南京航空航天大學
Nanjing University of Aeronautics and Astronautics

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

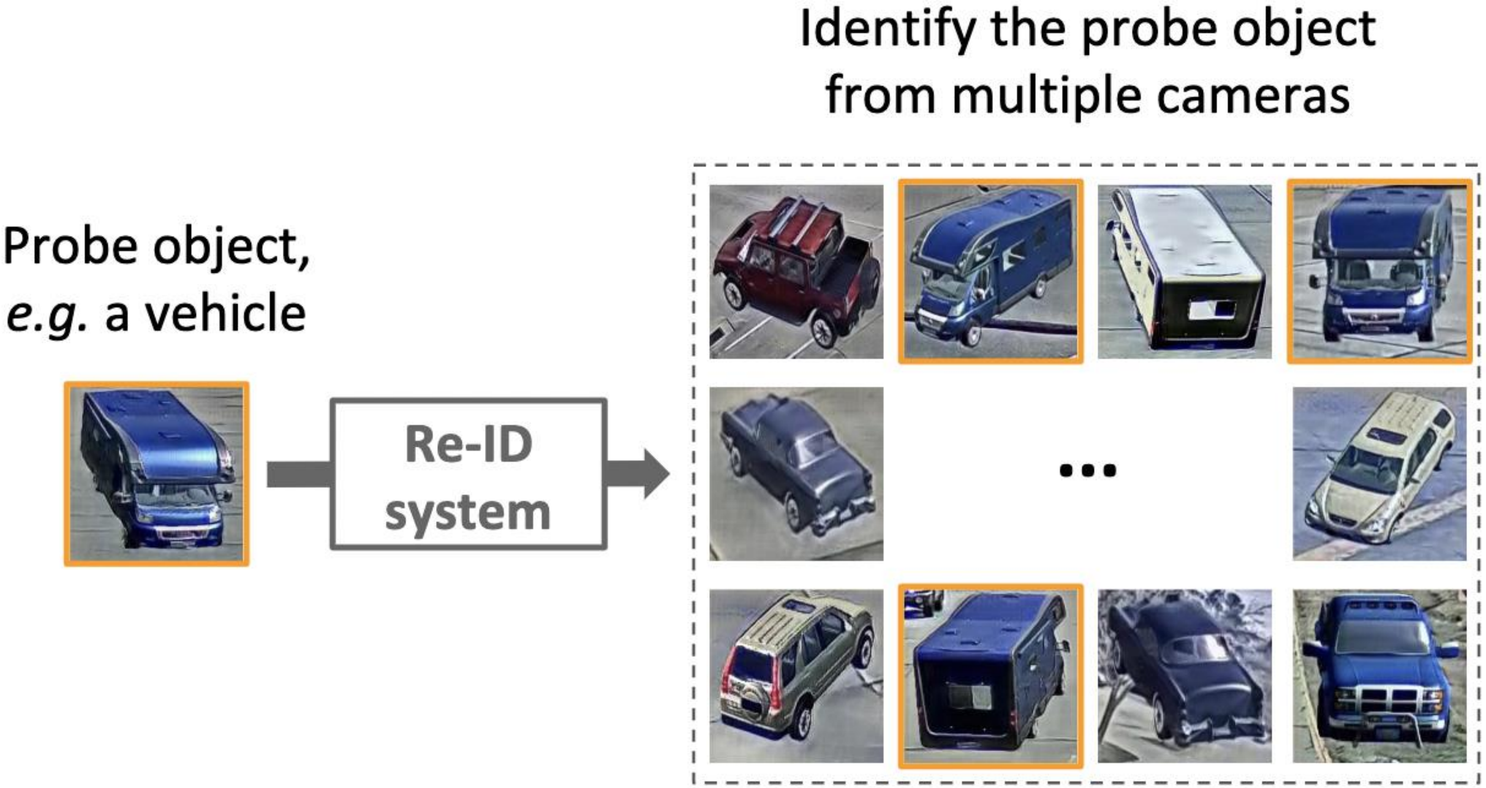
Yixiao Ge **Feng Zhu** **Dapeng Chen** **Rui Zhao** **Hongsheng Li**

Multimedia Laboratory
The Chinese University of Hong Kong

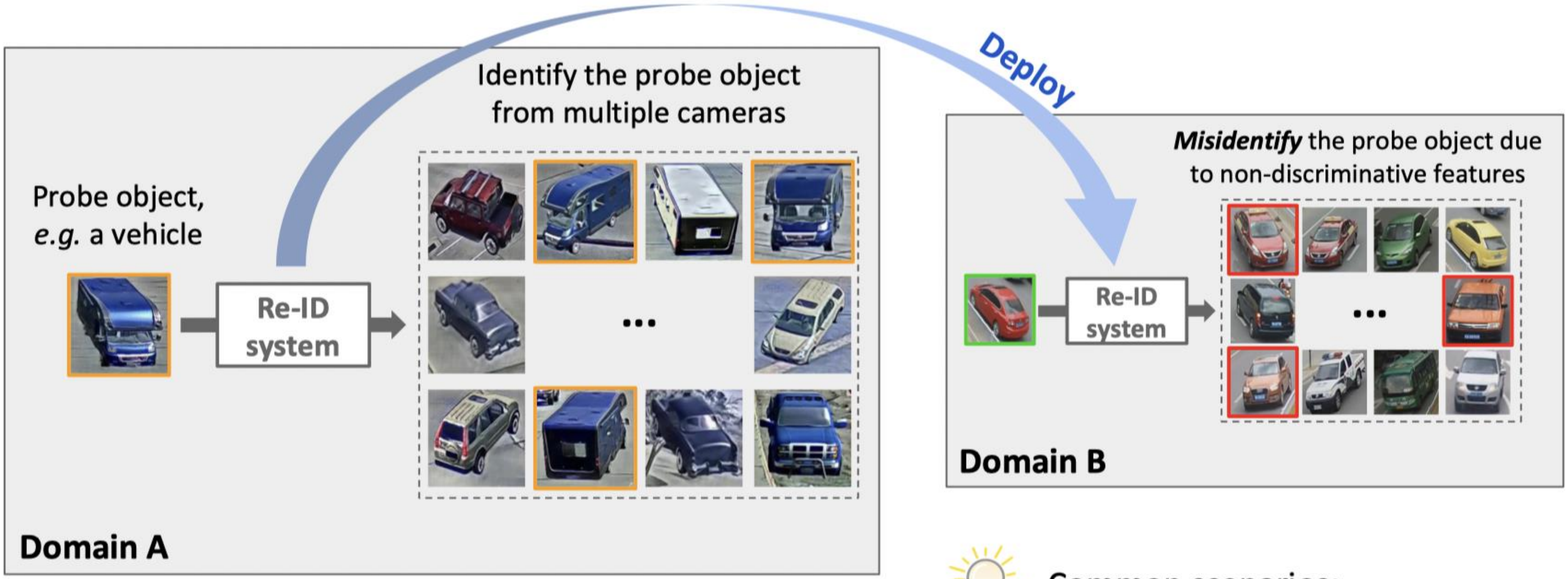
{yxge@link,hsli@ee}.cuhk.edu.hk dapengchenxjtu@gmail.com

NIPS 2020

Object Re-identification (Re-ID)



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



Common scenarios:

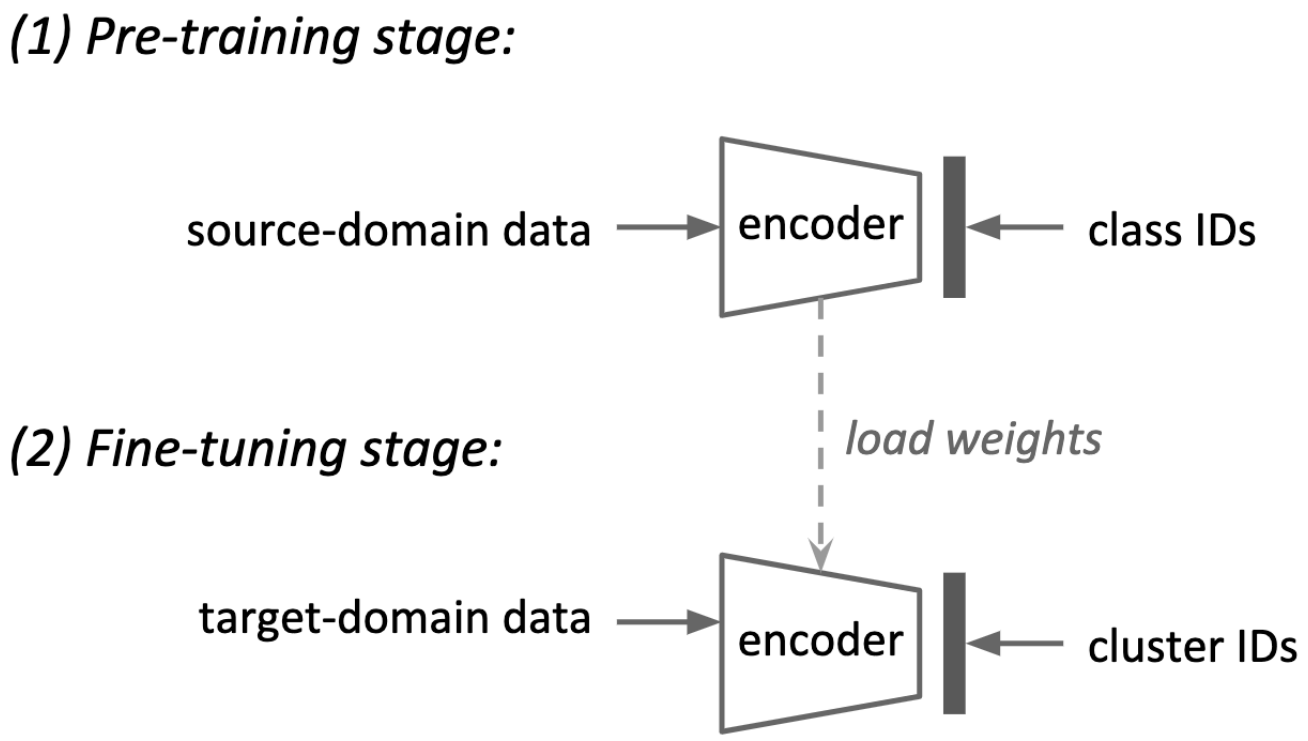
- City A → City B
- Synthetic → Real-world

Previous UDA Methods on Object Re-ID

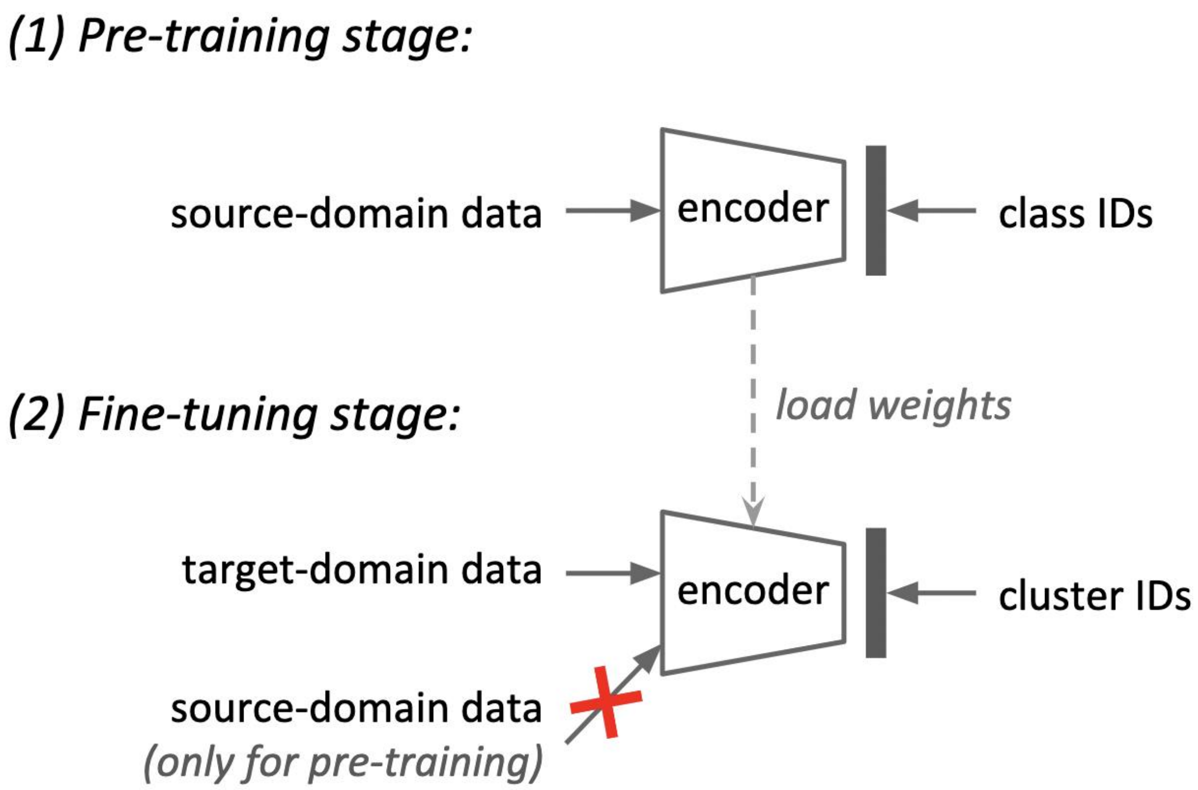
Unsupervised domain adaptation(UDA) re-ID : transfer the knowledge from source domain to target domain

Exist methods generally follow a two-stage scheme:

1. Supervised pre-train on source domain
2. Unsupervised fine-tuning on the target domain(generate pseudo-label by clustering)



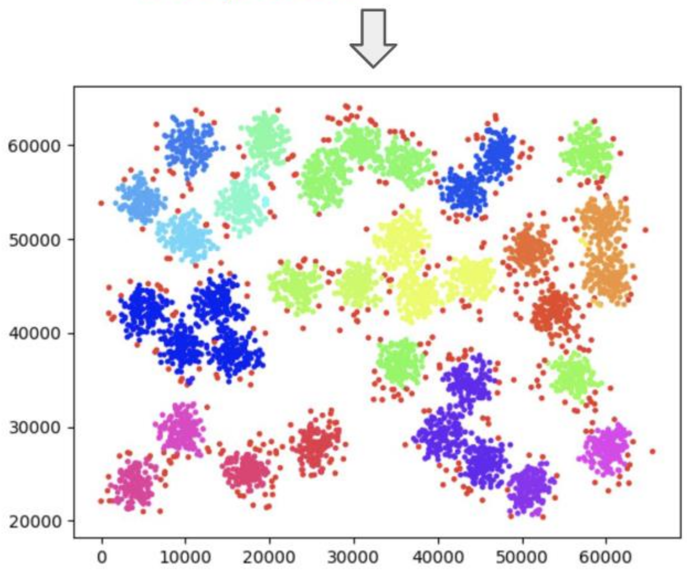
Previous UDA Methods on Object Re-ID



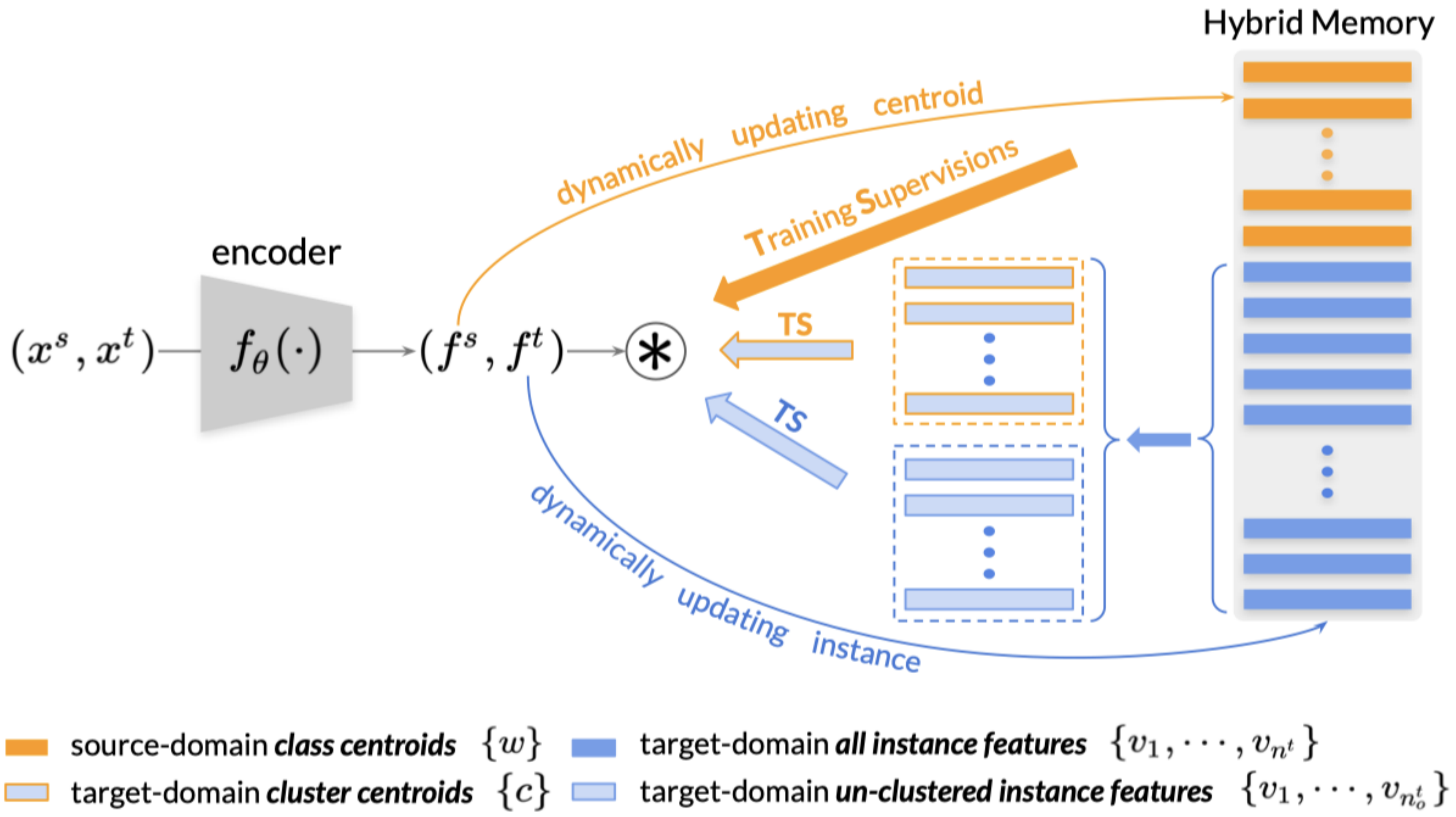
Limitations:

1. The accurate source-domain ground-truth labels are valuable but were ignored during target-domain training.
2. Discard difficult but valuable clustering outlier samples from being used for training.

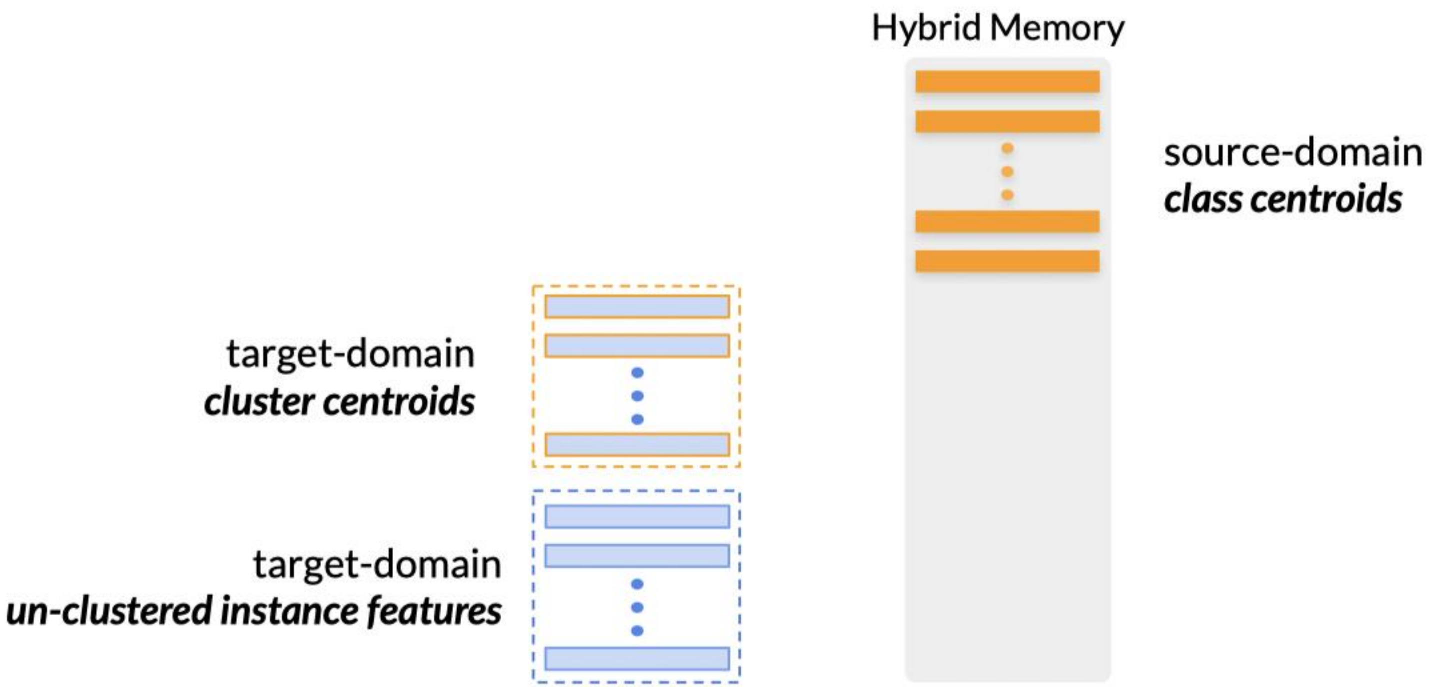
red points are outliers



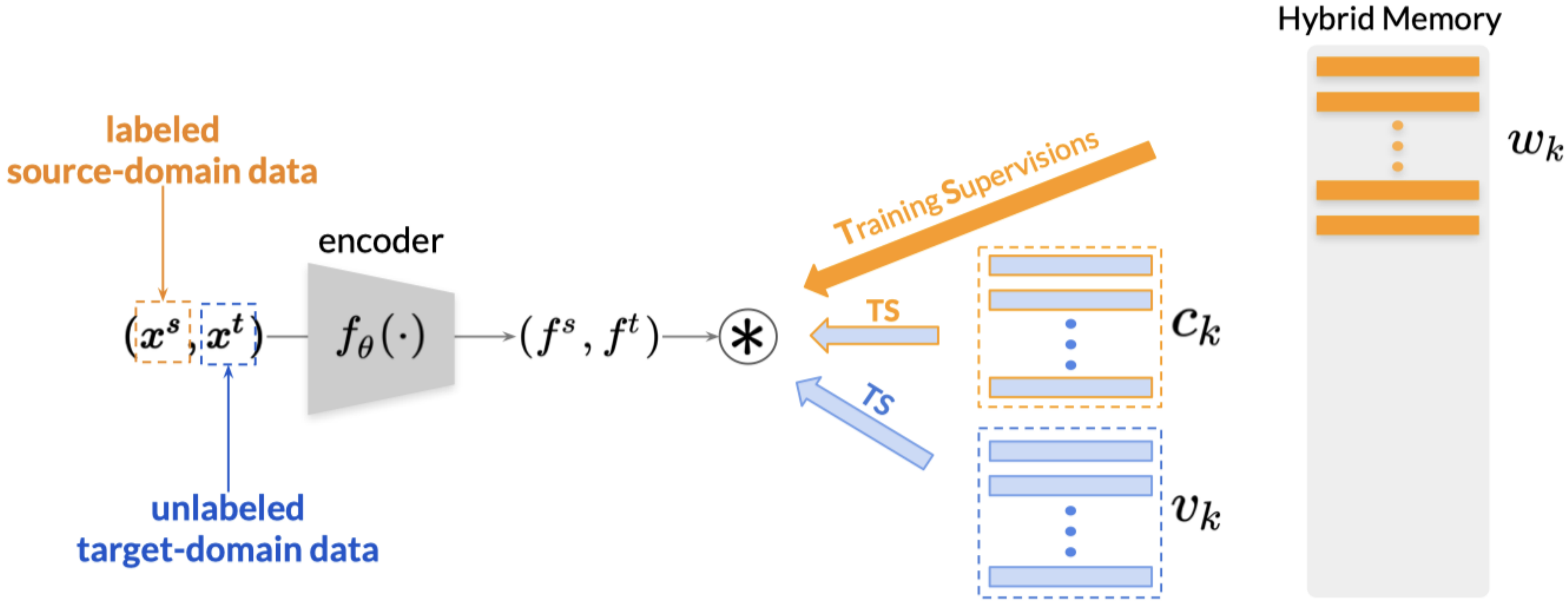
SpCL framework



Prototypes


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

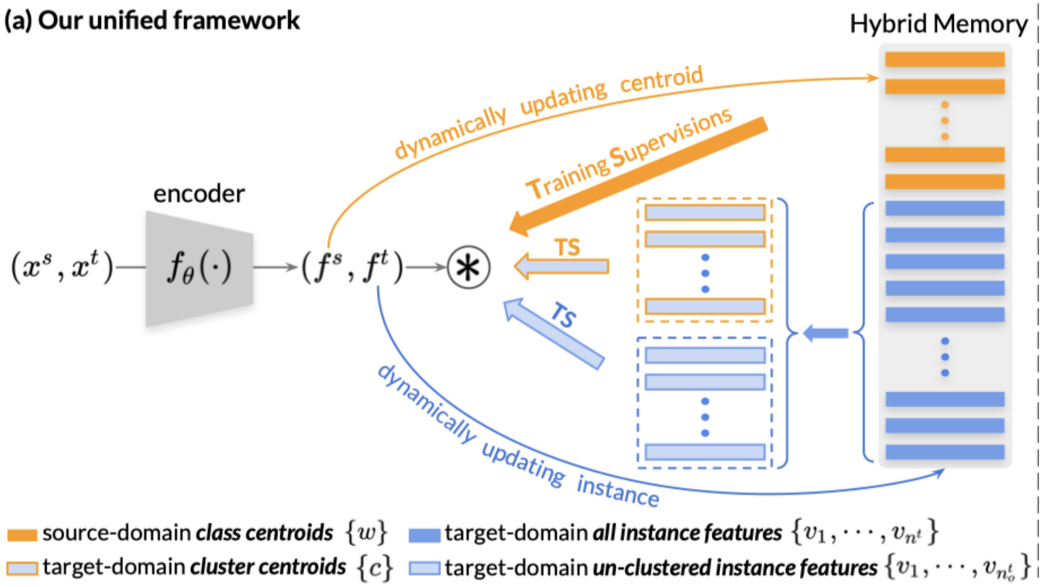
Contrast



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

Propose a self-paced contrastive learning framework:



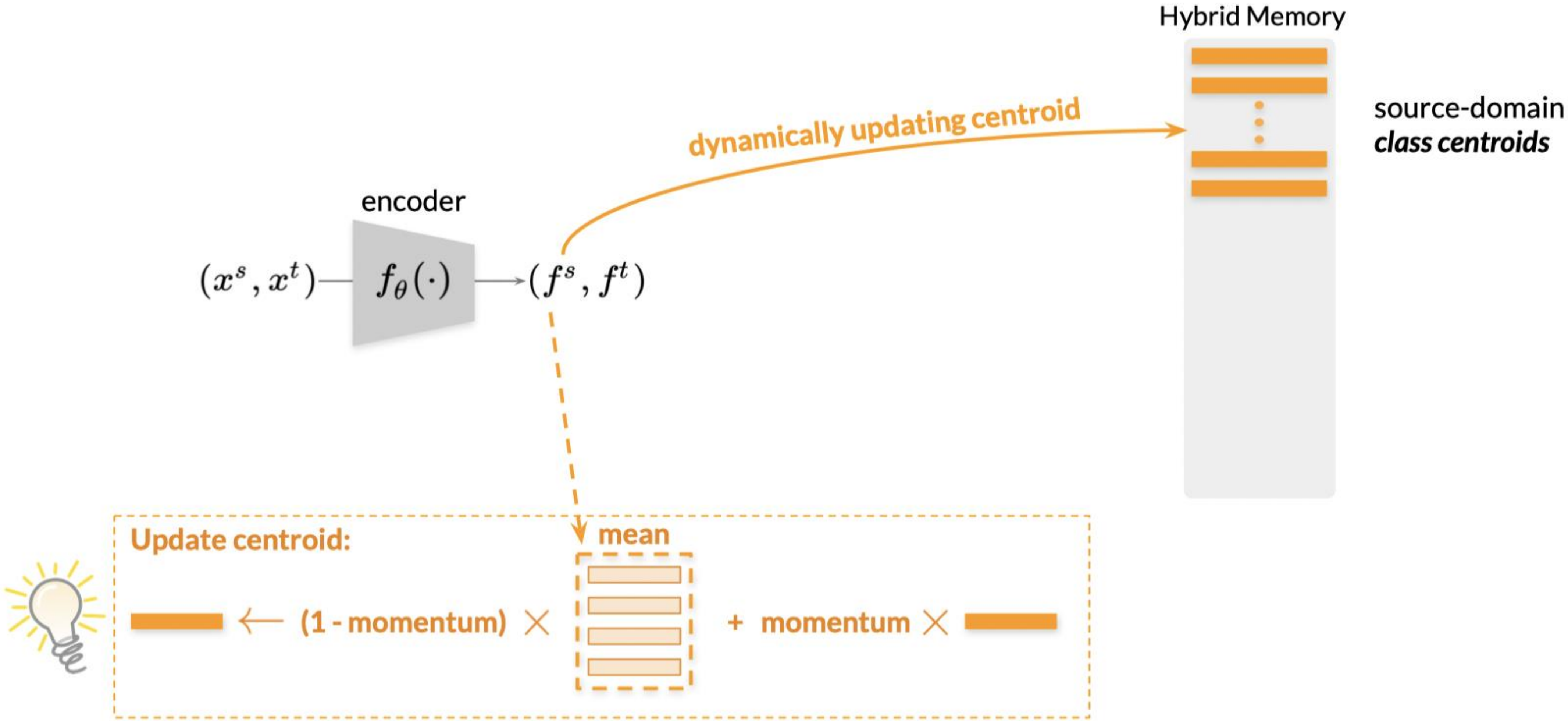
x^s, x_c^t, x_o^t:{源域，目标域簇，目标域离群点}样本
w_k源域类别k的中心（取平均）
c_k 目标域第k个簇的中心
v_k 目标域的离群样本k

Unified Contrastive Learning Loss:

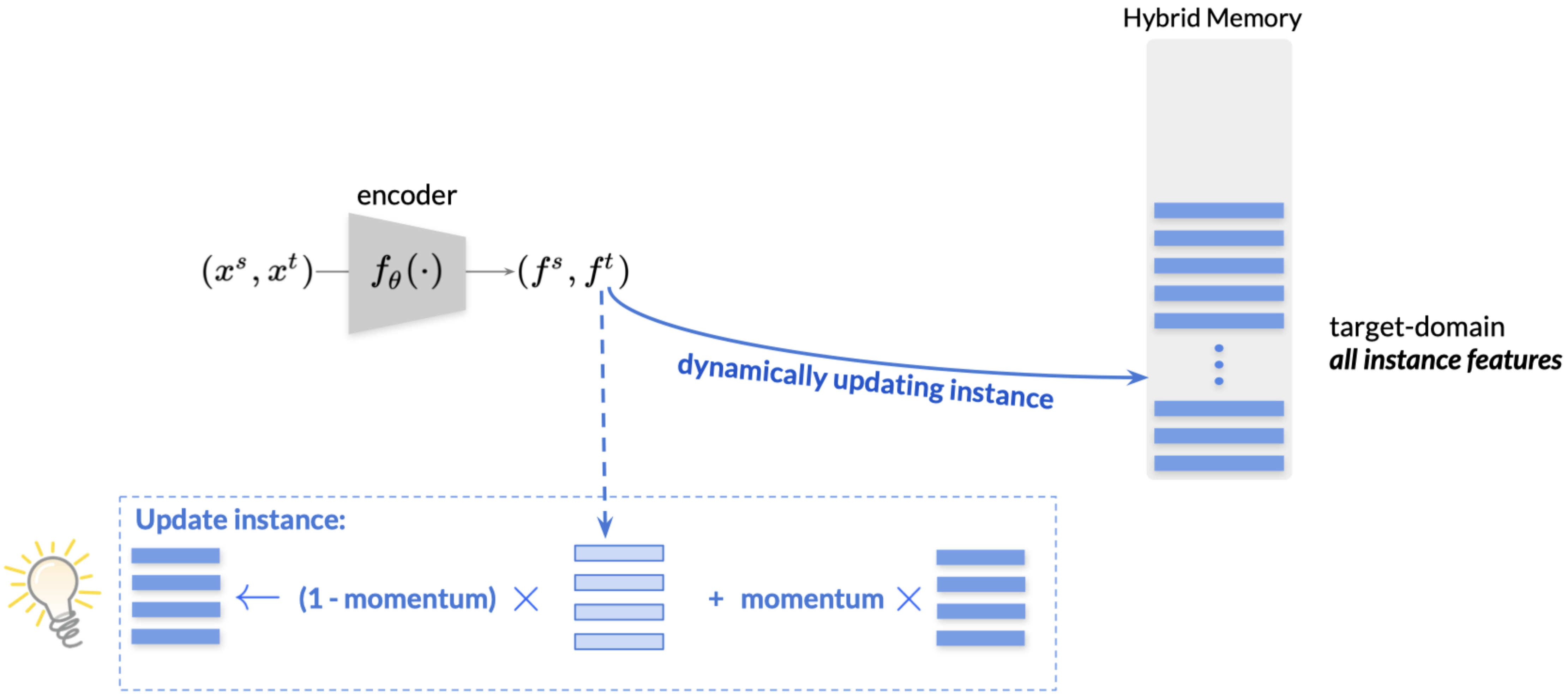
Given a general feature vector $\boldsymbol{f} = f_\theta(x), x \in \mathbb{X}^s \cup \mathbb{X}_c^t \cup \mathbb{X}_o^t$, our unified contrastive loss is

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

Update Memory -- Source-domain Class Centroids



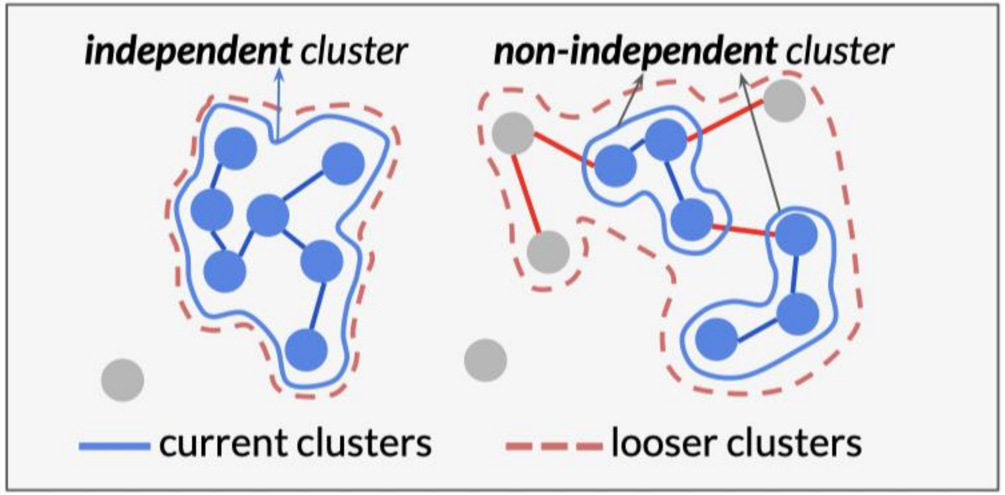
Update Memory -- Target-domain Instance Features



Cluster Reliability Criterion

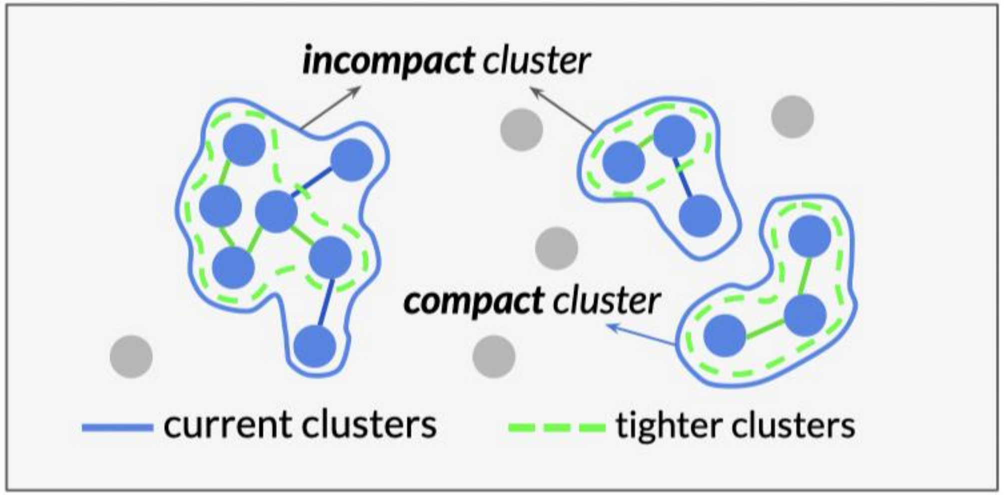
南京航空航天大学 Nanjing University of Aeronautics and Astronautics

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.

Algorithm

Algorithm 1 Self-paced contrastive learning algorithm on domain adaptive object re-ID

Require: Source-domain labeled data $\mathbb{X}^s$ and target-domain unlabeled data $\mathbb{X}^t$;
Require: Initialize the backbone encoder f_θ with ImageNet-pretrained ResNet-50;
Require: Initialize the hybrid memory with features extracted by f_θ;
Require: Temperature τ for Eq. (1), momentum m^s for Eq. (3), momentum m^t for Eq. (4);
for n in $[1, \text{num\_epochs}]$ **do**
  Group $\mathbb{X}^t$ into $\mathbb{X}^t_c$ and $\mathbb{X}^t_o$ by clustering $\{\boldsymbol{v}\}$ from the hybrid memory with the independence Eq. (5) and compactness Eq. (6) criterion;
  Initialize the cluster centroids $\{\boldsymbol{c}\}$ with Eq. (2) in the hybrid memory;
  for each mini-batch $\{x^s_i\} \subset \mathbb{X}^s$, $\{x^t_i\} \subset \mathbb{X}^t$ **do**
    1: Encode features $\{\boldsymbol{f}^s_i\}$, $\{\boldsymbol{f}^t_i\}$ for $\{x^s_i\}$, $\{x^t_i\}$ with f_θ;
    2: Compute the unified contrastive loss with $\{\boldsymbol{f}^s_i\}$, $\{\boldsymbol{f}^t_i\}$ by Eq. (1) and update the encoder f_θ by back-propagation;
    3: Update source-domain related class centroids $\{\boldsymbol{w}\}$ in the hybrid memory with $\{\boldsymbol{f}^s_i\}$ and momentum m^s (Eq. (3));
    4: Update target-domain related instance features $\{\boldsymbol{v}\}$ in the hybrid memory with $\{\boldsymbol{f}^t_i\}$ and momentum m^t (Eq. (4));
    5: Update target-domain related cluster centroids $\{\boldsymbol{c}\}$ with updated $\{\boldsymbol{v}\}$ in the hybrid memory (Eq. (2));
  end for
end for

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

$$\boldsymbol{c}_k = \frac{1}{|\mathcal{I}_k|}\sum_{\boldsymbol{v}_i \in \mathcal{I}_k} \boldsymbol{v}_i, \tag{2}$$

$$\boldsymbol{w}_k \leftarrow m^s \boldsymbol{w}_k + (1-m^s)\cdot \frac{1}{|\mathcal{B}_k|}\sum_{\boldsymbol{f}^s_i \in \mathcal{B}_k} \boldsymbol{f}^s_i, \tag{3}$$

$$\boldsymbol{v}_i \leftarrow m^t \boldsymbol{v}_i + (1-m^t)\boldsymbol{f}^t_i, \tag{4}$$

$$\mathcal{R}_{\text{indep}}(\boldsymbol{f}^t_i) = \frac{|\mathcal{I}(\boldsymbol{f}^t_i) \cap \mathcal{I}_{\text{loose}}(\boldsymbol{f}^t_i)|}{|\mathcal{I}(\boldsymbol{f}^t_i) \cup \mathcal{I}_{\text{loose}}(\boldsymbol{f}^t_i)|} \in [0,1], \tag{5}$$

$$\mathcal{R}_{\text{comp}}(\boldsymbol{f}^t_i) = \frac{|\mathcal{I}(\boldsymbol{f}^t_i) \cap \mathcal{I}_{\text{tight}}(\boldsymbol{f}^t_i)|}{|\mathcal{I}(\boldsymbol{f}^t_i) \cup \mathcal{I}_{\text{tight}}(\boldsymbol{f}^t_i)|} \in [0,1], \tag{6}$$

Domain Adaptive Object Re-ID Performance

南京航空航天大学
Nanjing University of Aeronautics and Astronautics

Table 2: Comparison with state-of-the-art methods on unsupervised domain adaptation for object re-ID. (*) the implementation is based on the authors' code.

(a) *Real*→*real* adaptation on person re-ID datasets.

Methods		Market-1501→MSMT17 mAP	top-1	top-5	top-10
PTGAN [46]	CVPR'18	2.9	10.2	-	24.4
ECN [62]	CVPR'19	8.5	25.3	36.3	42.1
SSG [10]	ICCV'19	13.2	31.6	-	49.6
ECN++ [63]	TPAMI'20	15.2	40.4	53.1	58.7
MMT-$k$means [11]	ICLR'20	22.9	49.2	63.1	68.8
MMCL [45]	CVPR'20	15.1	40.8	51.8	56.7
DG-Net++ [66]	ECCV'20	22.1	48.4	60.9	66.1
D-MMD [31]	ECCV'20	13.5	29.1	46.3	54.1
JVTC [23]	ECCV'20	20.3	45.4	58.4	64.3
GPR [29]	ECCV'20	20.4	43.7	56.1	61.9
NRMT [57]	ECCV'20	19.8	43.7	56.5	62.2
MMT-dbscan [11]*	ICLR'20	24.0	50.1	63.5	69.3
Ours		**26.8**	**53.7**	**65.0**	**69.8**

Methods		MSMT17→Market-1501 mAP	top-1	top-5	top-10
MAR [52]	CVPR'19	40.0	67.7	81.9	-
PAUL [50]	CVPR'19	40.1	68.5	82.4	87.4
CASCL [47]	ICCV'19	35.5	65.4	80.6	86.2
DG-Net++ [66]	ECCV'20	64.6	83.1	91.5	94.3
D-MMD [31]	ECCV'20	50.8	72.8	88.1	92.3
MMT-dbscan [11]*	ICLR'20	75.6	89.3	95.8	97.5
Ours		**77.5**	**89.7**	**96.1**	**97.6**

(b) *Synthetic*→*real* adaptation on person re-ID datasets.

Methods		PersonX→MSMT17 mAP	top-1	top-5	top-10
MMT-dbscan [11]*	ICLR'20	17.7	39.1	52.6	58.5
Ours		**22.7**	**47.7**	**60.0**	**65.5**

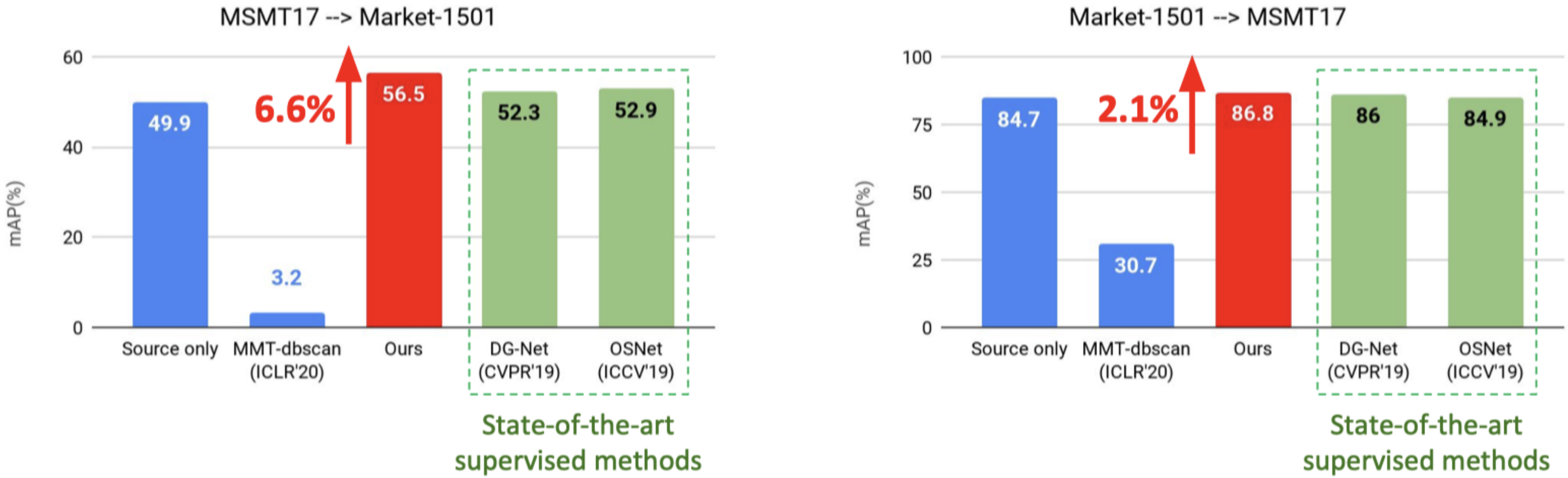
Methods		PersonX→Market-1501 mAP	top-1	top-5	top-10
MMT-dbscan [11]*	ICLR'20	71.0	86.5	94.8	**97.0**
Ours		**73.8**	**88.0**	**95.3**	96.9

(c) *Real*→*real* and *synthetic*→*real* adaptation on vehicle re-ID datasets.

Methods		VehicleID→VeRi-776 mAP	top-1	top-5	top-10
MMT-dbscan [11]*	ICLR'20	35.3	74.6	82.6	87.0
Ours		**38.9**	**80.4**	**86.8**	**89.6**

Methods		VehicleX→VeRi-776 mAP	top-1	top-5	top-10
MMT-dbscan [11]*	ICLR'20	35.6	76.0	83.1	87.4
Ours		**38.9**	**81.3**	**87.3**	**90.0**

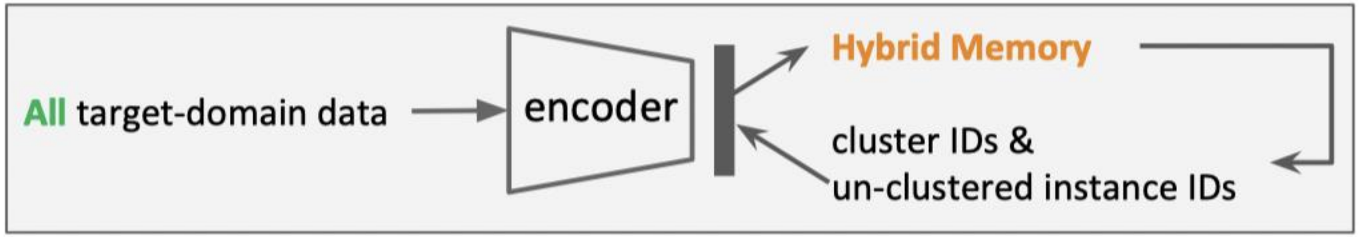
Performance on the Source Domain



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.

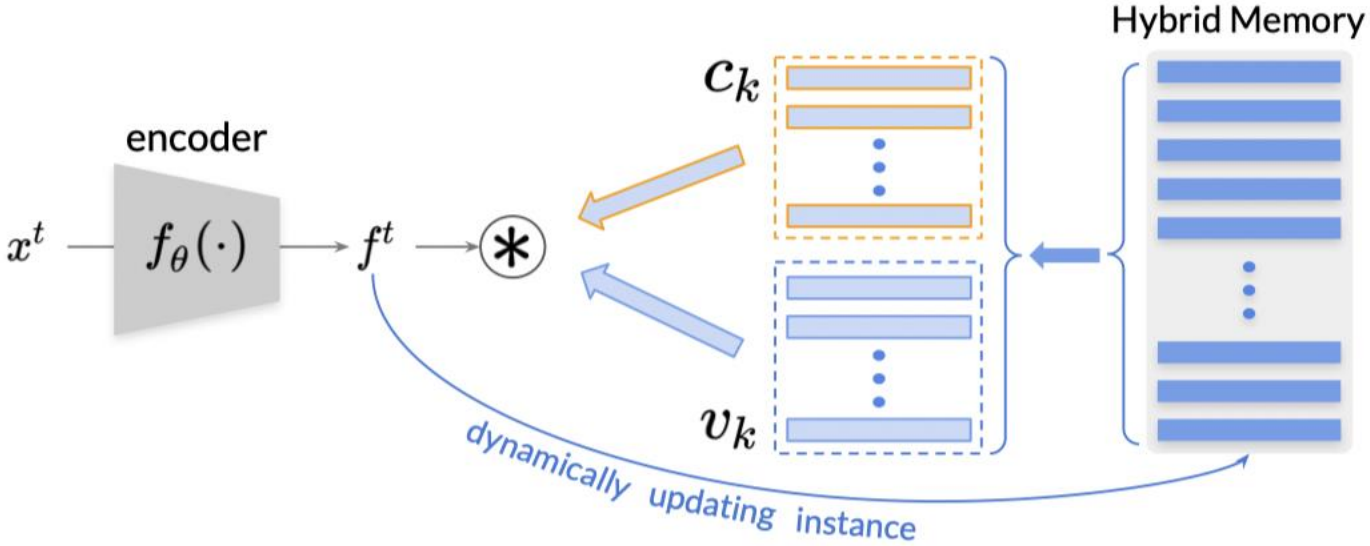
Unsupervised Person Re-ID

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*** $\{c\}$

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



Algorithm

Algorithm 2 Self-paced contrastive learning algorithm on unsupervised object re-ID

Require: Unlabeled data $\mathbb{X}^t$;
Require: Initialize the backbone encoder f_θ with ImageNet-pretrained ResNet-50;
Require: Initialize the hybrid memory with features extracted by f_θ;
Require: Temperature τ for Eq. (1), momentum m^t for Eq. (4);
for n in $[1, \text{num\_epochs}]$ **do**
Group $\mathbb{X}^t$ into $\mathbb{X}^t_c$ and $\mathbb{X}^t_o$ by clustering $\{\boldsymbol{v}\}$ from the hybrid memory with the independence Eq. (5) and compactness Eq. (6) criterion;
Initialize the cluster centroids $\{\boldsymbol{c}\}$ with Eq. (2) in the hybrid memory;
for each mini-batch $\{x^t_i\} \subset \mathbb{X}^t$ **do**
1: Encode features $\{\boldsymbol{f}^t_i\}$ for $\{x^t_i\}$ with f_θ;
2: Compute the unsupervised-version unified contrastive loss with $\{\boldsymbol{f}^t_i\}$ as below and update the encoder f_θ by back-propagation;

$$\mathcal{L}_{\boldsymbol{f}} = -\log \frac{\exp(\langle \boldsymbol{f}, \boldsymbol{z}^+ \rangle / \tau)}{\sum_{k=1}^{n^t_c} \exp(\langle \boldsymbol{f}, \boldsymbol{c}_k \rangle / \tau) + \sum_{k=1}^{n^t_o} \exp(\langle \boldsymbol{f}, \boldsymbol{v}_k \rangle / \tau)}$$

3: Update instance features $\{\boldsymbol{v}\}$ in the hybrid memory with $\{\boldsymbol{f}^t_i\}$ and momentum m^t (Eq. (4));
4: Update cluster centroids $\{\boldsymbol{c}\}$ with updated $\{\boldsymbol{v}\}$ in the hybrid memory (Eq. (2));
end for
end for

$$\mathcal{L}_{\boldsymbol{f}} = -\log \frac{\exp(\langle \boldsymbol{f}, \boldsymbol{z}^+ \rangle / \tau)}{\sum_{k=1}^{n^s} \exp(\langle \boldsymbol{f}, \boldsymbol{w}_k \rangle / \tau) + \sum_{k=1}^{n^t_c} \exp(\langle \boldsymbol{f}, \boldsymbol{c}_k \rangle / \tau) + \sum_{k=1}^{n^t_o} \exp(\langle \boldsymbol{f}, \boldsymbol{v}_k \rangle / \tau)}, \quad (1)$$

$$\boldsymbol{c}_k = \frac{1}{|\mathcal{I}_k|} \sum_{\boldsymbol{v}_i \in \mathcal{I}_k} \boldsymbol{v}_i, \quad (2)$$

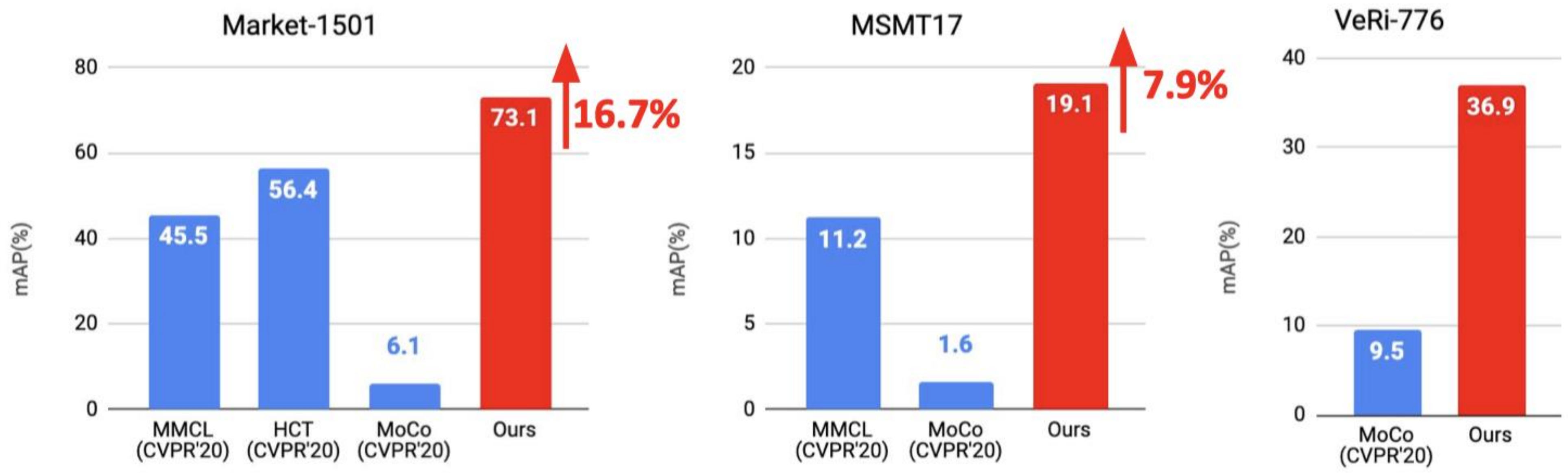
$$\boldsymbol{w}_k \leftarrow m^s \boldsymbol{w}_k + (1 - m^s) \cdot \frac{1}{|\mathcal{B}_k|} \sum_{\boldsymbol{f}^s_i \in \mathcal{B}_k} \boldsymbol{f}^s_i, \quad (3)$$

$$\boldsymbol{v}_i \leftarrow m^t \boldsymbol{v}_i + (1 - m^t) \boldsymbol{f}^t_i, \quad (4)$$

$$\mathcal{R}_{\text{indep}}(\boldsymbol{f}^t_i) = \frac{|\mathcal{I}(\boldsymbol{f}^t_i) \cap \mathcal{I}_{\text{loose}}(\boldsymbol{f}^t_i)|}{|\mathcal{I}(\boldsymbol{f}^t_i) \cup \mathcal{I}_{\text{loose}}(\boldsymbol{f}^t_i)|} \in [0, 1], \quad (5)$$

$$\mathcal{R}_{\text{comp}}(\boldsymbol{f}^t_i) = \frac{|\mathcal{I}(\boldsymbol{f}^t_i) \cap \mathcal{I}_{\text{tight}}(\boldsymbol{f}^t_i)|}{|\mathcal{I}(\boldsymbol{f}^t_i) \cup \mathcal{I}_{\text{tight}}(\boldsymbol{f}^t_i)|} \in [0, 1], \quad (6)$$

Unsupervised Object Re-ID Performance



THANKS