Continual Semantic Segmentation via Repulsion-Attraction of Sparse and Disentangled Latent Representations

Umberto Michieli and Pietro Zanuttigh - University of Padova

Università degli Studi di Padova

CVPR VIRTUAL JUNE 19-25

Abstract

Deep networks forget old tasks when learning new ones. We focus on *class incremental continual learning (CL) in semantic segmentation*. The proposed CL scheme shapes the latent space to reduce forgetting whilst improving the recognition of novel classes. Our method is driven by 3 novel components:

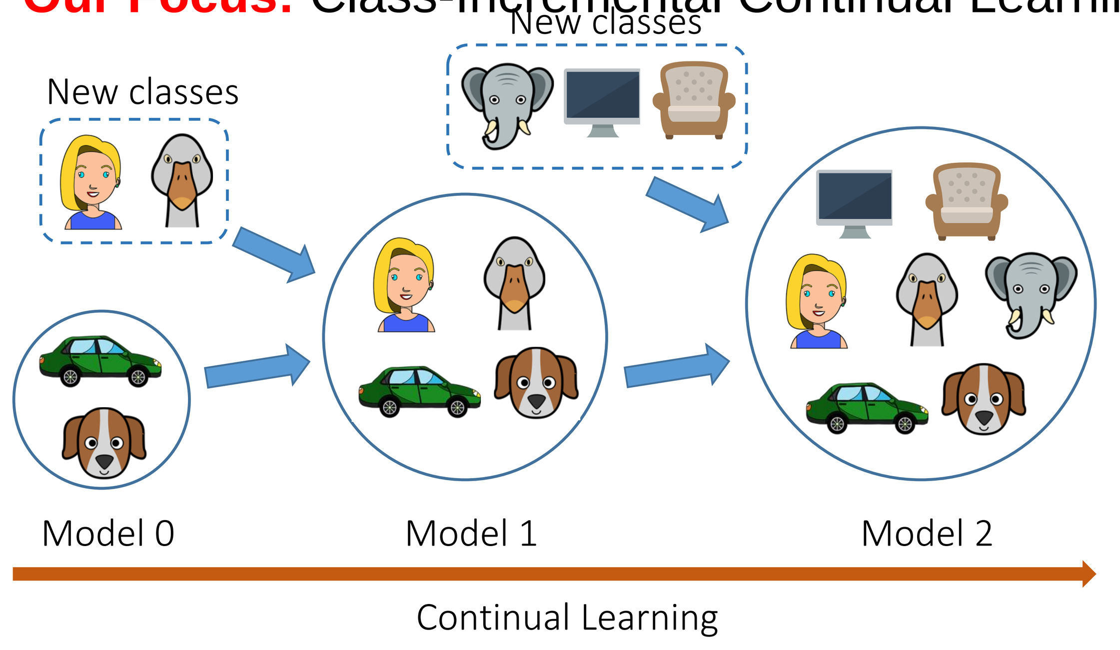
1) Prototype matching enforces latent space consistency, constraining the encoder to produce similar representations for old classes;

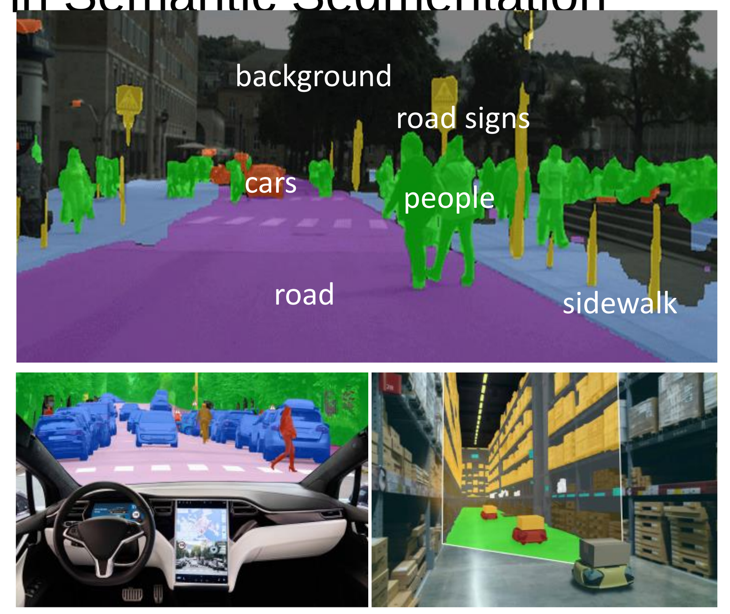
2) Features sparsity makes room in the latent space for new classes;

3) Contrastive learning clusters features according to their semantics while tearing apart those of different classes.

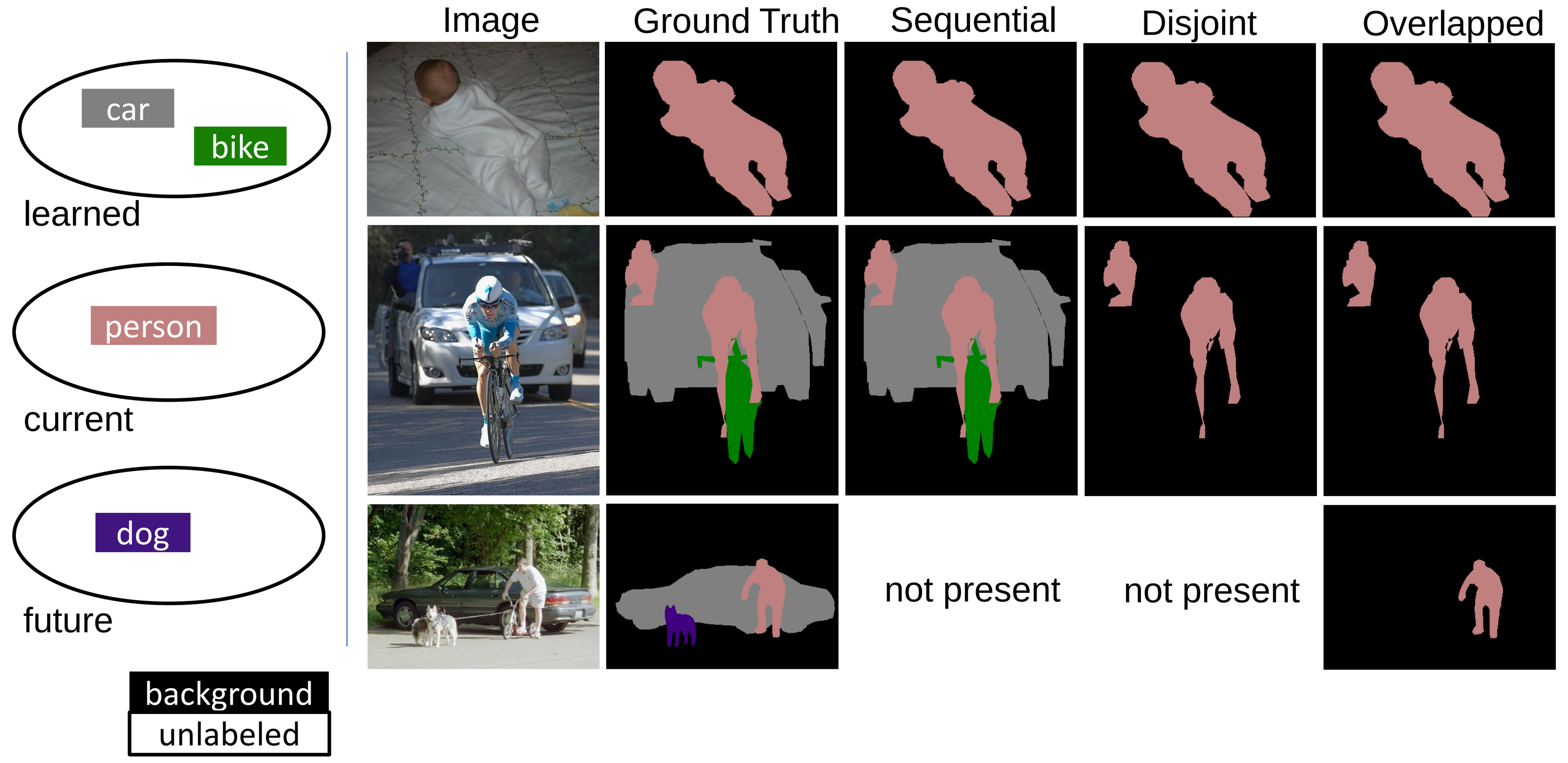
Continual Segmentation

Our Focus: Class-Incremental Continual Learning in Semantic Segmentation





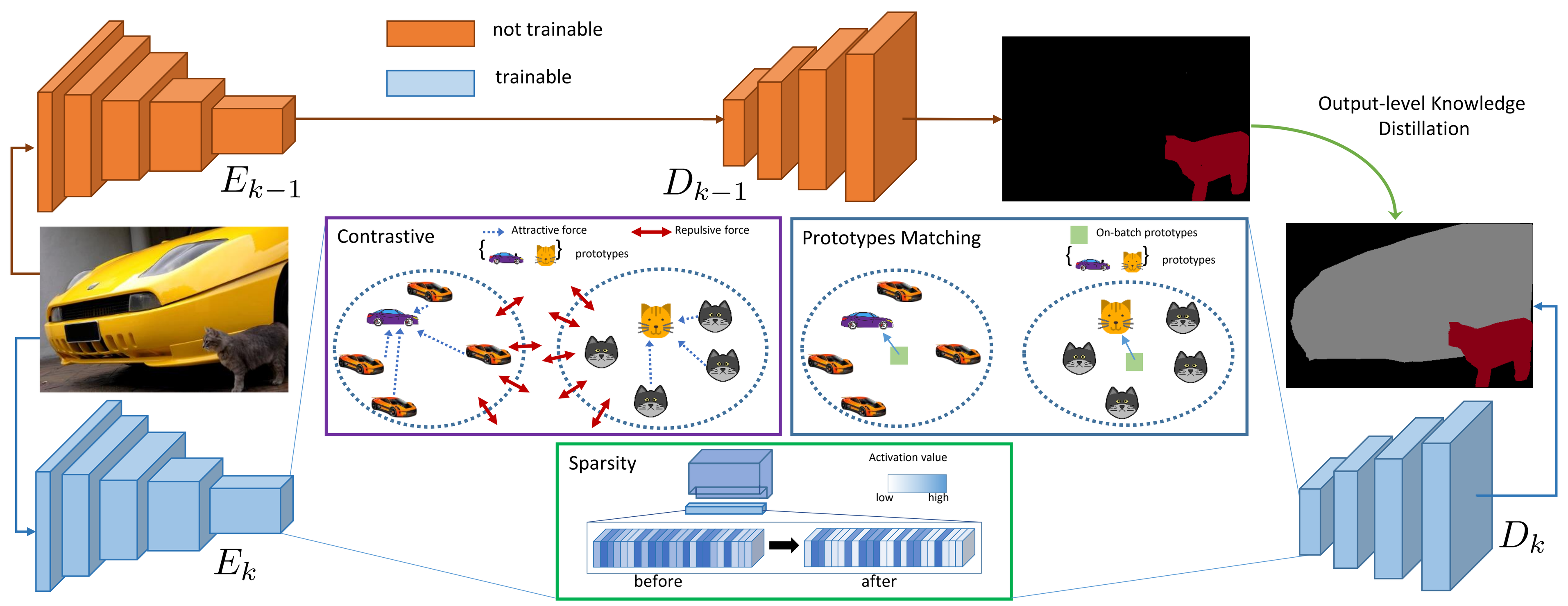
Many setups emerged to deal with the *background shift* and annotation of previous classes in future steps.



Our Approach (SDR)

SDR: **S**parse and **D**isentangled **R**epresentations

We combine task-related cross entropy loss with **4 constraints**:



Prototype matching:

$$\mathcal{L}_{pm} = \frac{1}{|\mathcal{C}_{k-1}|} ||\hat{\boldsymbol{p}}_c - \boldsymbol{p}_c||_F \qquad c \in \mathcal{C}_{k-1}$$

Set of previous classes

→ On-batch prototypes constrained to be close to representations learned from previous steps

Contrastive Learning:

- Attractive: $\mathcal{L}^a_{cl} = \frac{1}{|c_j \in \boldsymbol{y}^*_n|} \sum_{c_j \in \boldsymbol{y}^*_n} \sum_{\boldsymbol{f}_i \in \boldsymbol{F}_n} \left|\left| \left(\boldsymbol{f}_i - \boldsymbol{p}_{c_j}\right) \mathbb{I}[y^*_i = c_j] \right|\right|_F$

Features of the same class tightly clustered around prototype

- Repulsive: $\mathcal{L}^r_{cl} = \frac{1}{|c_j \in \boldsymbol{y}^*_n|} \sum_{c_j \in \boldsymbol{y}^*_n} \sum_{c_k \in \boldsymbol{y}^*_n, c_k \neq c_j} \frac{1}{||\hat{\boldsymbol{p}}_{c_j} - \hat{\boldsymbol{p}}_{c_k}||_F}$

Features of different classes separated from each other

Features Sparsity:

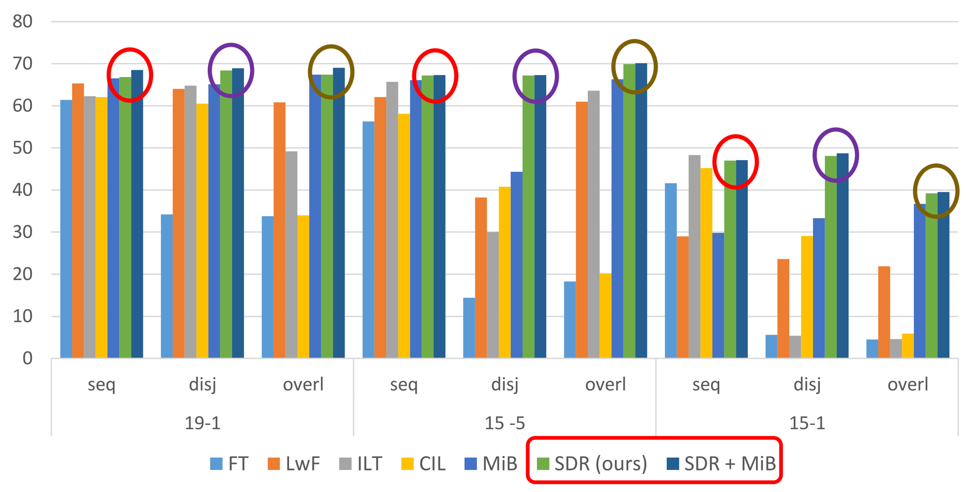
$$\mathcal{L}_{sp} = \frac{1}{|\boldsymbol{f}_i \in \boldsymbol{F}_n|} \sum_{\boldsymbol{f}_i \in \boldsymbol{F}_n} \frac{\sum_j \exp(\bar{f}_{i,j})}{\sum_j \bar{f}_{i,j}}$$

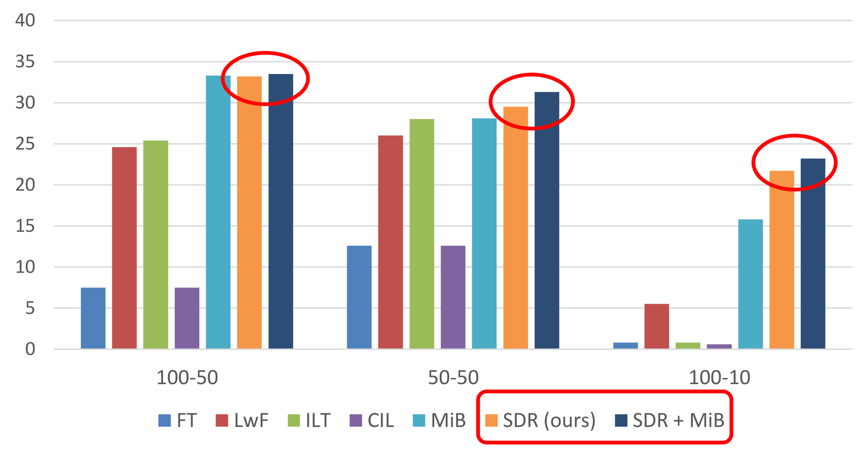
features normalized with respect to the class-conditional maximum value

Set of active channels is narrowed, letting room for the representation of upcoming classes

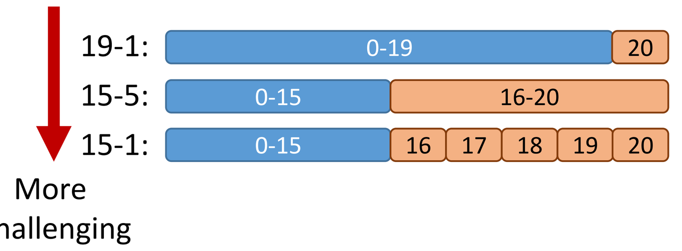
Results

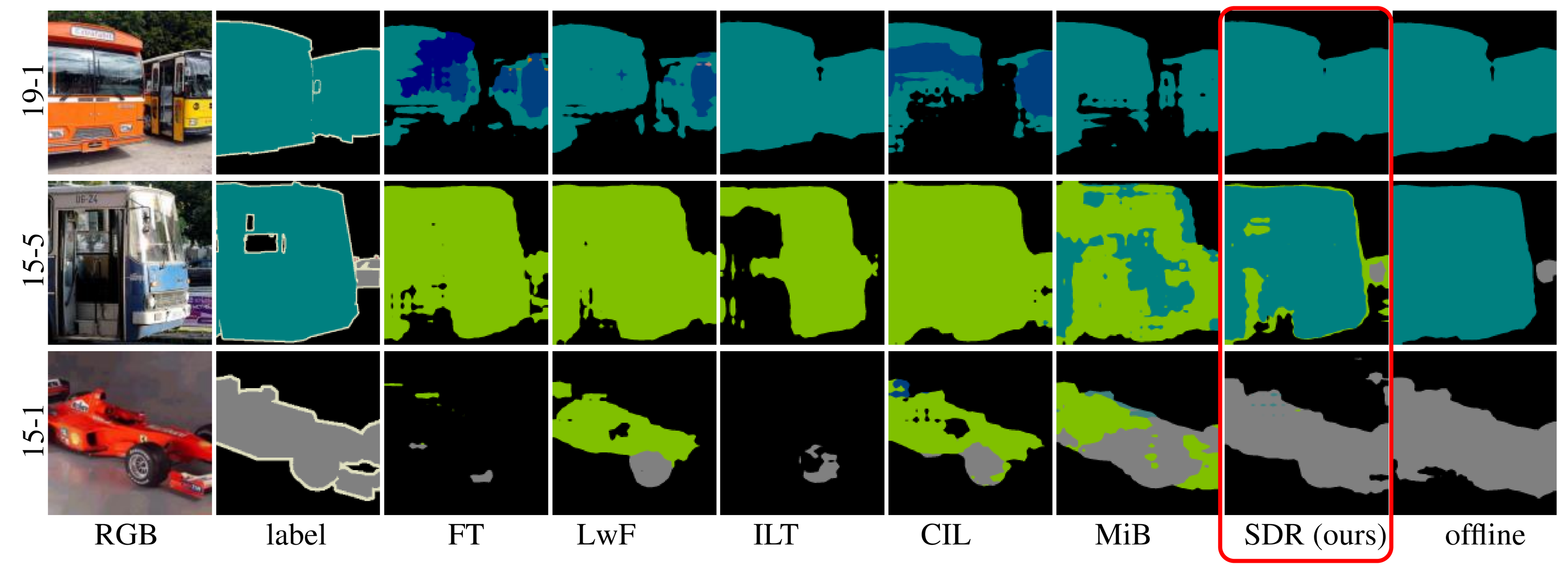
Pascal VOC 2012 **ADE20K**





Visual Results on Pascal VOC2012





Conclusion

- ✓ We propose 3 novel latent space shaping techniques to avoid forgetting and promote learning of new concepts: **prototype matching**, **contrastive learning** and **sparsity**
- ✓ We jointly tackle sequential, disjoint and overlapped scenarios
- ✓ We achieve state-of-the-art results on a variety of tasks and datasets

Paper website: https://lttm.dei.unipd.it/paper_data/SDR/

Code available: https://github.com/LTTM/SDR