NEURAL INFORMATION PROCESSING SYSTEMS

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

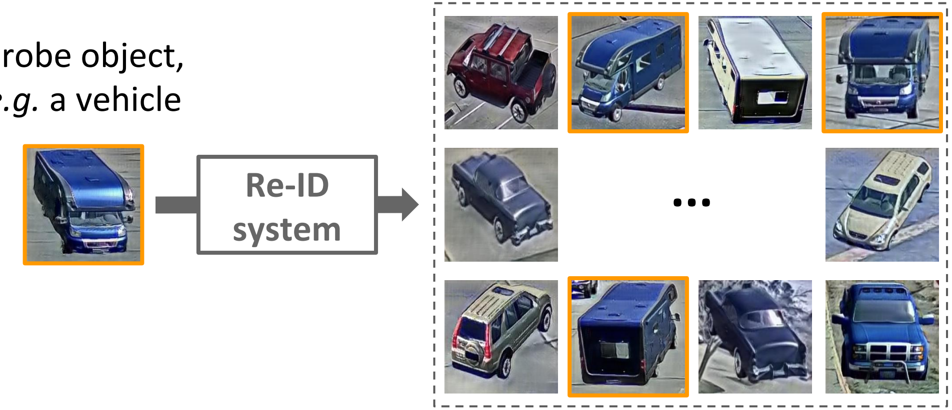
Yixiao Ge, Feng Zhu, Dapeng Chen, Rui Zhao, Hongsheng Li

Multimedia Laboratory
The Chinese University of Hong Kong

Object Re-identification (Re-ID)

Identify the probe object from multiple cameras

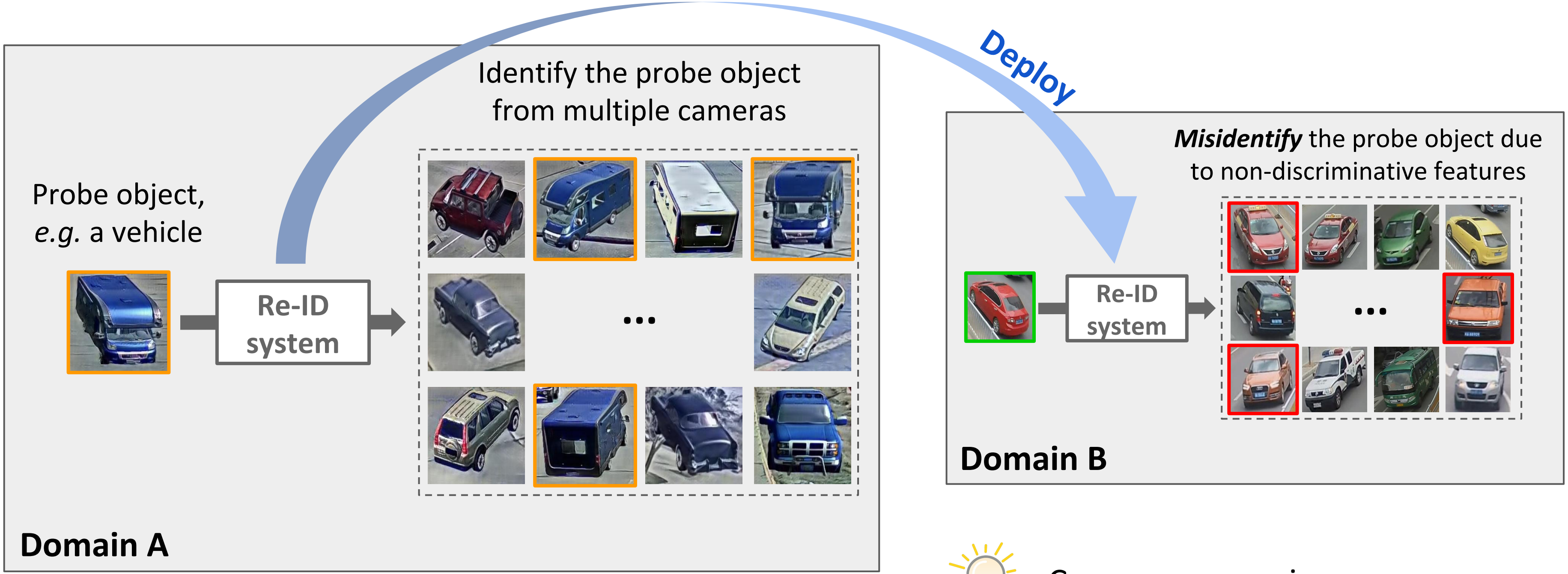
Probe object, *e.g.* a vehicle



Learn discriminative features in varying conditions.

Example images from *VehicleX* synthetic dataset, introduced in the 4th AI City Challenge.

Object Re-identification (Re-ID) -- Domain Gaps

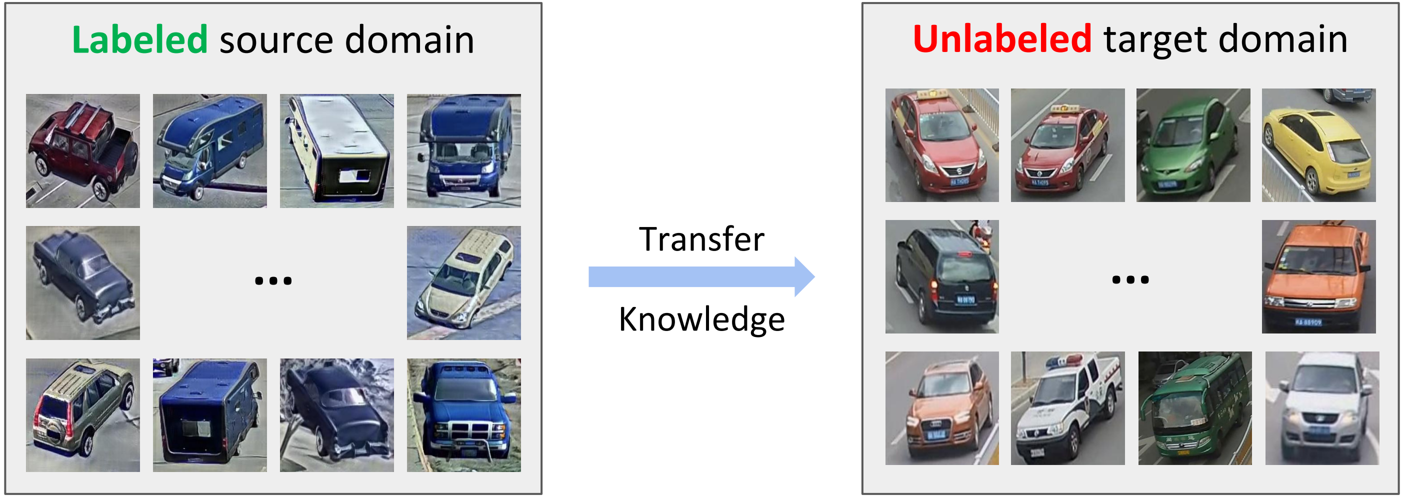


Common scenarios:

- City A ⟶ City B
- Synthetic ⟶ Real-world

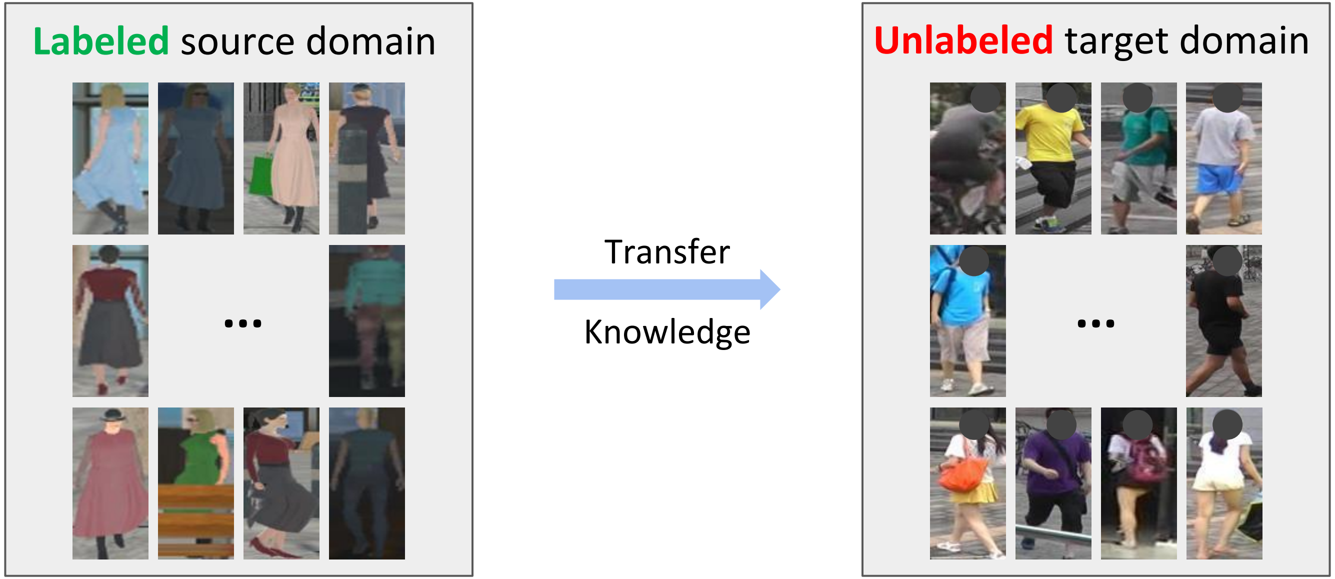
Example images from *VehicleX* synthetic dataset for domain A and VeRi-776 real-world dataset for domain B.

Open-class Domain Adaptive Vehicle Re-ID



Example images from *VehicleX* synthetic dataset for the source domain and VeRi-776 real-world dataset for the target domain.

Open-class Domain Adaptive Person Re-ID

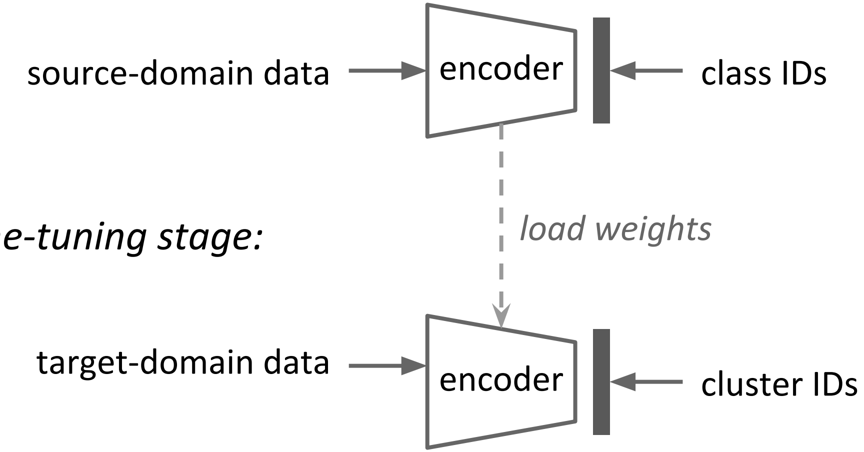


Example images from *PersonX* synthetic dataset for the source domain and Market-1501 real-world dataset for the target domain.

Previous UDA Methods on Object Re-ID

(1) Pre-training stage:

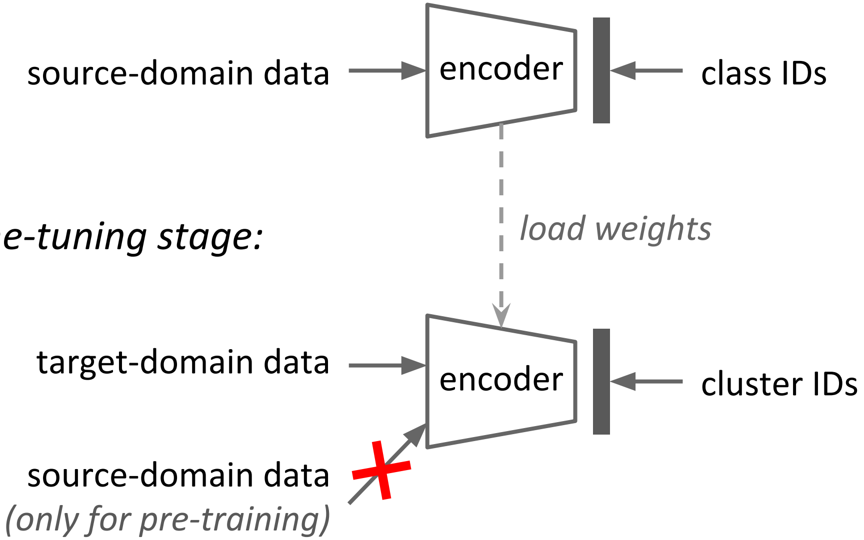
(2) Fine-tuning stage:



Previous UDA Methods on Object Re-ID

(1) Pre-training stage:

(2) Fine-tuning stage:



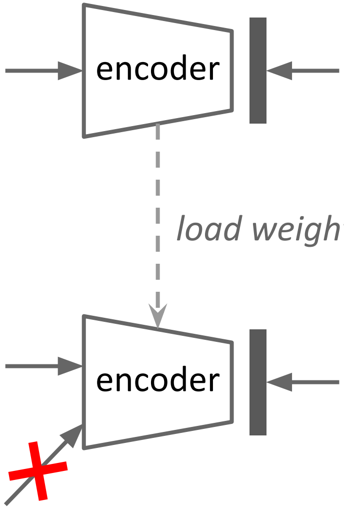
Limitation #1:

*The <u>accurate</u> **source-domain ground-truth labels** are <u>valuable</u> but were ignored during target-domain training.*

Previous UDA Methods on Object Re-ID

(1) Pre-training stage:

source-domain data → encoder ← class IDs

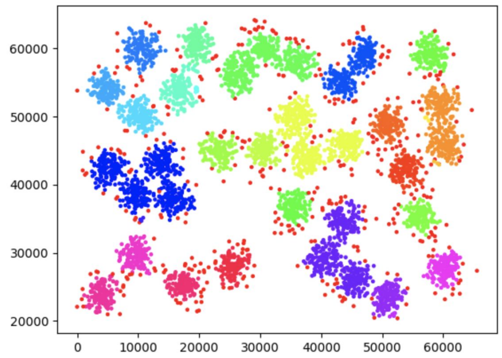


(2) Fine-tuning stage:

Partial target-domain data
(only clustered data are used) → encoder ← cluster IDs

source-domain data
(only for pre-training)

red points are outliers



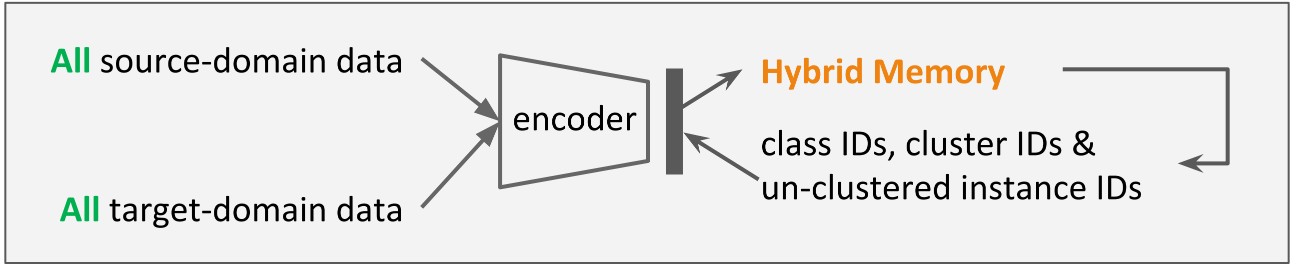
Limitation #2:

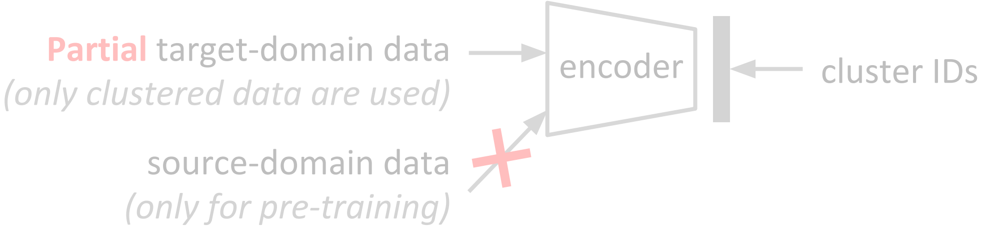
*Discard <u>difficult but valuable</u> **clustering outlier samples** from being used for training. Note that there are generally many outliers especially in early epochs.*

Solution

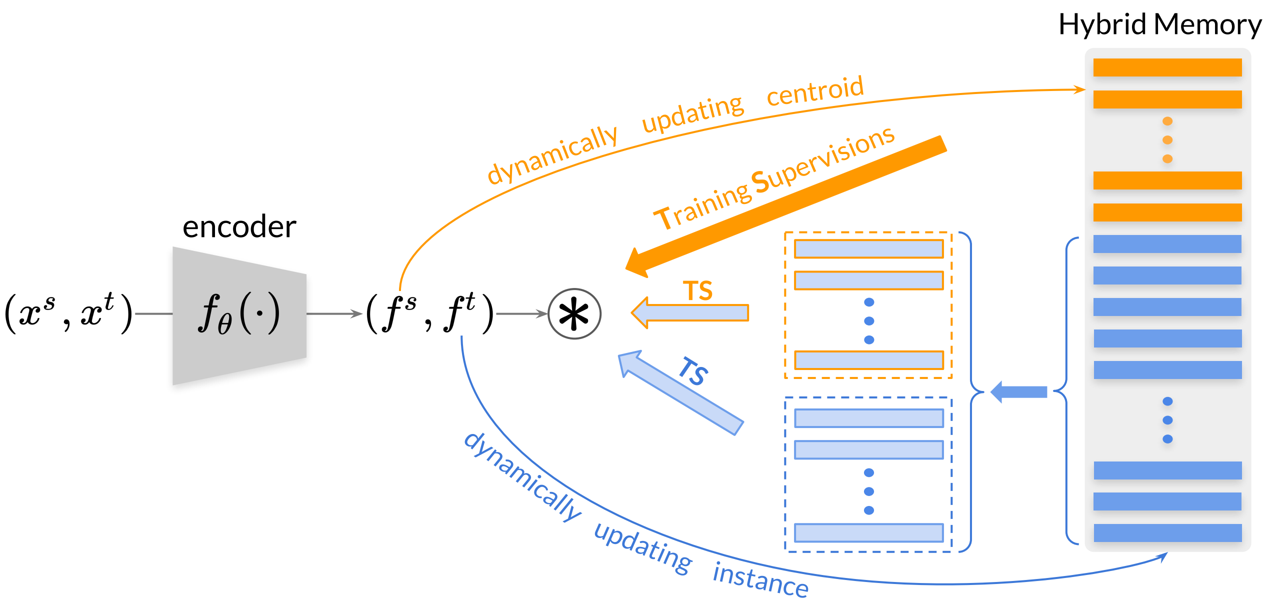
Encode all available information,

i.e. source data, clustered target data, un-clustered target data



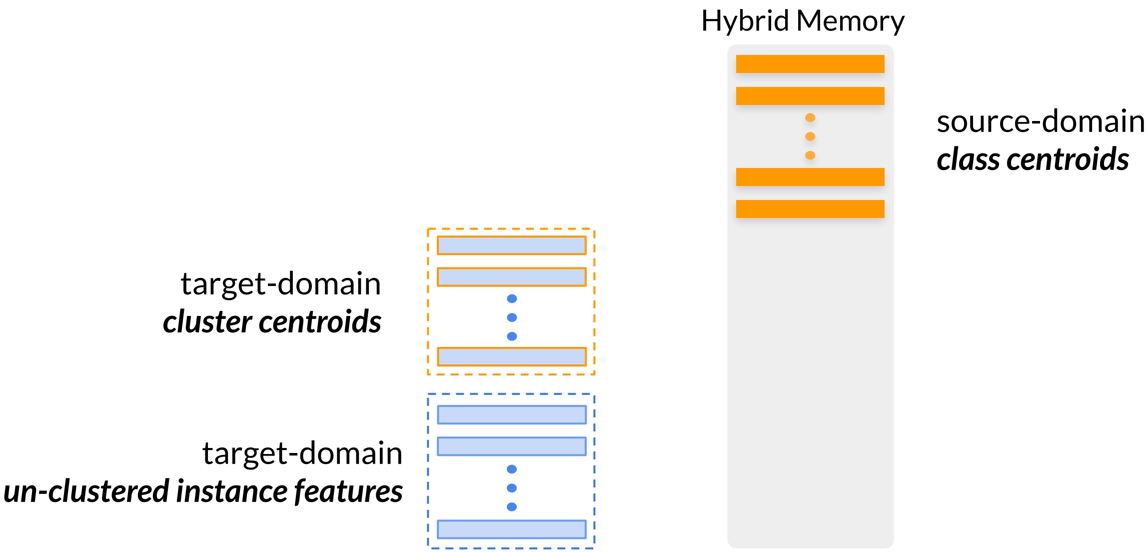


SpCL Framework

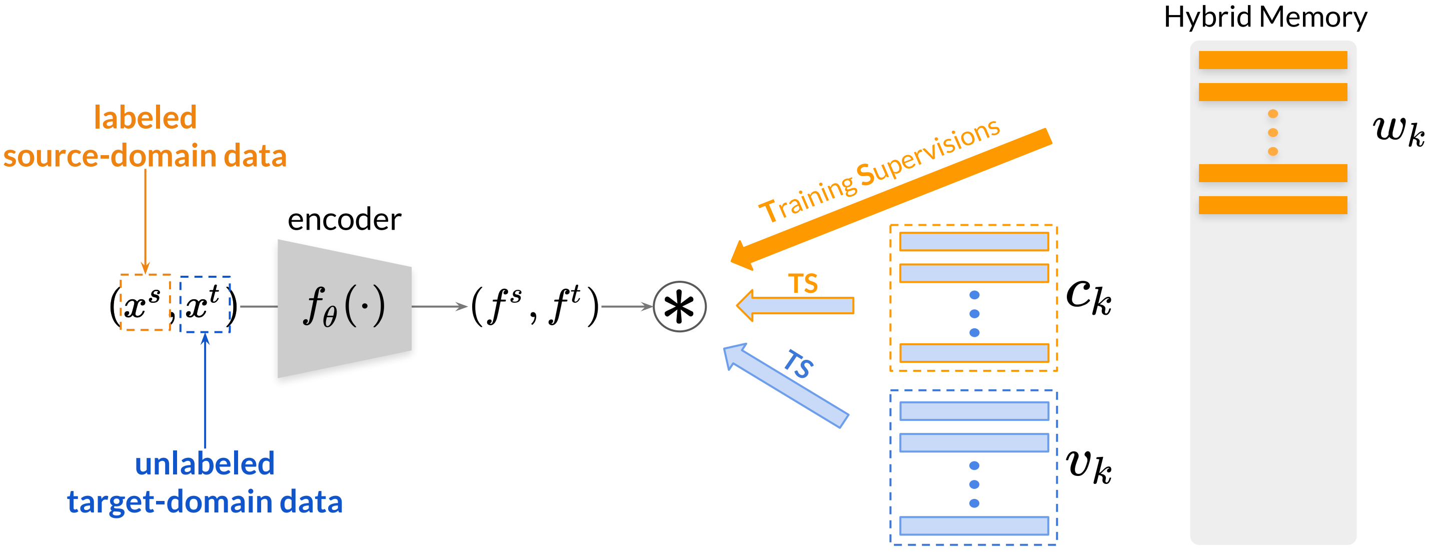


source-domain ***class centroids*** $\{w\}$
target-domain ***cluster centroids*** $\{c\}$
target-domain ***all instance features*** $\{v_1, \cdots, v_{n^t}\}$
target-domain ***un-clustered instance features*** $\{v_1, \cdots, v_{n^t_o}\}$

Prototypes

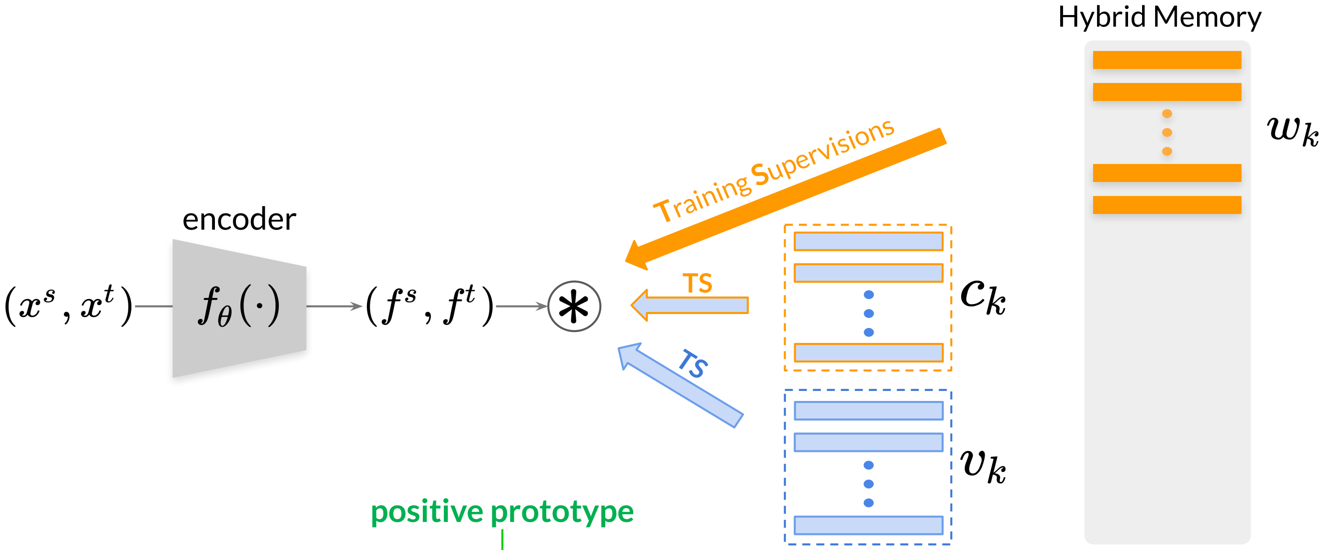

Hybrid Memory
source-domain
class centroids
target-domain
cluster centroids
target-domain
un-clustered instance features

Contrast



labeled
source-domain data
encoder
(x^s, x^t)
$f_\theta(\cdot)$
(f^s, f^t)
unlabeled
target-domain data
Training Supervisions
TS
TS
c_k
v_k
Hybrid Memory
w_k

Unified Contrastive Loss



positive prototype

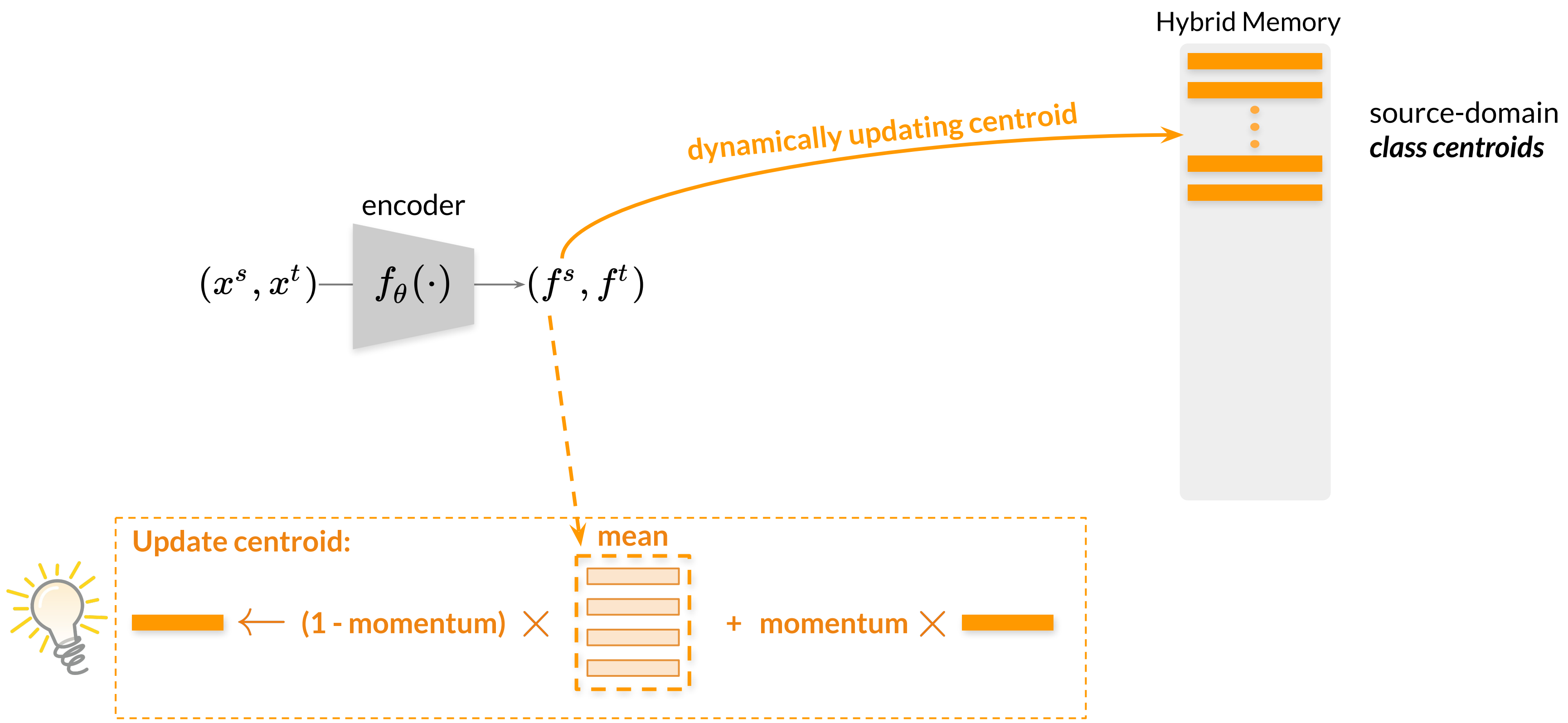
$$\mathcal{L}_{\boldsymbol{f}} = -\log \frac{\exp\left(\langle \boldsymbol{f}, \boxed{\boldsymbol{z}^+} \rangle / \tau\right)}{\sum_{k=1}^{n^s} \exp\left(\langle \boldsymbol{f}, \boxed{\boldsymbol{w}_k} \rangle / \tau\right) + \sum_{k=1}^{n_c^t} \exp\left(\langle \boldsymbol{f}, \boxed{\boldsymbol{c}_k} \rangle / \tau\right) + \sum_{k=1}^{n_o^t} \exp\left(\langle \boldsymbol{f}, \boxed{\boldsymbol{v}_k} \rangle / \tau\right)}$$

class centroids

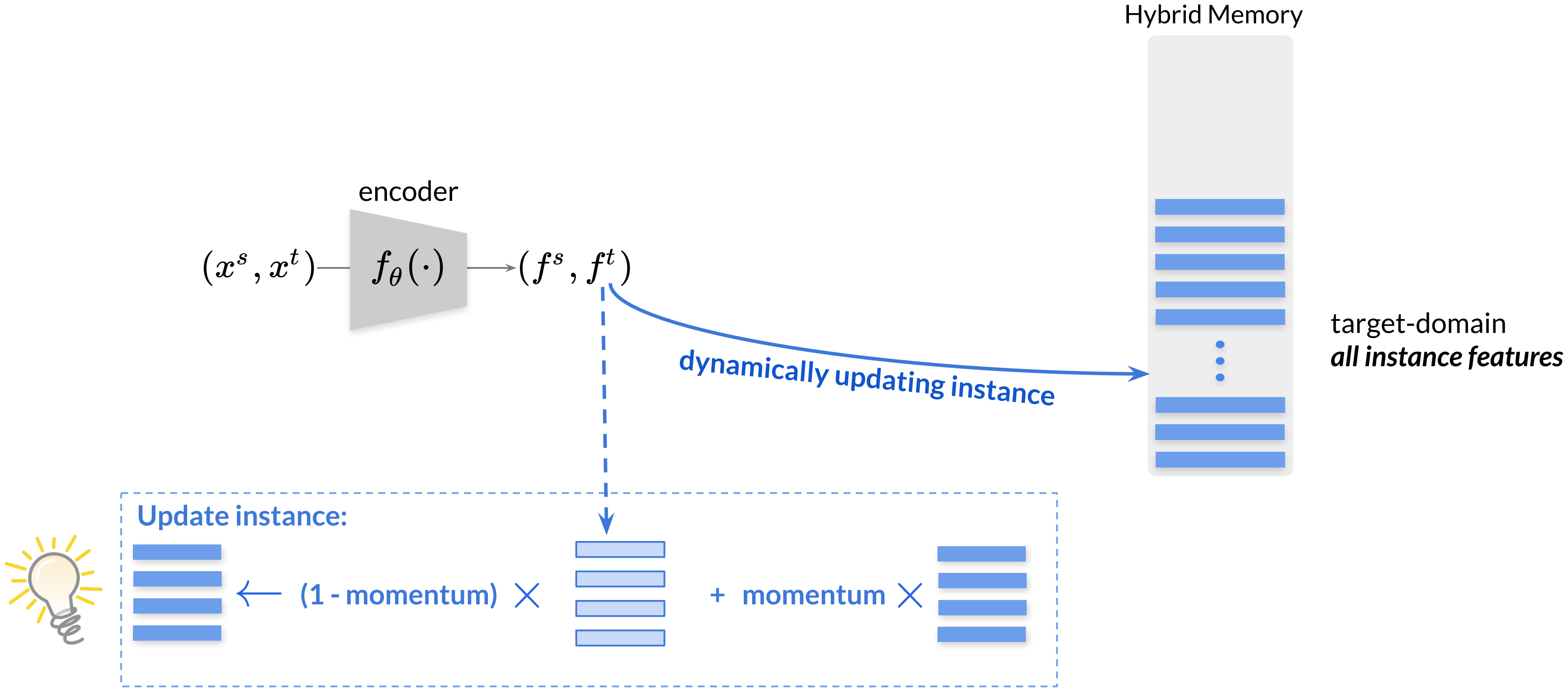
cluster centroids

instance features

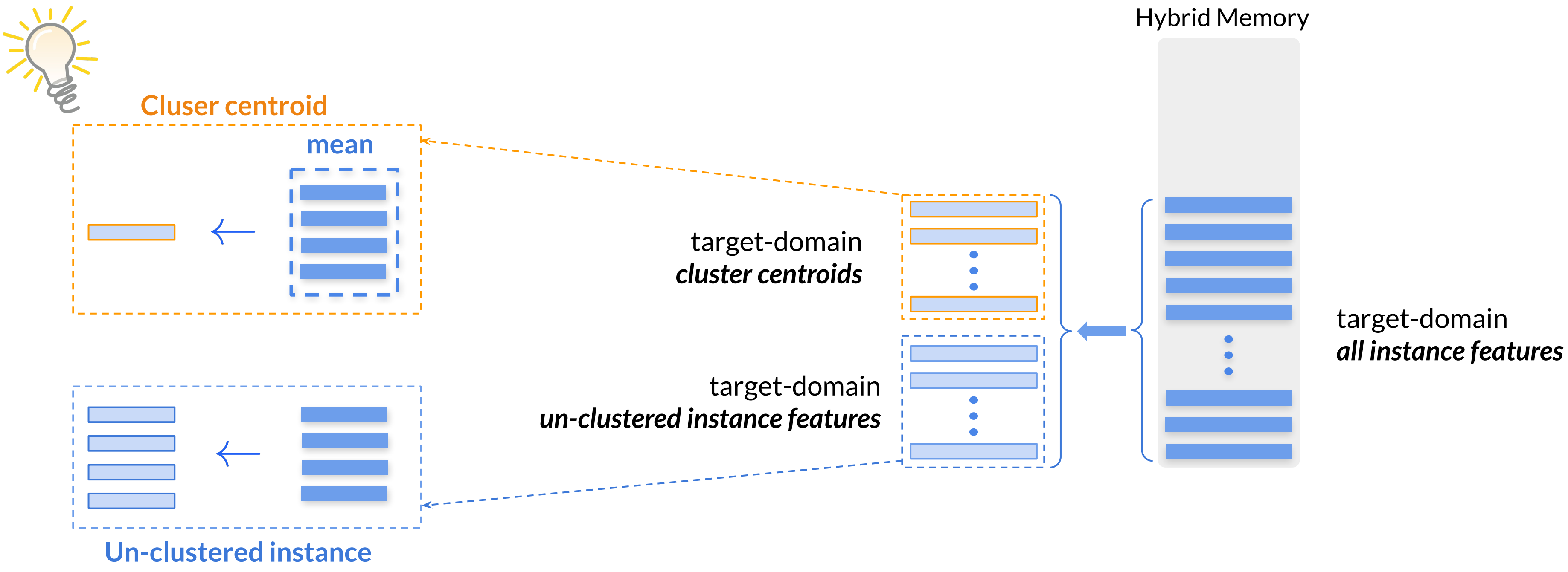
Update Memory -- Source-domain Class Centroids



Update Memory -- Target-domain Instance Features

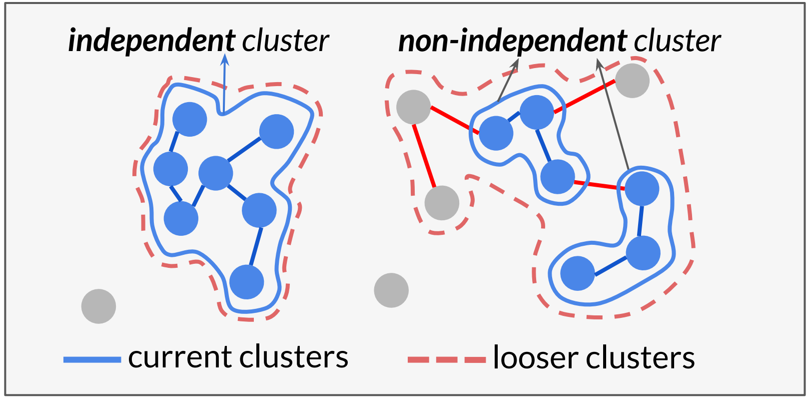


Target-domain Cluster Centroids & Un-clustered Instances



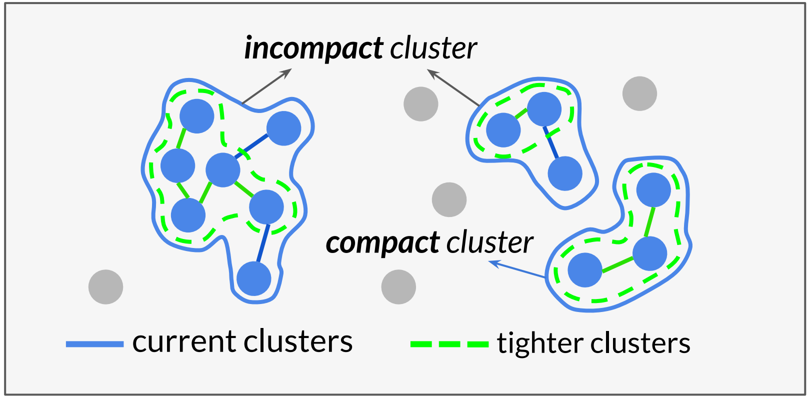
Cluster Reliability Criterion

Cluster independence*



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

Cluster campactness



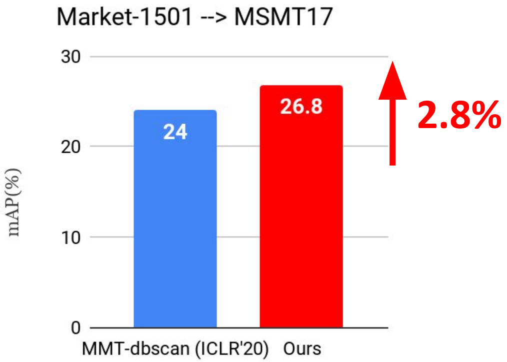
$$\mathcal{R}_{\text{comp}}(\boldsymbol{f}_i^t) = \frac{|\mathcal{I}(\boldsymbol{f}_i^t) \cap \mathcal{I}_{\text{tight}}(\boldsymbol{f}_i^t)|}{|\mathcal{I}(\boldsymbol{f}_i^t) \cup \mathcal{I}_{\text{tight}}(\boldsymbol{f}_i^t)|} \in [0,1]$$

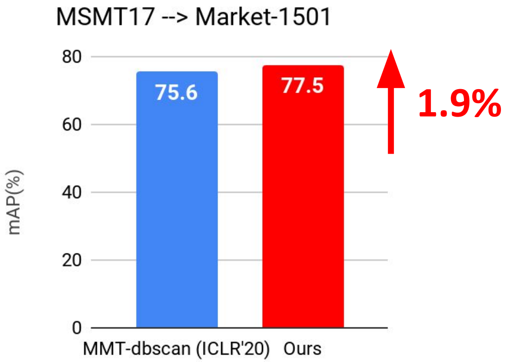
We preserve independent clusters with compact data points whose $\mathcal{R}_{\text{indep}} > \alpha$ and $\mathcal{R}_{\text{comp}} > \beta$, while the remaining data are treated as un-clustered outlier instances.

* *"Independence" is used in its idiomatic sense rather than the statistical sense.*

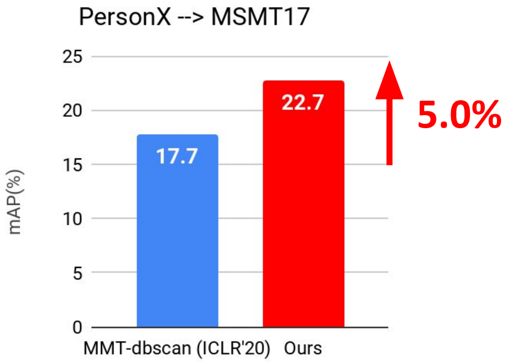
Domain Adaptive Object Re-ID Performance

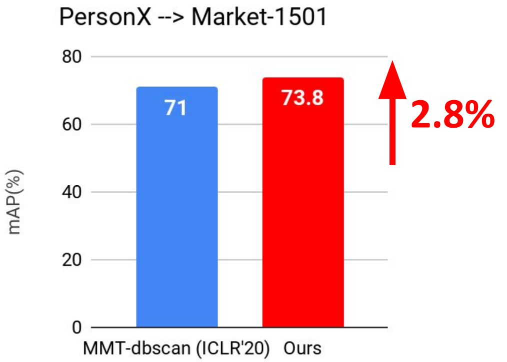
(a) Real → real adaptation on person re-ID tasks





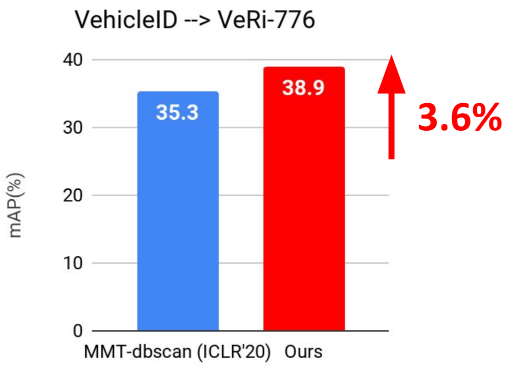
(b) Synthetic → real adaptation on person re-ID tasks



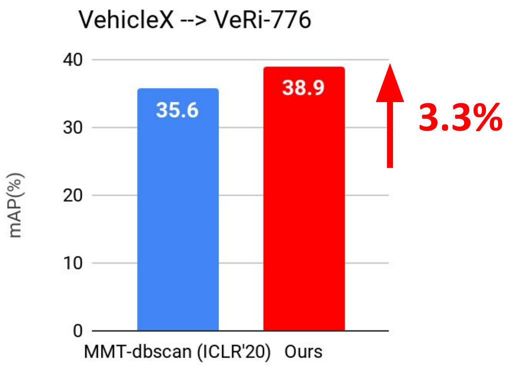


Domain Adaptive Object Re-ID Performance

(c) Real → real adaptation on vehicle re-ID tasks

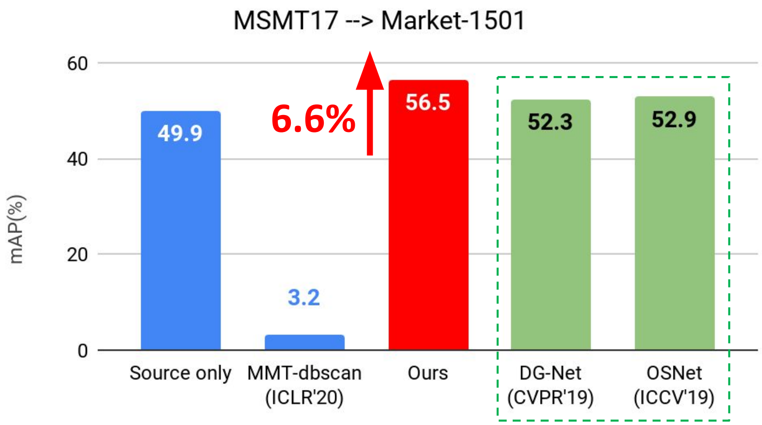


(d) Synthetic → real adaptation on vehicle re-ID tasks

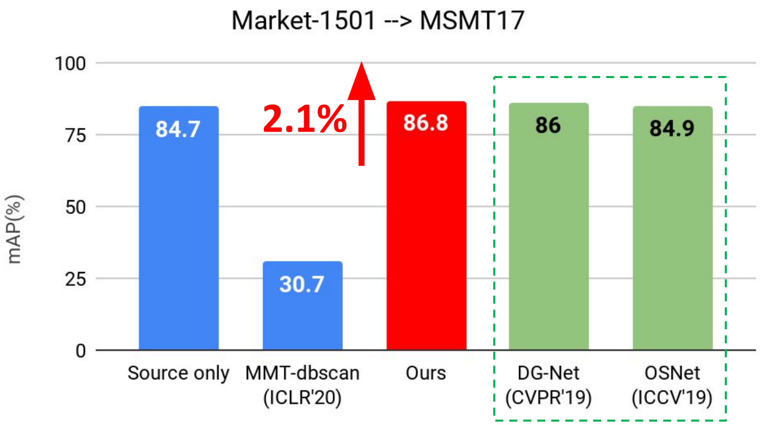


An inspiring discovery: synthetic → real task could achieve competitive performance (38.9%) as the real → real task with the same target-domain dataset (VeRi-776), which indicates that we are one more step closer towards **no longer needing any manually annotated real-world images** in the future.

Performance on the Source Domain



State-of-the-art supervised methods



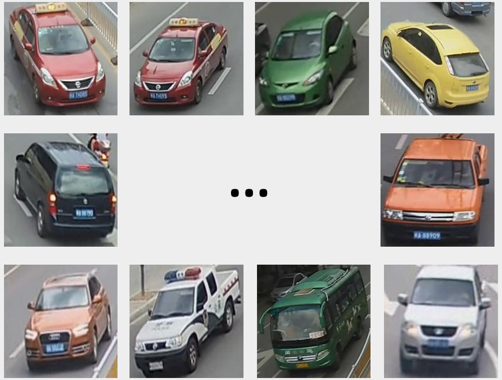
State-of-the-art supervised methods

Our method could even **boost the source-domain performance**, while previous UDA methods (*e.g.* MMT) inevitably forget the source-domain knowledge. Our method also outperforms state-of-the-art supervised re-ID methods (*e.g.* DG-Net, OSNet), indicates that our method could be applied to **improve the supervised training by incorporating unlabeled data** without extra human labor.

Unsupervised Vehicle Re-ID

NO source domain

Unlabeled target domain

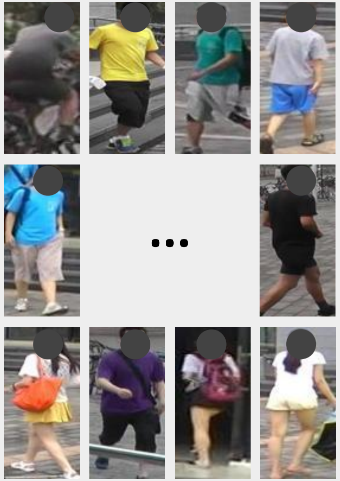


Example images from VeRi-776 dataset for the target domain.

Unsupervised Person Re-ID

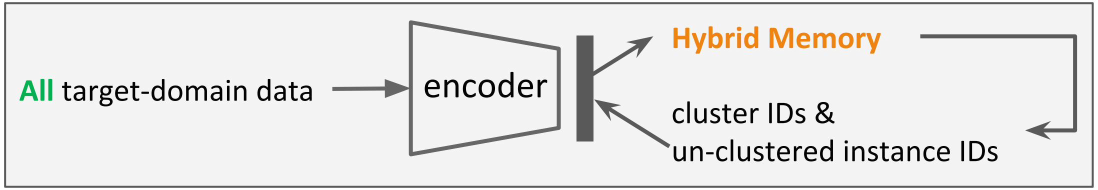
NO source domain

Unlabeled target domain



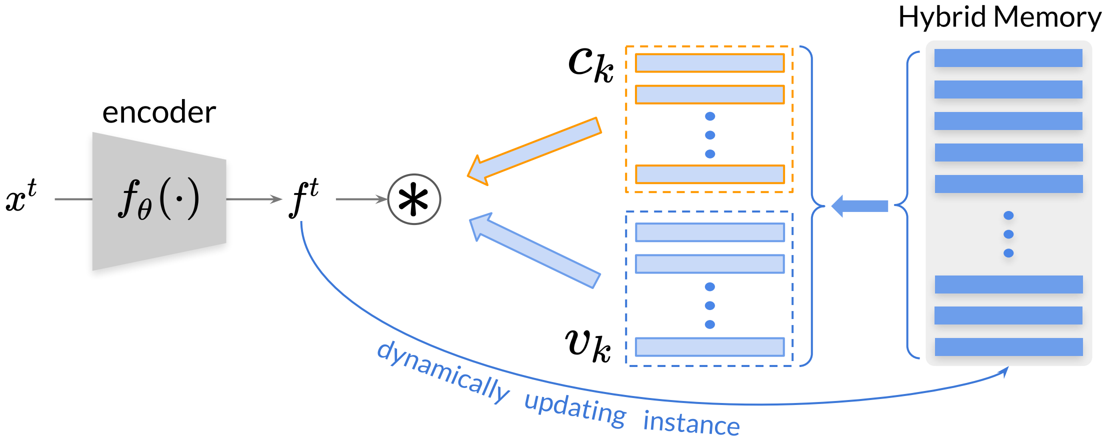
Example images from Market-1501 dataset for the target domain.

Generalized Version of *SpCL* for Unsupervised Object Re-ID

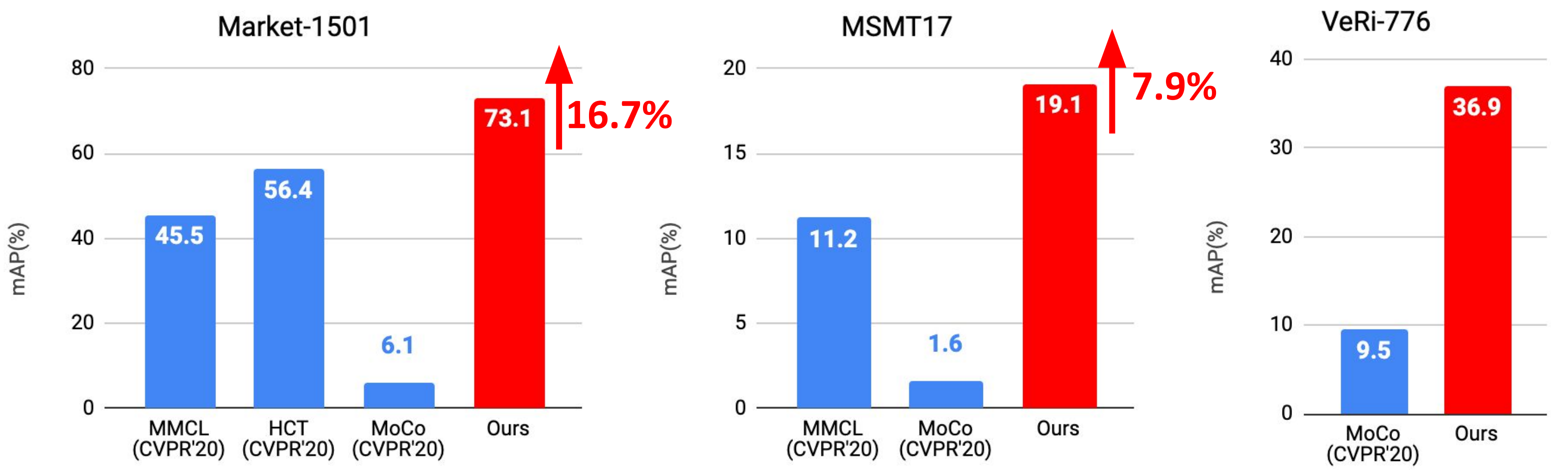


- target-domain ***all instance features*** $\{v_1, \cdots, v_{n^t}\}$
- target-domain ***un-clustered instance features*** $\{v_1, \cdots, v_{n_o^t}\}$
- target-domain ***cluster centroids*** $\{c\}$

$$\mathcal{L}_f = -\log \frac{\exp(\langle f, z^+ \rangle/\tau)}{\sum_{k=1}^{n_c^t} \exp(\langle f, c_k \rangle/\tau) + \sum_{k=1}^{n_o^t} \exp(\langle f, v_k \rangle/\tau)}$$



Unsupervised Object Re-ID Performance



MoCo is inapplicable on unsupervised re-ID tasks, because it treats each instance as a single class, while the core of re-ID tasks is to encode and model intra-/inter-class variations.

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

Yixiao Ge, Feng Zhu, Dapeng Chen, Rui Zhao, Hongsheng Li

Multimedia Laboratory, The Chinese University of Hong Kong

Code available at

https://github.com/yxgeee/SpCL