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

Umberto Michieli and Pietro Zanuttigh
umberto.michieli@dei.unipd.it

University of Padova
Our Focus: Class-Incremental Continual Learning in Semantic Segmentation
Continual Segmentation - Different Setups

- **Image**
- **Ground Truth**
- **Sequential**
- **Disjoint**
- **Overlapped**

**Learned**
- car
- bike

**Current**
- person

**Future**
- dog

**Background**
- unlabeled

- **not present**
- **not present**
SDR: Sparse and Disentangled Representations

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

- Contrastive
  - Attractive force
  - Repulsive force

- Prototypes Matching
  - On-batch prototypes

- Sparsity
  - Activation value
    - Low
    - High

Before
After

Output-level Knowledge Distillation [1,2]

SDR Architecture

\[ \mathcal{L}_{pm} = \frac{1}{|\mathcal{C}_{k-1}|} \left\| \hat{p}_c - p_c \right\|_F \quad c \in \{\mathcal{C}_{k-1}\} \]

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

- Attractive: \[ \mathcal{L}_{cl}^a = \frac{1}{|c_j \in \mathcal{Y}_n^*|} \sum_{c_j \in \mathcal{Y}_n^*} \sum_{f_i \in \mathcal{F}_n} \left\| \left( f_i - p_{c_j} \right) \mathbb{I}[y_i^* = c_j] \right\|_F \]

Features of the same class tightly clustered around prototype

- Repulsive: \[ \mathcal{L}_{cl}^r = \frac{1}{|c_j \in \mathcal{Y}_n^*|} \sum_{c_j \in \mathcal{Y}_n^*} \sum_{c_k \in \mathcal{Y}_n^*} \left\| \hat{p}_{c_j} - \hat{p}_{c_k} \right\|_F \]

Features of different classes separated from each other

\[ \mathcal{L}_{sp} = \frac{1}{|f_i \in \mathcal{F}_n|} \sum_{f_i \in \mathcal{F}_n} \frac{\sum_j \exp(\hat{f}_{i,j})}{\sum_j \hat{f}_{i,j}} \]

Set of active channels is narrowed, letting room for the representation of upcoming classes
Results – Pascal VOC2012

More challenging
→ SDR outperforms competitors, especially when multiple steps are involved.
We propose 3 novel latent space shaping techniques to avoid forgetting and promote learning of new concepts:
- prototype matching
- contrastive learning
- 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