Self-paced Contrastive Learning with Hybrid Memory for Domain Adaptive Object Re-ID

香 港 中 文 大 學
The Chinese University of Hong Kong

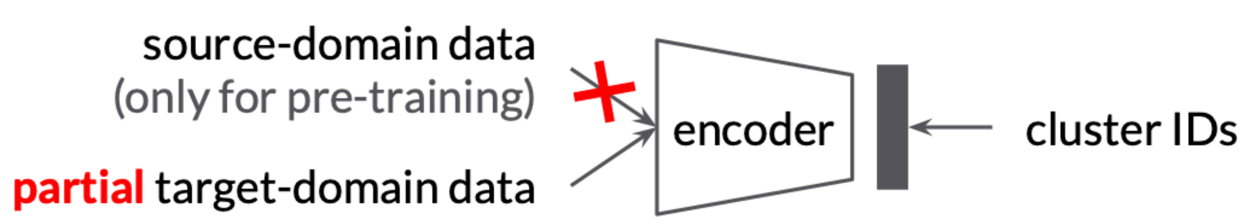
Yixiao Ge, Feng Zhu, Dapeng Chen, Rui Zhao, Hongsheng Li
Multimedia Laboratory, The Chinese University of Hong Kong

NEURAL INFORMATION PROCESSING SYSTEMS

Existing Domain Adaptive Methods on Object Re-ID

Two-stage training scheme:

1. Supervised pre-training on the source domain with ground-truth labels;
2. Unsupervised fine-tuning on the target domain with pseudo labels, which can be generated by clustering instance features.



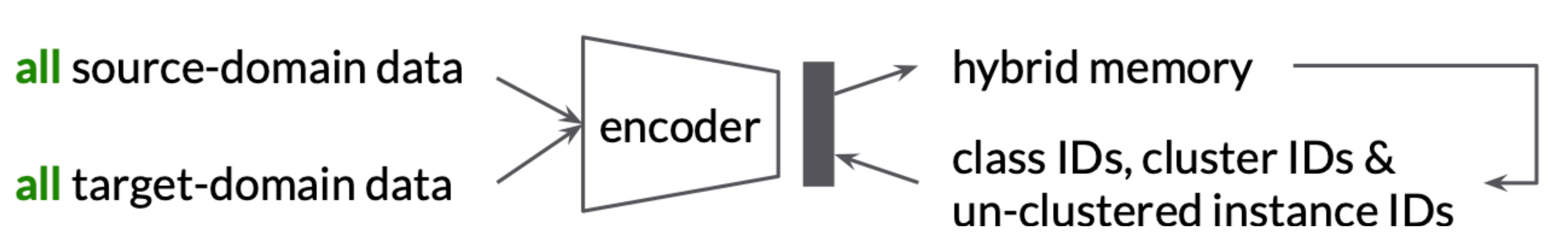
Limitations:

- The <u>accurate</u> **source-domain ground-truth labels** are <u>valuable</u> but were ignored during target-domain training.
- Discard <u>difficult but valuable</u> **clustering outlier samples** from being used for training. Note that there are generally many outliers especially in early epochs.

Motivations & Contributions

Motivations:

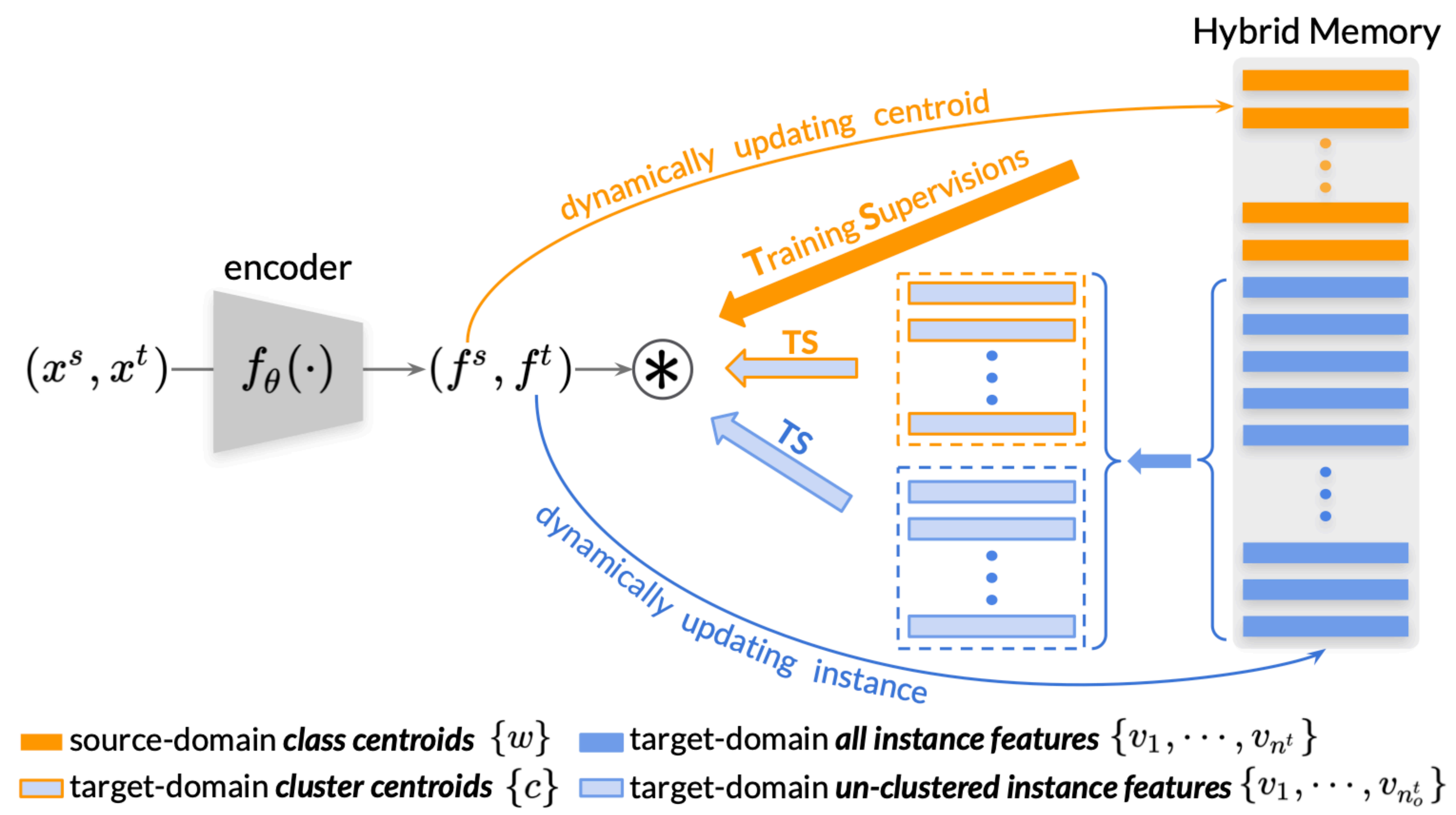
- Encode **all available information** from both source and target domains;
- Treat all source-domain classes, target-domain clusters and un-clustered outlier instances as **equal classes** for training.



Contributions:

- Propose **a unified contrastive learning framework with hybrid memory** for joint feature learning with class-level, cluster-level and un-clustered instance-level supervisions;
- Design **a self-paced learning strategy** with a clustering reliability criterion to gradually provide more confident learning targets for training;
- Significantly outperform state-of-the-arts with up to **5.0%** mAP gains on <u>domain adaptive object re-ID</u> tasks, and up to **16.7%** mAP gains on <u>unsupervised object re-ID</u> tasks. Our method can even boost the <u>source-domain performance</u> with up to **6.6%** mAP gains by incorporating unlabeled target-domain data for joint training.

Self-paced Contrastive Learning (*SpCL*) Framework



Unified contrastive Loss:

$$\mathcal{L}_{\boldsymbol{f}} = -\log \frac{\exp\left(\langle \boldsymbol{f}, \overbrace{\boldsymbol{z}^+}^{\text{positive prototype}} \rangle / \tau\right)}{\sum_{k=1}^{n^s} \exp\left(\langle \boldsymbol{f}, \underbrace{\boldsymbol{w}_k}_{\text{class centroids}} \rangle / \tau\right) + \sum_{k=1}^{n_c^t} \exp\left(\langle \boldsymbol{f}, \underbrace{\boldsymbol{c}_k}_{\text{cluster centroids}} \rangle / \tau\right) + \sum_{k=1}^{n_o^t} \exp\left(\langle \boldsymbol{f}, \underbrace{\boldsymbol{v}_k}_{\text{instance features}} \rangle / \tau\right)}$$

Positive prototypes:

- *For source-domain images:* class centroids
- *For target-domain clustered images:* cluster centroids
- *For target-domain un-clustered images:* instance features

Hybrid memory (momentum update):

- All the <u>source-domain</u> images are cached in the form of ground-truth **classes**:

$$\boldsymbol{w}_k \leftarrow m^s \boldsymbol{w}_k + (1 - m^s) \cdot \frac{1}{|\mathcal{B}_k|} \sum_{\boldsymbol{f}_i^s \in \mathcal{B}_k} \boldsymbol{f}_i^s$$

- All the <u>target-domain</u> images are cached in the form of **instances**:

$$\boldsymbol{v}_i \leftarrow m^t \boldsymbol{v}_i + (1 - m^t) \boldsymbol{f}_i^t$$

Then the cluster centroids can be calculated on-the-fly:

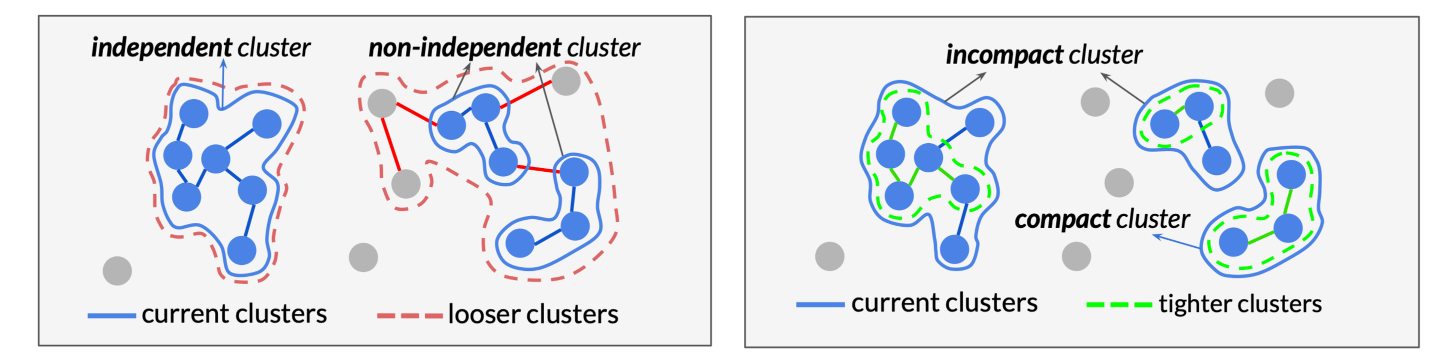
$$\boldsymbol{c}_k = \frac{1}{|\mathcal{I}_k|} \sum_{\boldsymbol{v}_i \in \mathcal{I}_k} \boldsymbol{v}_i$$

And the un-clustered instance features are directly loaded from the memory.

Self-paced Learning Strategy

Initialize the training process with the most reliable clusters and gradually incorporate more un-clustered instances to form new reliable clusters.

Cluster reliability criterion:

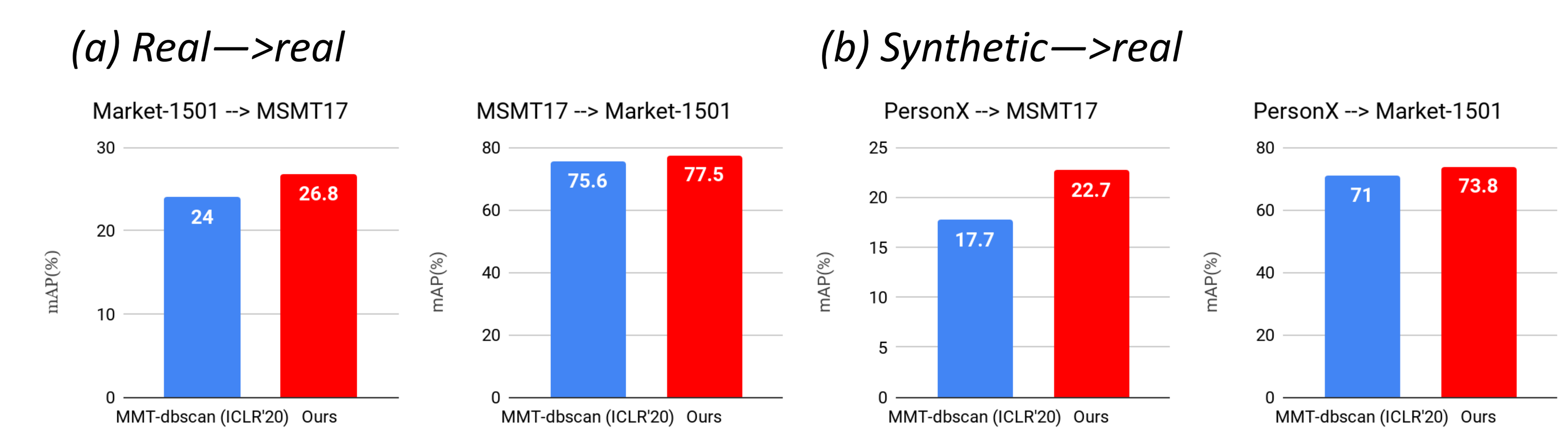


$$\mathcal{R}_{\text{indep}}(\boldsymbol{f}_i^t) = \frac{|\mathcal{I}(\boldsymbol{f}_i^t) \cap \mathcal{I}_{\text{loose}}(\boldsymbol{f}_i^t)|}{|\mathcal{I}(\boldsymbol{f}_i^t) \cup \mathcal{I}_{\text{loose}}(\boldsymbol{f}_i^t)|} \in [0, 1] \qquad \mathcal{R}_{\text{comp}}(\boldsymbol{f}_i^t) = \frac{|\mathcal{I}(\boldsymbol{f}_i^t) \cap \mathcal{I}_{\text{tight}}(\boldsymbol{f}_i^t)|}{|\mathcal{I}(\boldsymbol{f}_i^t) \cup \mathcal{I}_{\text{tight}}(\boldsymbol{f}_i^t)|} \in [0, 1]$$

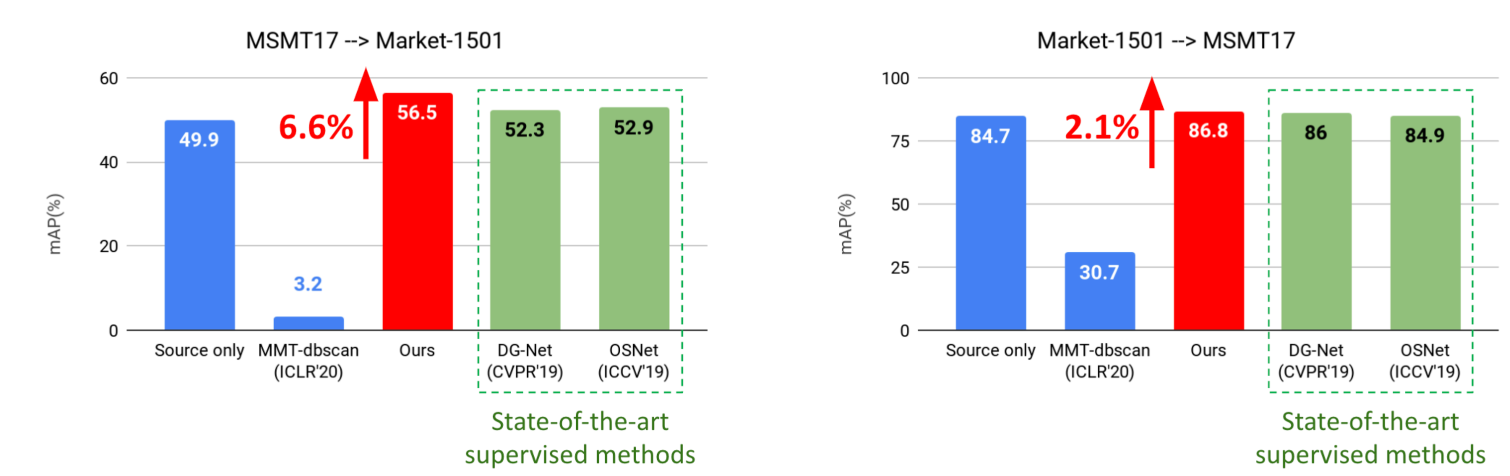
Only reliable clusters are preserved and other confusing clusters are disassembled back to un-clustered instances.

Experimental Results

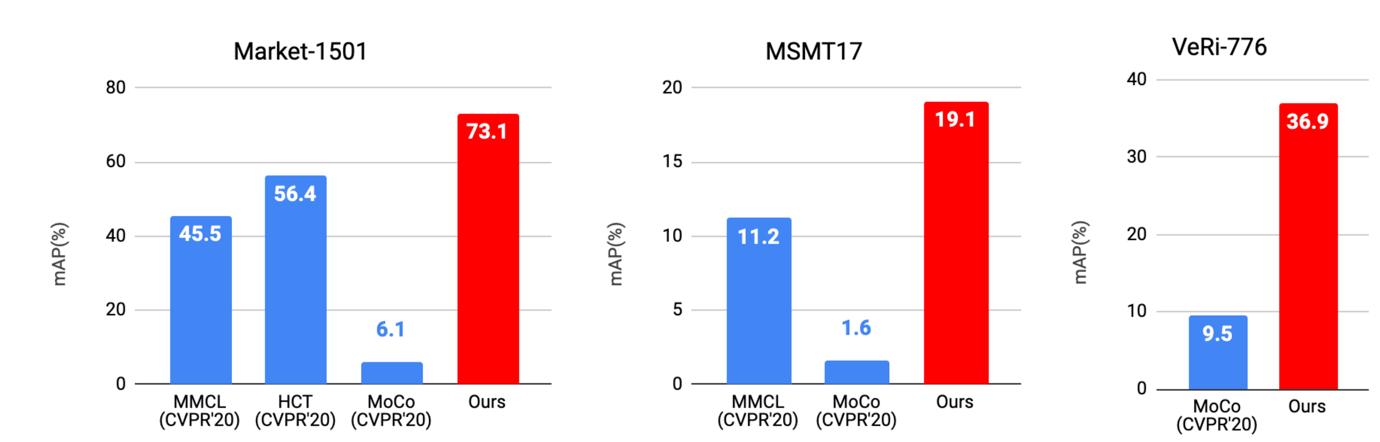
Domain adaptive object re-ID benchmarks, e.g.



Performance of domain adaptive models on the source domain



Unsupervised object re-ID benchmarks, e.g.



Project Page: https://geyixiao.com/projects/spcl.html