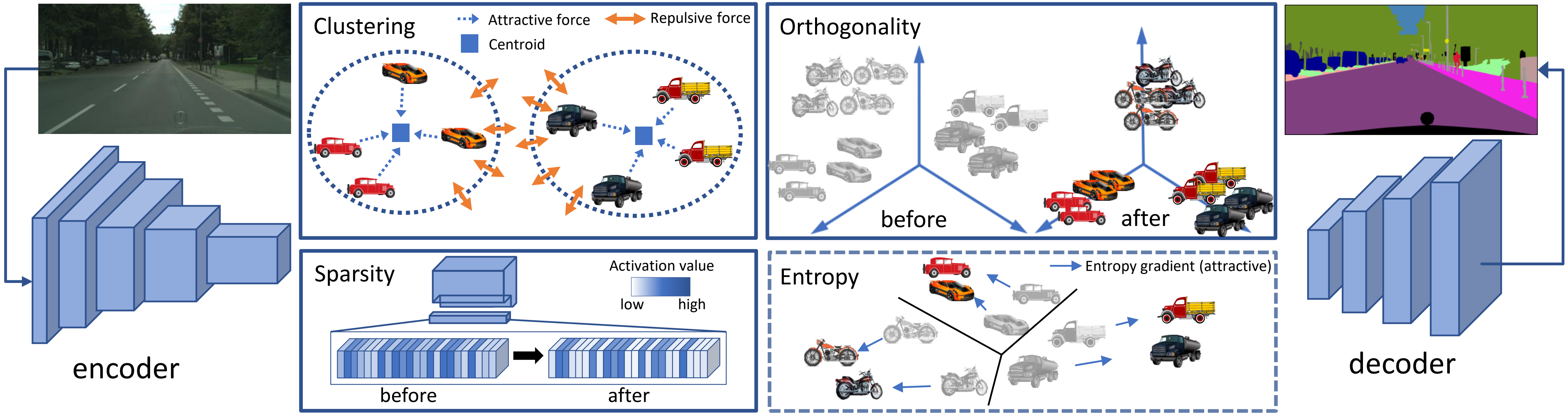
[2] M. Toldo et al. , "Unsupervised domain adaptation in semantic segmentation: a review", *Technologies*, 2020.

Model Architecture - Overview

→ *Class-conditional* domain alignment of feature distribution

Clustering: group together features of similar semantics

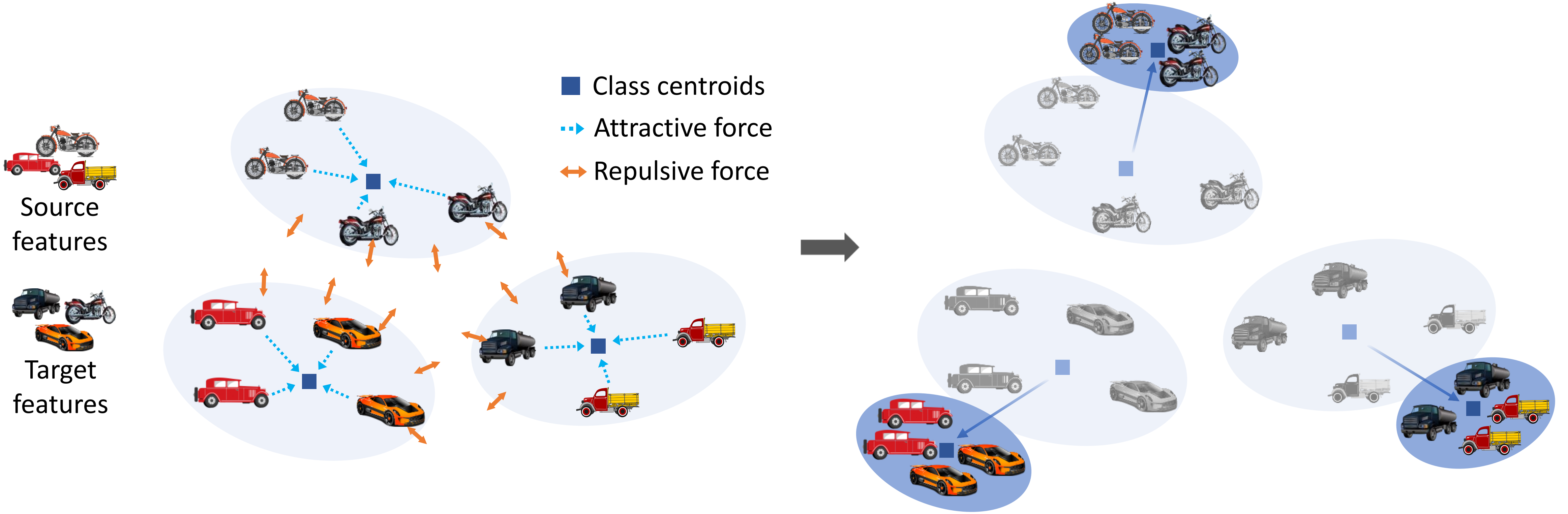
Orthogonality: different active channels for distinct semantic classes



Sparsity: suppress weak activations

Entropy minimization (EM)**:** push away features from decision boundaries

Model Architecture - Clustering



Clustering loss: $$\mathcal{L}_{cl} = \overbrace{\frac{1}{|\mathbf{F}_n^{s,t}|}\sum_{\substack{\mathbf{f}_i \in \mathbf{F}_n^{s,t} \\ \hat{y}_i \in \mathbf{S}_n^{s,t}}} d(\mathbf{f}_i, \mathbf{c}_{\hat{y}_i})}^{\text{Intra-class contraction}} - \overbrace{\frac{1}{|\mathcal{C}|(|\mathcal{C}|-1)}\sum_{j \in \mathcal{C}}\sum_{\substack{k \in \mathcal{C} \\ k \neq j}} \boxed{d}(\mathbf{c}_j, \mathbf{c}_k)}^{\text{Inter-class spacing}}$$

L2 norm

Class centroid: (barycenter) $$\mathbf{c}_j = \frac{\sum_{\mathbf{f}_i}\sum_{\hat{y}_i}\delta_{j,\hat{y}_i}\,\mathbf{f}_i}{\sum_{\hat{y}_i}\delta_{j,\hat{y}_i}}, \quad j \in \mathcal{C}$$

→ Group together features of *same class* from *both domains*

Model Architecture - Orthogonality & Sparsity

- **Orthogonality loss:** $$\mathcal{L}_{or} = -\sum_{\mathbf{f}_i \in F(\mathbf{X}_n^{s,t})} \sum_{j \in \mathcal{C}} p_j(\mathbf{f}_i) \log p_j(\mathbf{f}_i)$$

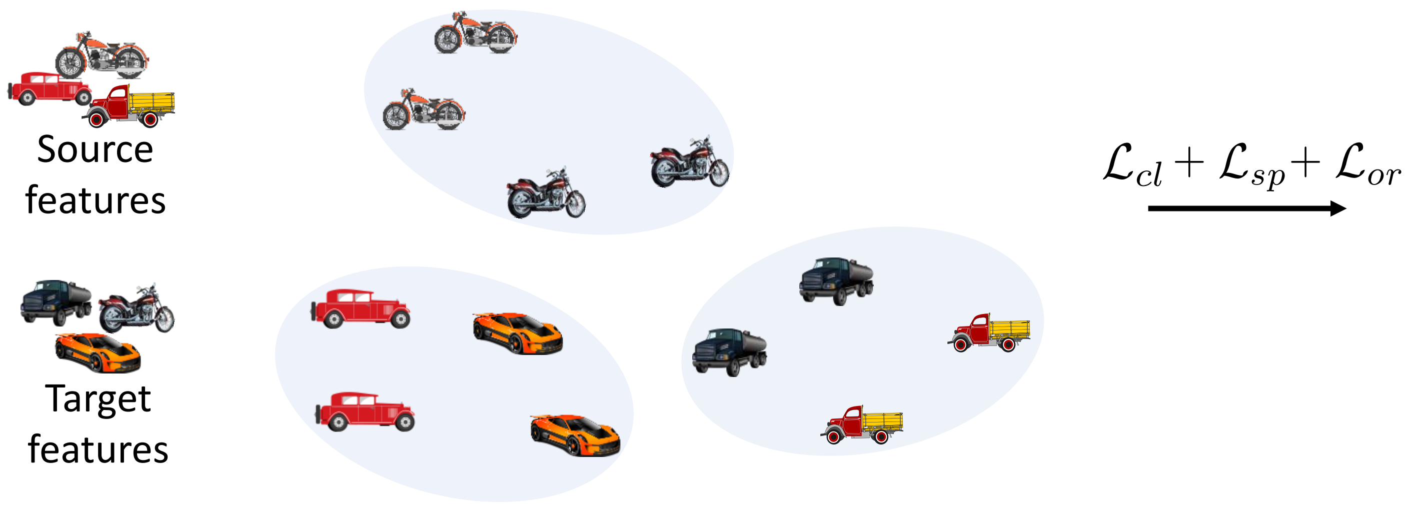
Similarity based distribution: $$p_j(\mathbf{f}_i) = \frac{e^{\langle \mathbf{f}_i, \mathbf{c}_j \rangle}}{\sum_{k \in \mathcal{C}} e^{\langle \mathbf{f}_i, \mathbf{c}_k \rangle}}, \quad j \in \mathcal{C}$$

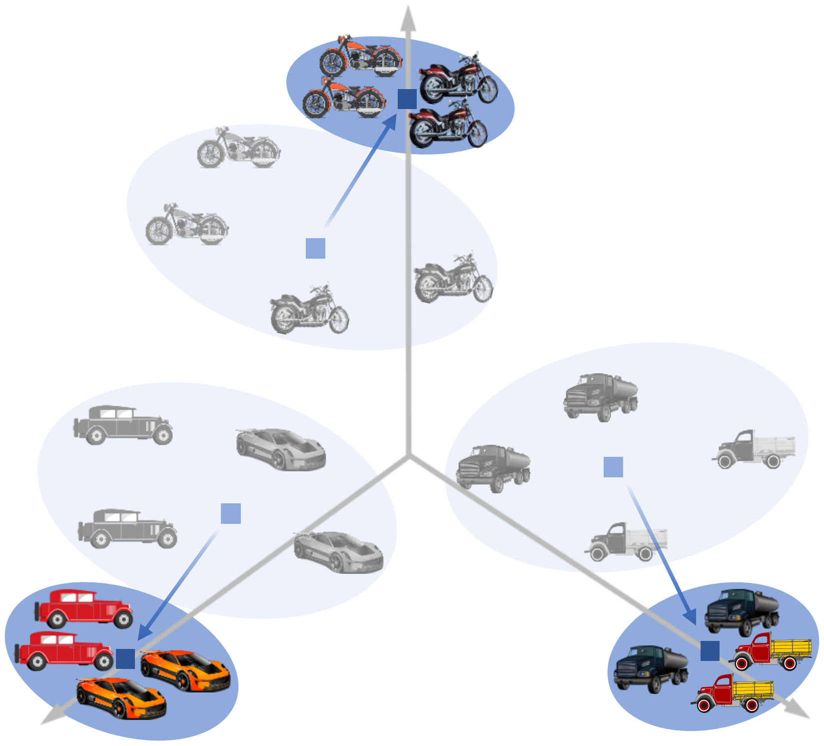
→ Orthogonality on *distinct* classes, similarity within *same* class

- **Sparsity loss:** $$\mathcal{L}_{sp} = -\sum_{i \in \mathcal{C}} ||\tilde{\mathbf{c}}_i - \boldsymbol{\rho}||_2^2$$

Indirect action over centroids → *class-wise* influence

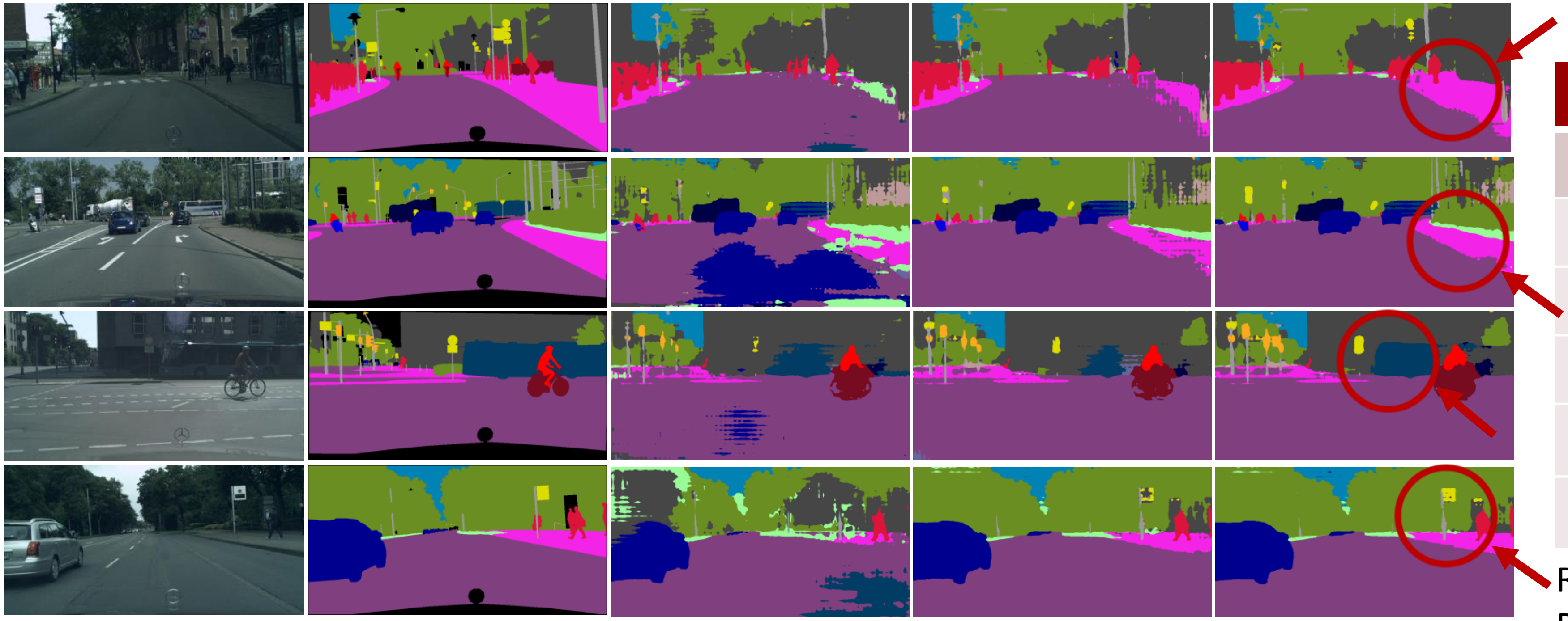
→ Lower volume of active feature channels





Experiments: GTA → Cityscapes

Source domain: *synthetic* GTA dataset **Target domain**: *real-world* Cityscapes dataset



Target Image | Annotation | Source Only | MaxSquare IW [5] | Ours (with EM)

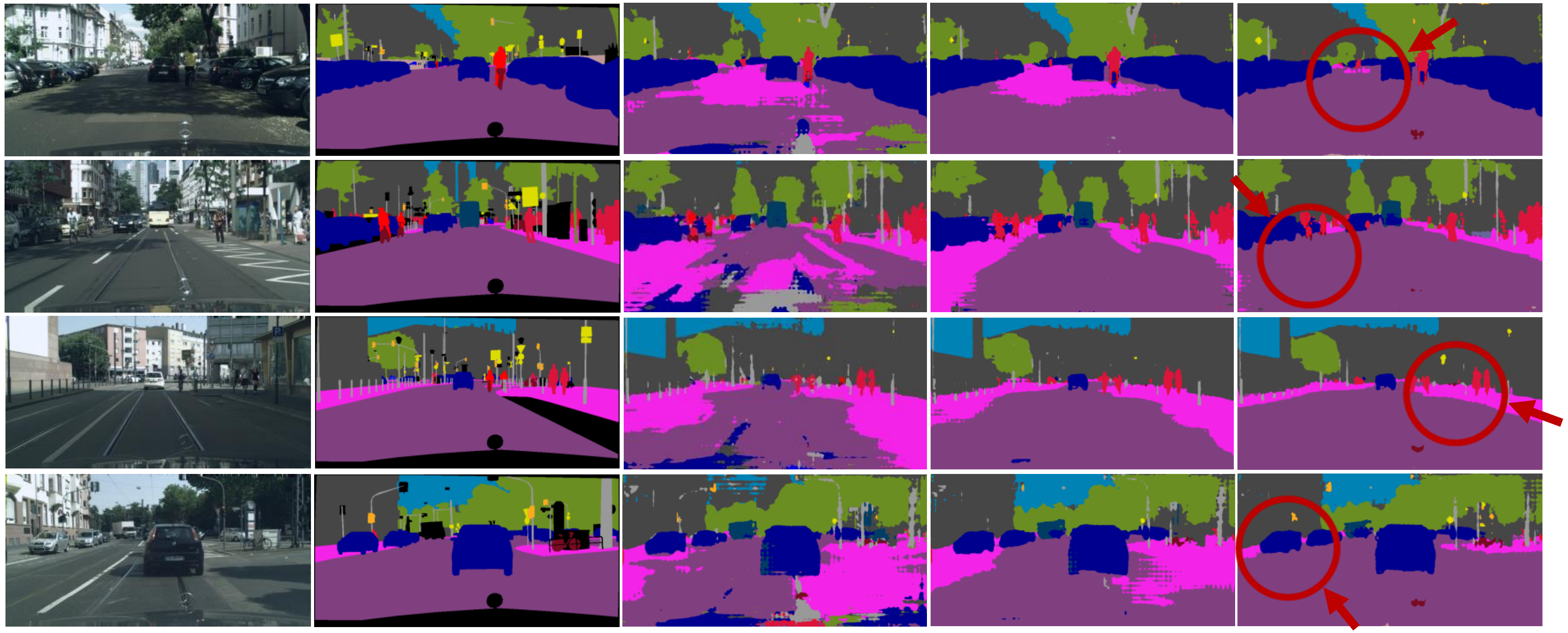
Method	mIoU
Source Only	37.0
Tsai et al. [3] (feat)	39.3
MinEnt [4]	42.3
MaxSquare IW [5]	45.2
$\mathcal{L}_{cl} + \mathcal{L}_{or} + \mathcal{L}_{sp}$	45.3
$\mathcal{L}_{cl} + \mathcal{L}_{or} + \mathcal{L}_{sp} + \mathrm{EM}$	**45.9**

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

[3] Y. Tsai et al., "Learning to adapt structured output space for semantic segmentation", In *CVPR*, 2018.
[4] T. Vu et al., "ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation ", In *CVPR*, 2019.
[5] M. Chen et al. , "Domain adaptation for semantic segmentation with maximum squares loss", In *ICCV*, 2019.

Experiments: SYNTHIA → Cityscapes

Source domain: *synthetic* SYNTHIA dataset **Target domain**: *real-world* Cityscapes dataset



Target Image | Annotation | Source Only | MaxSquare IW [5] | Ours (with EM)

Method	mIoU
Source Only	40.5
Tsai et al. [3] (feat)	40.8
MinEnt [4]	44.2
MaxSquare IW [5]	46.9
$\mathcal{L}_{cl} + \mathcal{L}_{or} + \mathcal{L}_{sp}$	44.2
$\mathcal{L}_{cl} + \mathcal{L}_{or} + \mathcal{L}_{sp} + \text{EM}$	**48.2**

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

[3] Y. Tsai et al., "Learning to adapt structured output space for semantic segmentation", In *CVPR*, 2018.
[4] T. Vu et al., "ADVENT: Adversarial Entropy Minimization for Domain Adaptation in Semantic Segmentation ", In *CVPR*, 2019.
[5] M. Chen et al. , "Domain adaptation for semantic segmentation with maximum squares loss", In *ICCV*, 2019.

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/

Arxiv: https://arxiv.org/abs/2011.12616

Code: https://github.com/LTTM/UDAclustering

Contact: toldomarco@dei.unipd.it