Learning Feature Fusion for Unsupervised Domain Adaptive Person Re-identification

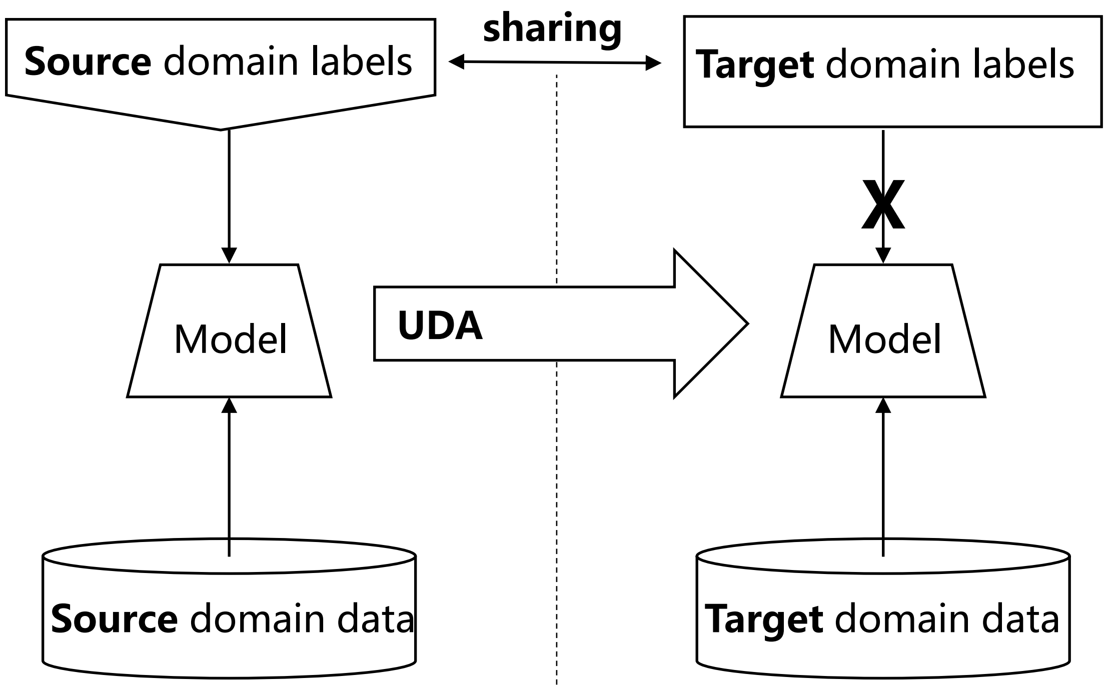
Jin Ding[1], **Xue Zhou**[1,2*]

[1]University of Electronic Science and Technology of China (UESTC),
[2]Shenzhen Institute of Advanced Study, UESTC
[*]Corresponding author:zhouxue@uestc.edu.cn

Contents

- Background
- Motivation
- Method
- Experiments
- Analysis
- Conclusion

Background



$$\mathcal{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$$
$$P(\mathbf{X}_s, \mathbf{Y}_s)$$

$$\mathcal{D}_t = \{x_j^t\}_{j=1}^{n_t}$$
$$Q(\mathbf{X}_t, \mathbf{Y}_t)$$

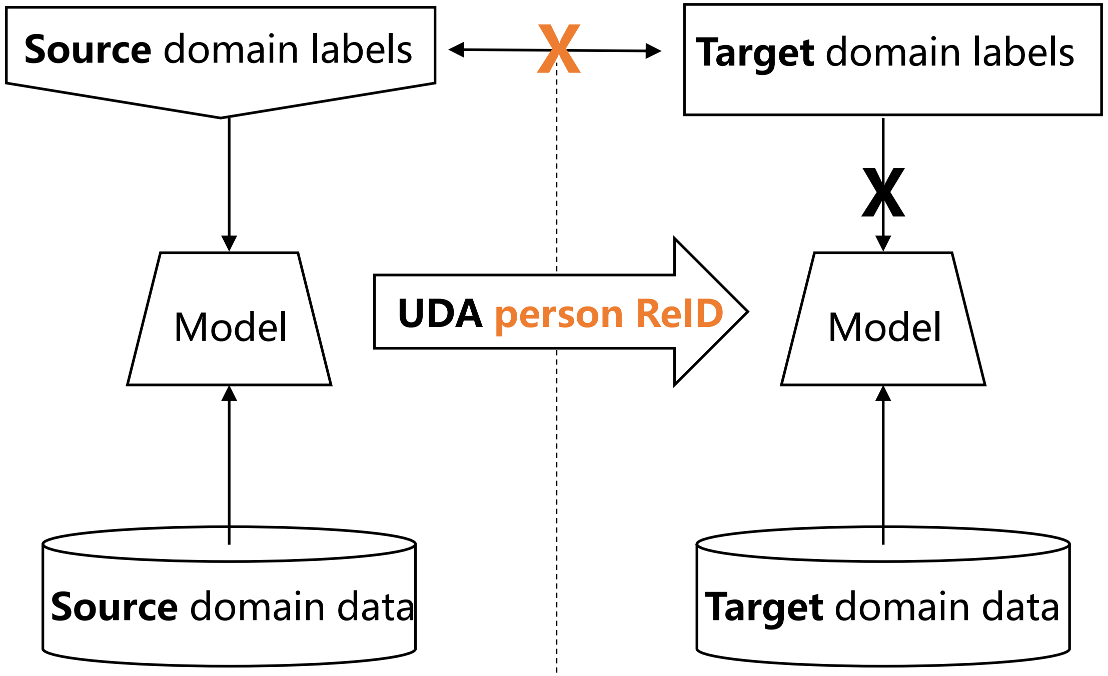
$$P(\mathbf{X}_s, \mathbf{Y}_s) \neq Q(\mathbf{X}_t, \mathbf{Y}_t)$$
$$\mathbf{Y}_s = \mathbf{Y}_t$$

[Office-home dataset (AAAI 2019)] Wen, Jun, et al. Exploiting local feature patterns for unsupervised domain adaptation.

Background

电子科技大学
University of Electronic Science and Technology of China

- **Unsupervised Domain Adaptation (UDA)**
- **Unsupervised Domain Adaptive person Person Re-Identification (ReID)**



$$\mathcal{D}_s = \{(x_i^s, y_i^s)\}_{i=1}^{n_s}$$
$$P(\mathbf{X}_s, \mathbf{Y}_s)$$

$$\mathcal{D}_t = \{x_j^t\}_{j=1}^{n_t}$$
$$Q(\mathbf{X}_t, \mathbf{Y}_t)$$

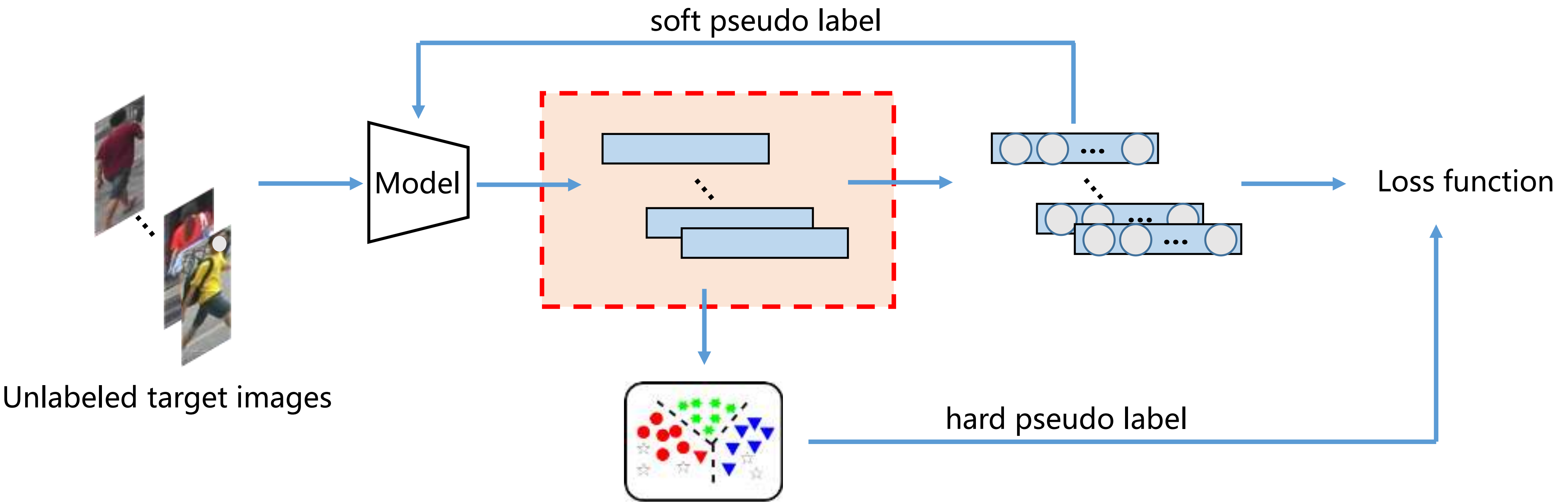
Source

Target

$$P(\mathbf{X}_s, \mathbf{Y}_s) \neq Q(\mathbf{X}_t, \mathbf{Y}_t)$$
$$\mathbf{Y}_s \neq \mathbf{Y}_t$$

Motivation

- Limitation of **Fine-tuning based** UDA person ReID

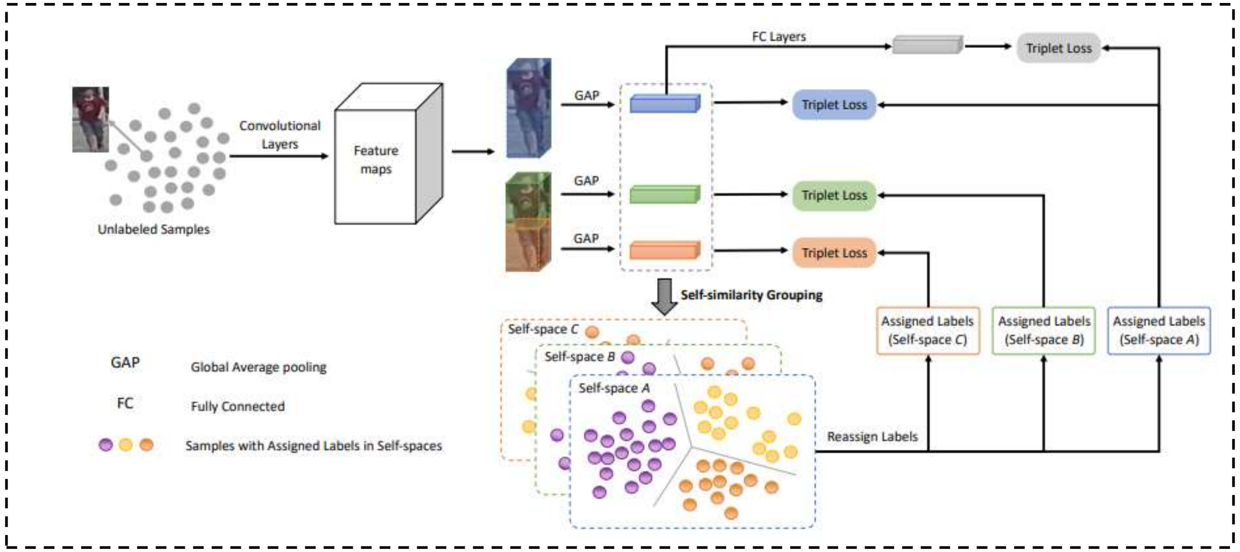


Can we use local features for multi-level clustering?

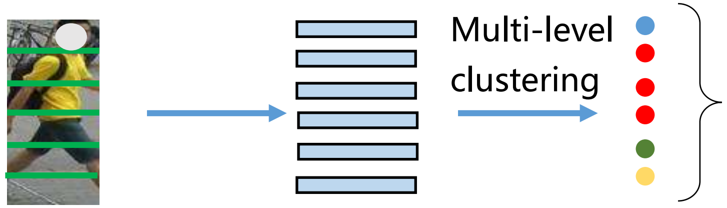
Motivation

电子科技大学 University of Electronic Science and Technology of China

Limitation of **Fine-tuning based** UDA person ReID



- **Noisy pseudo labels**
- **Obscure learning**



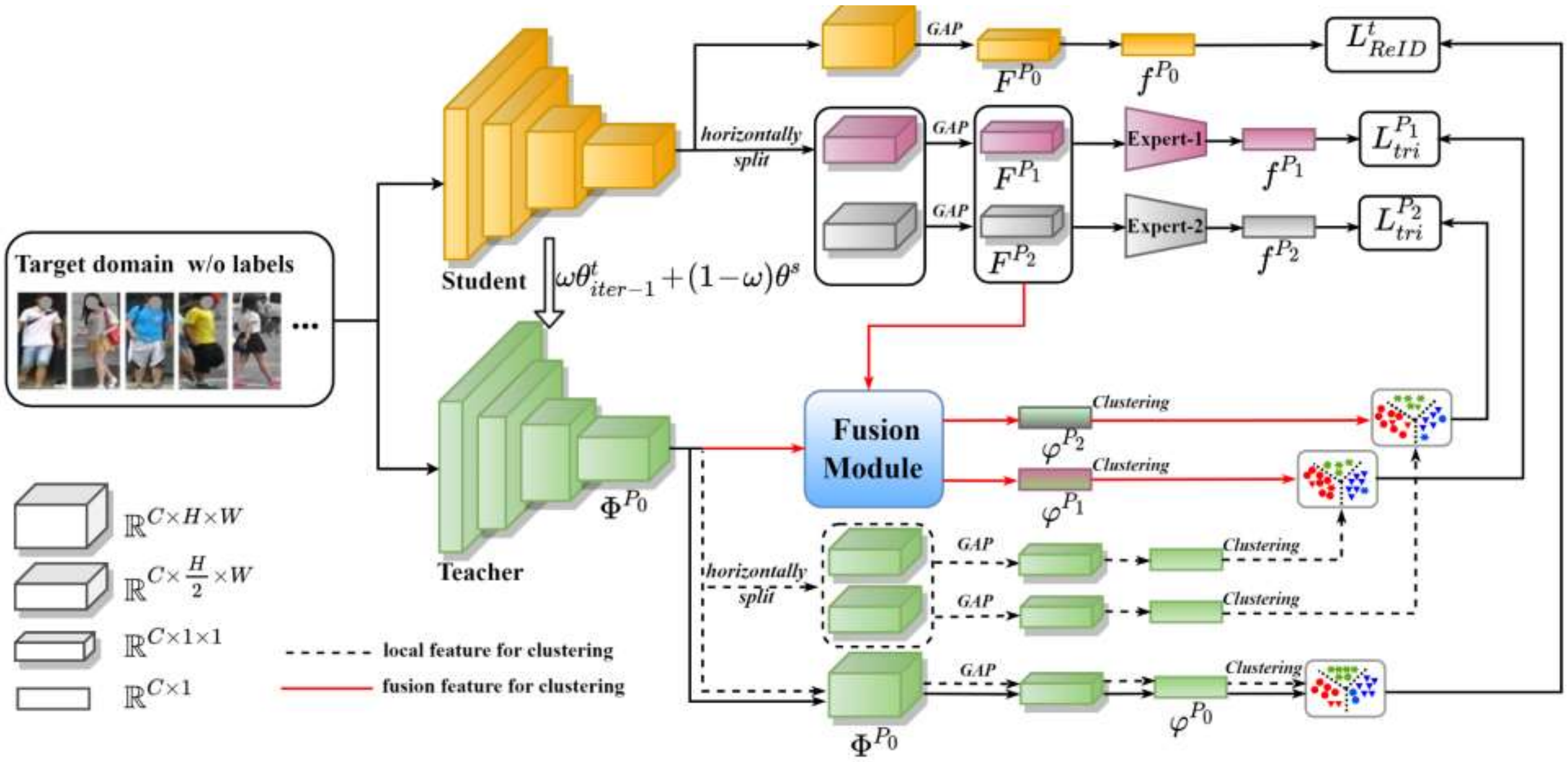
obscure learning

How to avoid obscure learning?

[SSG (ICCV 2019)] Fu, Yang, et al. Self-similarity grouping: A simple unsupervised cross domain adaptation approach for person re-identification.

Method

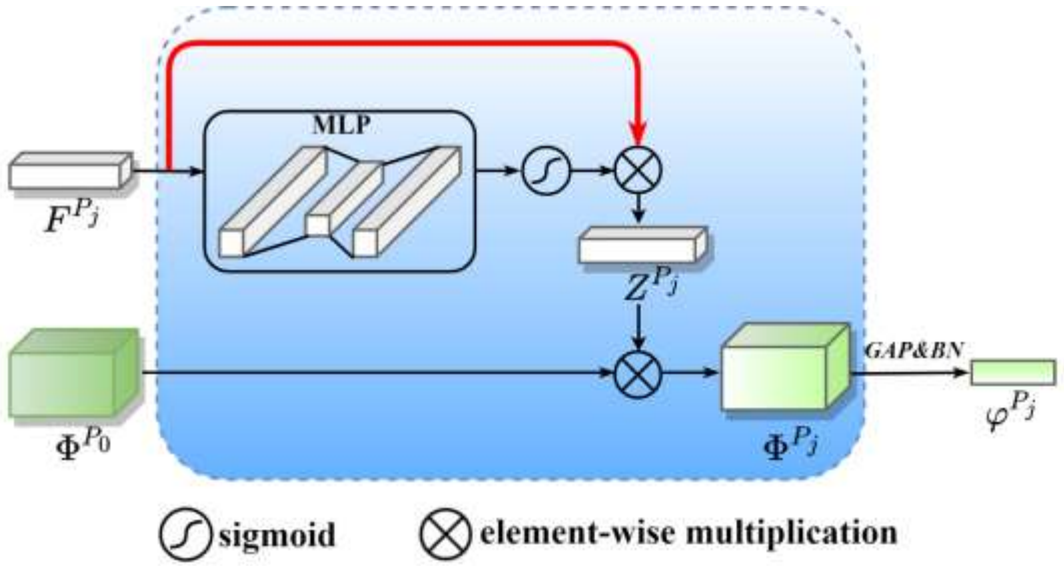
■ Overview



- **Fusion Module:** Fuse the student network's local feature maps and the teacher network's global feature maps.
- **Expert-i:** Align the student network's local feature and the fusion feature.

Method

Fusion Module (FM)



$$
\begin{aligned}
\Phi^{P_j} &= \sigma\left(Z^{P_j}\right) \otimes \Phi^{P_0} \\
&= \sigma\left(MLP\left(F^{P_j}\right) \otimes F^{P_j}\right) \otimes \Phi^{P_0} \\
&= \sigma\left(W_2\left(\mathrm{ReLU}\left(W_1\left(F^{P_j}\right)\right)\right) \otimes F^{P_j}\right) \otimes \Phi^{P_0}
\end{aligned}
$$

- Only the student network’s local feature maps are forwarded to a MLP for adaptively learning fusion.
- The red line is a residual structure to obtain the learned attention map.

Method

Optimization

$$L_{\text{total}} = \alpha L_{\text{ReID}}^{t} + \gamma \sum_{j=1}^{K} L_{tri}^{P_j} = \alpha \left(L_{cls}^{t} + \lambda L_{tri}^{t} \right) + \gamma \sum_{tri}^{K} L_{ti}^{P_j}$$

- **Classification loss:** $$L_{cls}^{t} = \frac{1}{N_t} \sum_{i=1}^{N_t} L_{ce}\left(C^t \left(f^{P_0}(x_i) \right), \hat{y}_{i,0} \right)$$

- **Softmax triplet loss:** $$L_{tri}^{P_j} = -\frac{1}{N_t} \sum_{i=1}^{N_t} \log \mathcal{H}_j(x_i \mid \theta^s)$$

$$\mathcal{H}_j(x_i \mid \theta^s) = \frac{e^{\left\| f^{P_j}(x_i|\theta^s) - f^{P_i}(x_{i,-}|\theta^s) \right\|_2}}{e^{\left\| f^{P_j}(x_i|\theta^s) - f^{P_j}(x_{i,+}|\theta^s) \right\|_2} + e^{\left\| f^{P_j}(x_i|\theta^s) - f^{P_j}(x_{i,-}|\theta^s) \right\|_2}}$$

Experiments

电子科技大学 University of Electronic Science and Technology of China

Training stage

1) First stage: Pretrain on the source domain.

2) Second stage: Fine-tune on the target domain.

Component-wise analysis of the proposed model.

Methods	D-to-M		M-to-D	
	mAP	Rank1	mAP	Rank1
Direct transfer	27.8	55.6	26.9	42.6
Baseline(only L^t_{ReID})	69.0	86.6	61.3	75.6
LF^2 w/o FM	78.5	90.5	68.5	81.5
$LF^2(M_{t,j}$=500)	79.9	91.8	68.7	81.7
$LF^2(M_{t,j}$=700)	**83.2**	**92.8**	72.2	82.9
$LF^2(M_{t,j}$=900)	82.3	92.4	**73.5**	**83.7**

- **D-to-M**: pretrain on Duke and fine-tine on Market.
- **M-to-D**: pretrain on Market and fine-tine on Duke.
- $\boldsymbol{M_{t,j}}$: the number of pseudo identities

- **Direct transfer**: directly using the source-domain pre-trained model to adapt the target domain.
- **Baseline(only L^t_{ReID})**: It only uses the teacher network's global feature for clustering.
- LF^2 w/o FM **:** Replace the fusion features with the teacher network's local features for clustering.

Experiments

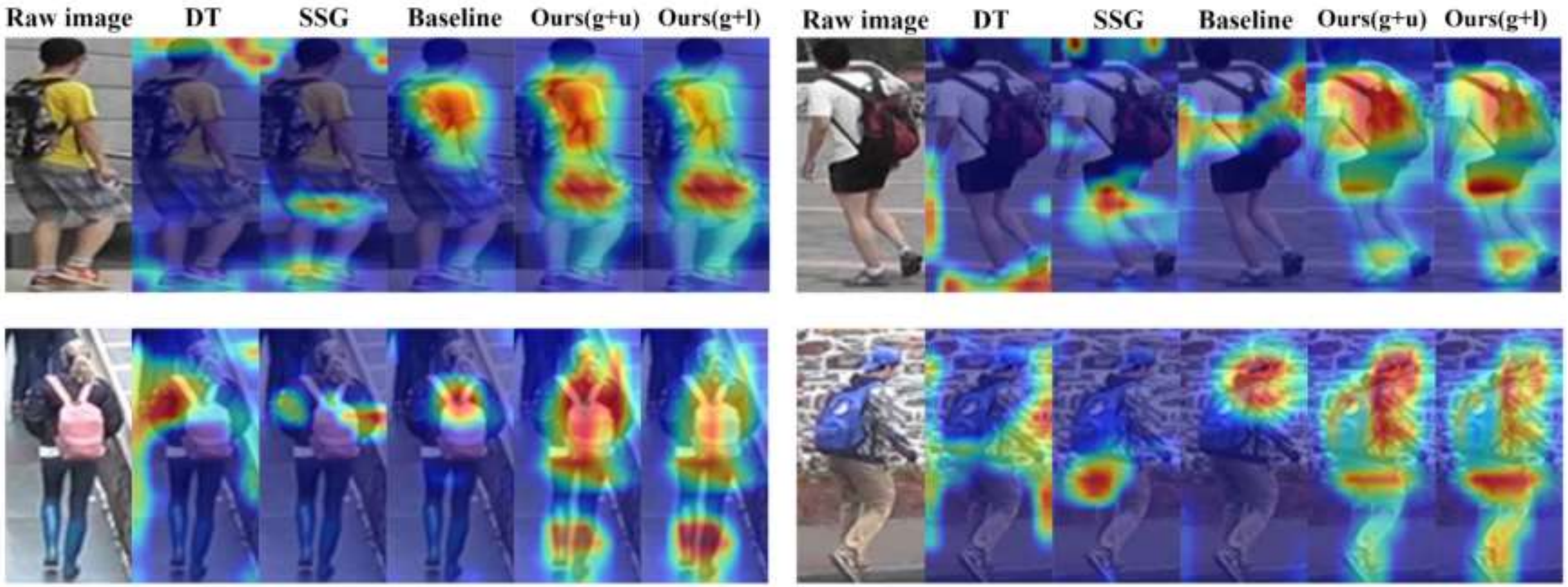
State-of-the-art Comparison

Categories	Methods	Reference	D-to-M				M-to-D			
			mAP	Rank1	Rank5	Rank10	mAP	Rank1	Rank5	Rank10
GAN transferring	SPGAN+LMP [23]	CVPR'18	26.7	57.7	75.8	82.4	26.2	46.4	62.3	68.0
	PDA-Net [24]	ICCV'19	47.6	75.2	86.3	90.2	45.1	63.2	77.0	82.5
Joint learning	ECN [25]	CVPR'19	43.0	75.1	87.6	91.6	40.4	63.3	75.8	80.4
	MMCL [27]	CVPR'20	60.4	84.4	92.8	95.0	51.4	72.4	82.9	85.0
	JVTC+ [26]	ECCV'20	67.2	86.8	95.2	97.1	66.5	80.4	89.9	92.2
	IDM [28]	ICCV'21	**82.8**	**93.2**	**97.5**	**98.1**	**70.5**	**83.6**	**91.5**	**93.7**
Fine-tuning	SSG [7]	ICCV'19	58.3	80.0	90.0	92.4	53.4	73.0	80.6	83.2
	ADTC [9]	ECCV'20	59.7	79.3	90.8	94.1	52.5	71.9	84.1	87.5
	AD-Cluster [8]	CVPR'20	68.3	86.7	94.4	96.5	54.1	72.6	82.5	85.5
	MMT [10]	ICLR'20	71.2	87.7	94.9	96.9	65.1	78.0	88.8	92.5
	MEB-Net [11]	ECCV'20	76.0	89.9	96.0	97.5	66.1	79.6	88.3	92.2
	Dual-Refinement [15]	TIP'21	78.0	90.9	96.4	97.7	67.7	82.1	90.1	92.5
	UNRN [14]	AAAI'21	78.1	91.9	96.1	97.8	69.1	82.0	90.7	**93.5**
	GLT [12]	CVPR'21	79.5	92.2	96.5	97.8	69.2	82.0	90.2	92.8
	HCD [13]	ICCV'21	80.0	91.5	–	–	70.1	82.2	–	–
	P^2LR [32]	AAAI'22	81.0	92.6	97.4	**98.3**	**70.8**	**82.6**	**90.8**	**93.7**
	RDSBN+MDIF [33]	CVPR'21	**81.5**	**92.9**	**97.6**	**98.4**	66.6	80.3	89.1	92.6
	$\boldsymbol{LF^2}$**(Ours)**	This paper	**83.2**	**92.8**	**97.8**	**98.4**	**73.5**	**83.7**	**91.9**	**94.3**

- **Top three** performance values are highlighted in **RED**, **BLUE** and **ORANGE** colors respectively.

Analysis

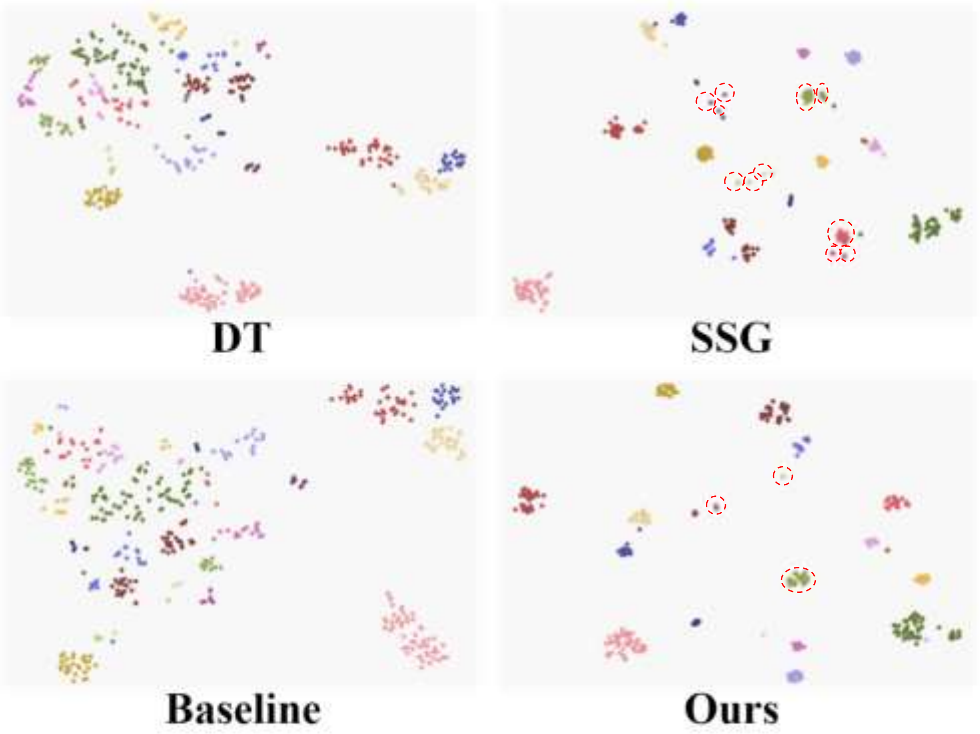
Visualization of feature maps



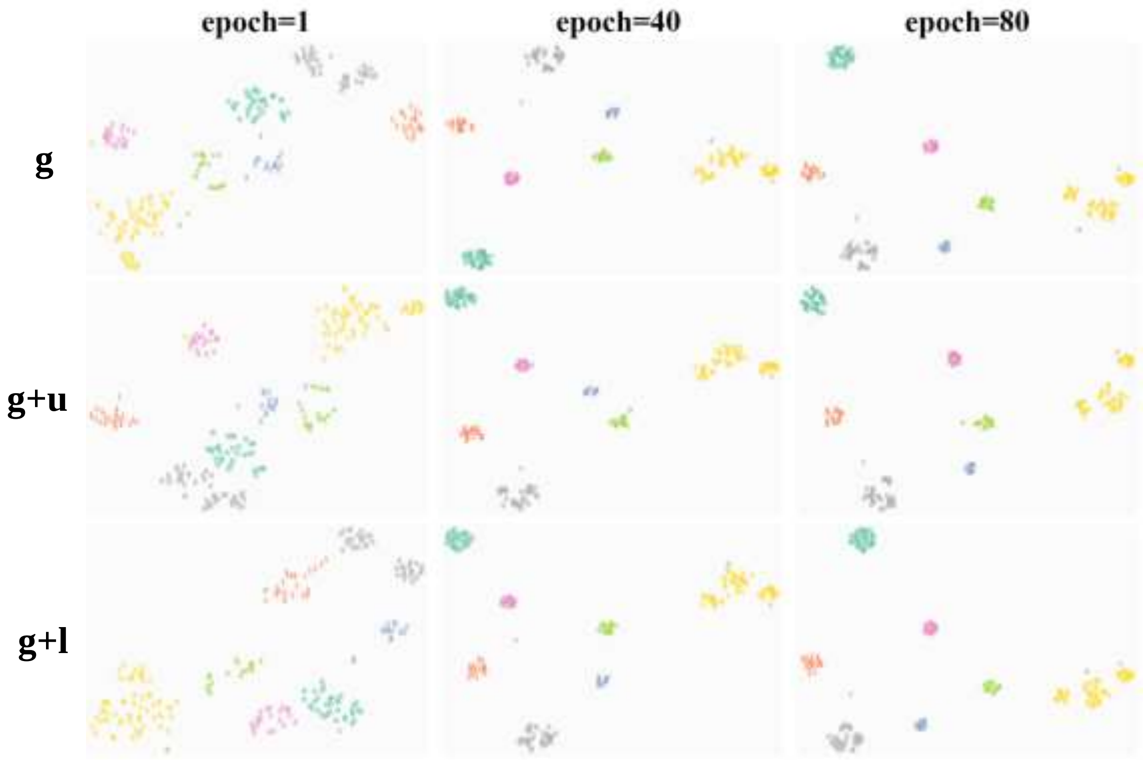
- **DT:** Direct transfer
- **Ours(g+u)**: Fuse the teacher network's **global** feature map and the student network's **upper** local feature map.
- **Ours(g+l)**: Fuse the teacher network's **global** feature map and the student network's **lower** local feature map.

Analysis

■ **Visualization of clustering features**



Visualization of 20 pedestrians on target domain.



Visualization of 10 pedestrians during target-domain fine-tuning with our framework

➢ Different colors indicate different identities

Conclusion

电子科技大学
University of Electronic Science and Technology of China

- **We propose a Learning Feature Fusion (LF2) framework that adaptively learns to fuse global and local features to obtain more comprehensive representations.**

- **A learnable Fusion Module (FM) is proposed to avoid obscure learning of multiple pseudo labels.**

- **Experiments conducted on two common UDA ReID settings show that our method achieves significant performance gain over the state-of-the-arts.**

https://github.com/DJEddyking/LF2

电子科技大学
University of Electronic Science and Technology of China

August 21-25 • Montréal, Québec
ICPR 2022
26TH International Conference on Pattern Recognition

Thank you for your attention!