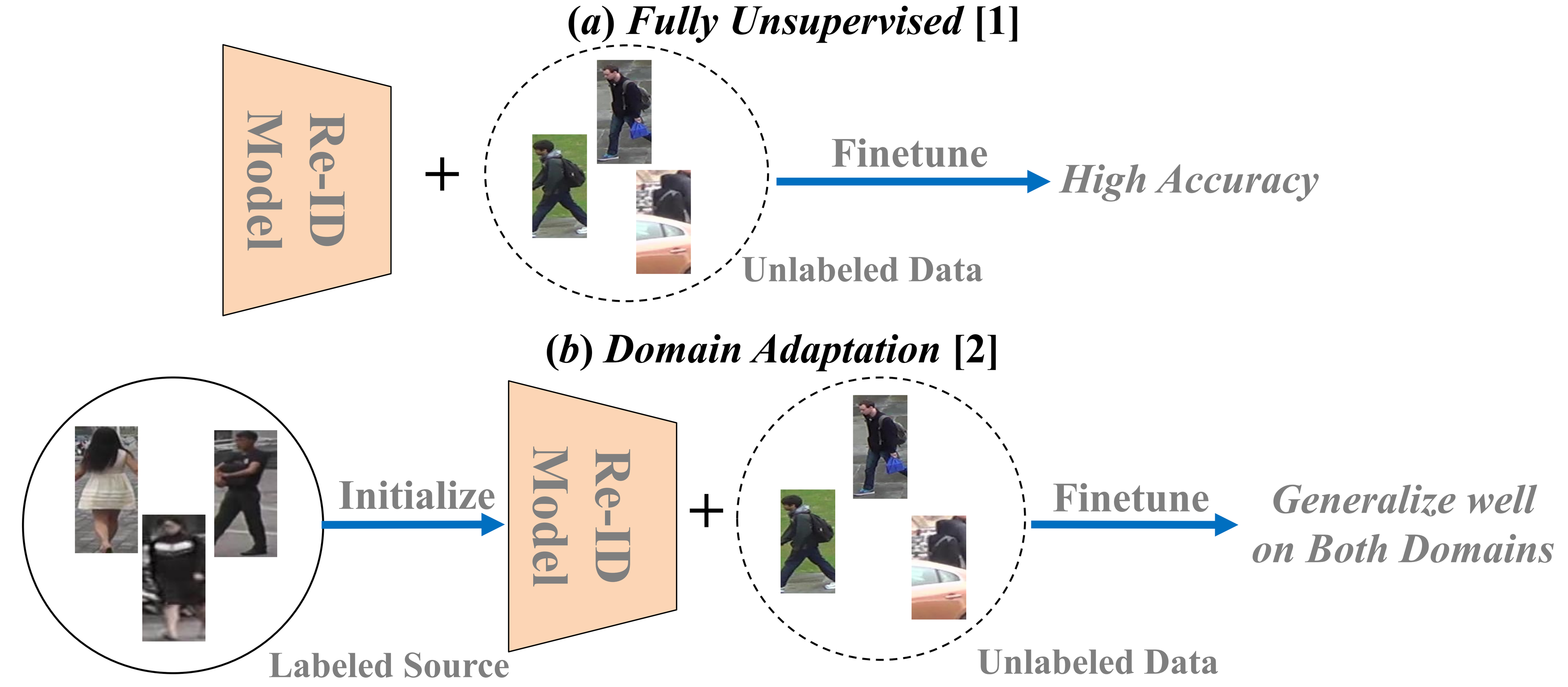
Joint Noise-Tolerant Learning and Meta Camera Shift Adaptation for Unsupervised Person Re-Identification

Fengxiang Yang[1]*, Zhun Zhong[2]*, Zhiming Luo[1], Yuanzheng Cai[3], Yaojin Lin[4], Shaozi Li[1], Nicu Sebe[2]

[1]Xiamen University [2]University of Trento [3]Minjiang University [4]Minnan Normal University

CVPR VIRTUAL JUNE 19-25

Problem Definition



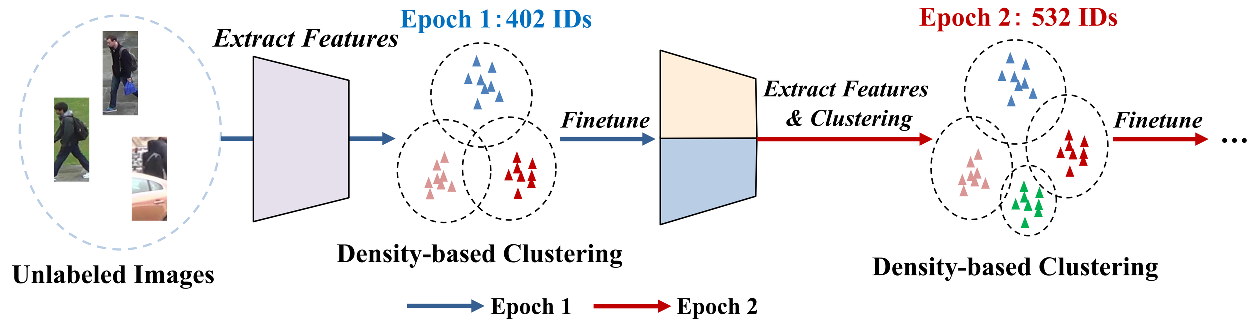
Motivation: Optimizing (Generalizing) models with unlabeled data.

Challenges & Solutions

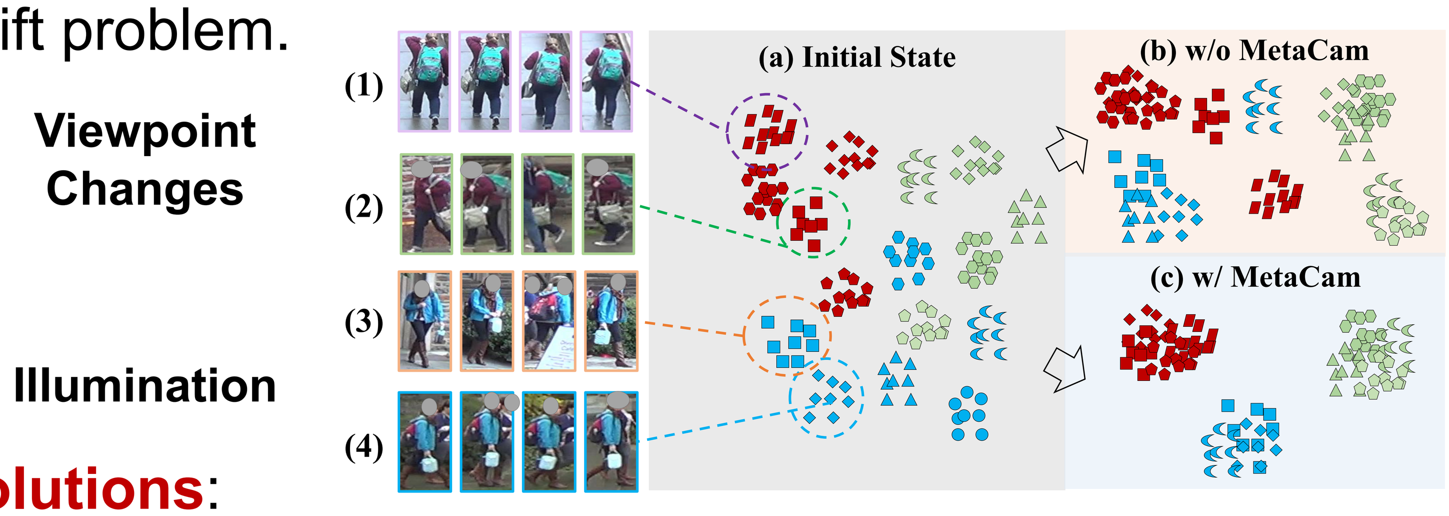
Goal: (1) Noisy-tolerant Optimization **(2)** Overcome camera-shift.

Challenges:

(1) Changing identities during clustering prevents the utilization of theories about noisy label learning like [3].



(2) Re-ID models are supposed to be camera-invariant, which is hard to achieve due to large intra-class variation brought by camera-shift problem.

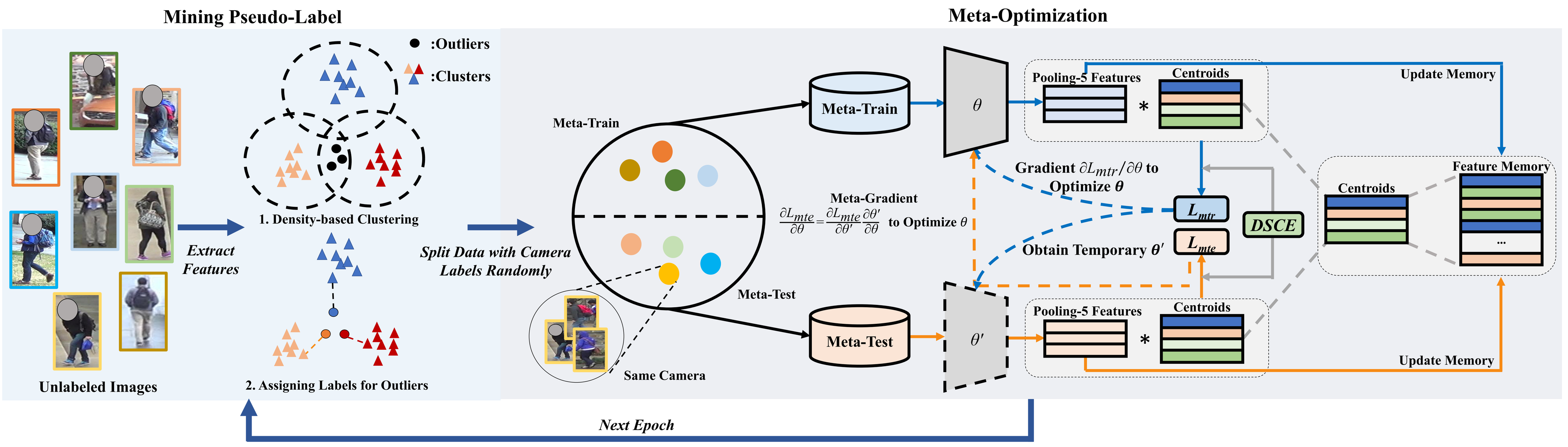


Solutions:

(1) Dynamic & Symmetric Cross-Entropy Loss (DSCE).

(2) Camera-aware meta-learning.

Framework



General Idea:

(1) Loss functions should satisfy "symmetric constriant"[3] to become noise-tolerant. To accommodate to changing IDs, we adopt memory bank.

(2) A good unsupervised re-ID model should not only discern the pedestrians from seen cameras, but also samples in unseen cameras. This idea can be achieved through camera-aware meta-learning.

t-SNE Visualization

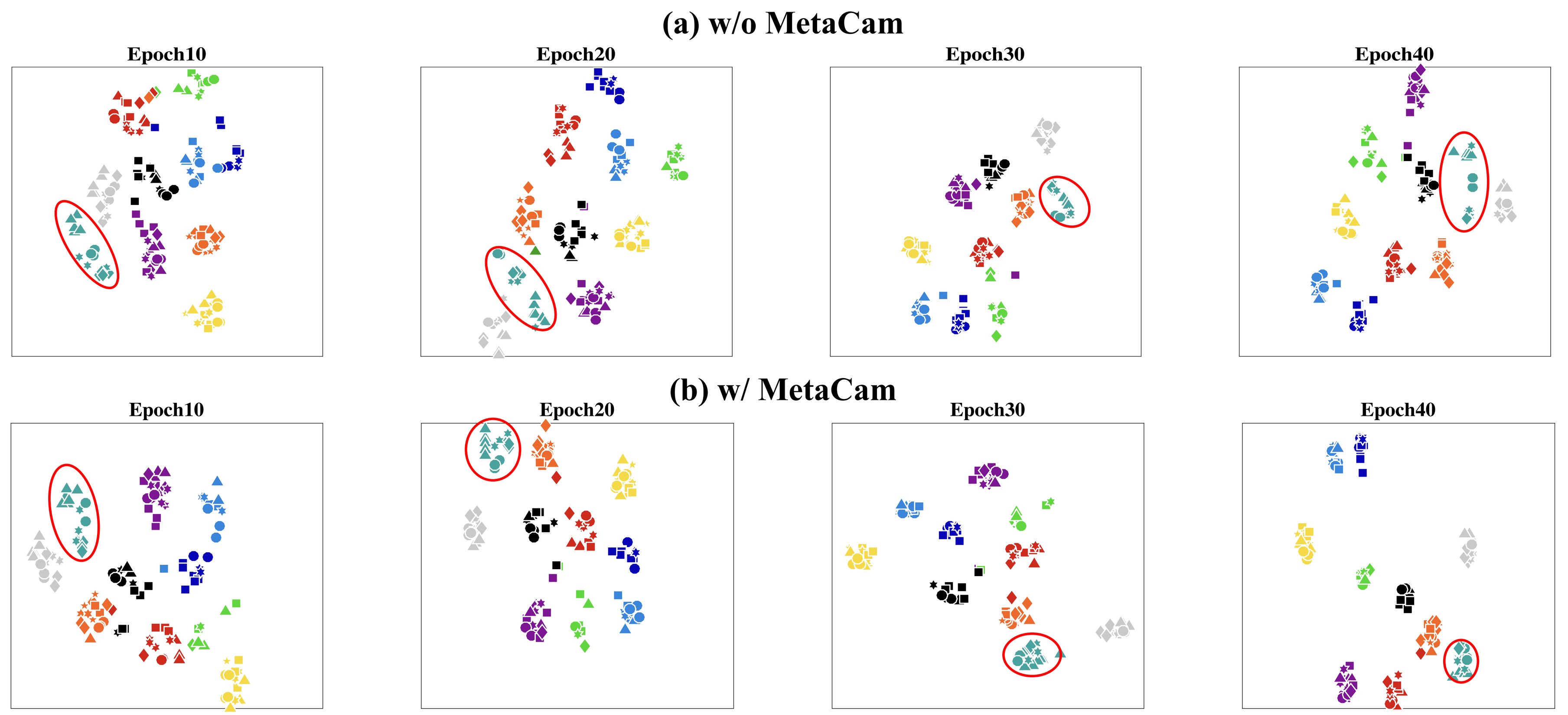


Fig 1. *t*-SNE plot of 10 persons under different settings (model trained w/o MetaCam and model trained w/ MetaCam). We use different colors to denote identities and different shapes to indicate camera IDs. The algorithm with MetaCam generates intra-class features that are close to each other, indicating that our MetaCam can guide the model to learn camera-invariant features.

Contact Us

If you have any problem, please send email to us (**yangfx@stu.xmu.edu.cn**) or ask in Github.

Scan the right QR code for code and other resources.

Experimental Results

Tab 1. Comparison with state-of-the-arts (fully unsupervised). Our method out performs current unsupervised re-ID algorithms. "*": Reproduced by [3], "†": Reproduced based on the authors' code.

Methods	Venue	DukeMTMC-reID			Market-1501			MSMT-17		
		mAP	rank-1	rank-5	mAP	rank-1	rank-5	mAP	rank-1	rank-5
OIM [38]	CVPR'17	11.3	24.5	38.8	14.0	38.0	58.0	-	-	-
BUC [19]	AAAI'19	27.5	47.4	62.6	38.3	66.2	79.6	3.4*	11.5*	18.6*
SSL [20]	CVPR'20	28.6	52.5	63.5	37.8	71.1	83.8	-	-	-
MMCL [34]	CVPR'20	40.2	65.2	75.9	45.5	80.3	89.4	-	-	-
HCT [42]	CVPR'20	50.7	69.6	83.4	56.4	80.0	91.6	-	-	-
ECN† [49]	CVPR'19	24.5	49.0	61.7	30.3	63.5	79.0	3.1	10.2	15.5
AE [3]	TOMM'20	39.0	63.2	75.4	54.0	77.5	89.8	8.5	26.6	37.0
WFDR† [41]	CVPR'20	42.4	62.0	75.1	50.1	72.1	80.5	8.6	22.3	32.5
Ours	This work	**53.8**	**73.8**	**84.2**	**61.7**	**83.9**	**92.3**	**15.5**	**35.2**	**48.3**

Tab 2. Ablation study on the proposed method. "Outliers": Including outliers into training data. "DSCE": training with DSCE loss. "MetaCam": training with MetaCam.

No.	Attributes			DukeMTMC-reID		Market-1501	
	Outliers	DSCE	MetaCam	mAP	rank-1	mAP	rank-1
1	×	×	×	6.8	16.6	6.6	17.5
2	✓	×	×	39.2	59.7	51.2	73.2
3	✓	✓	×	43.4	62.8	53.9	74.8
4	✓	×	✓	51.1	71.2	59.4	82.1
5	✓	✓	✓	**53.8**	**73.8**	**61.7**	**83.9**

Tab 3. Results on domain adaptation. M: Market-1501, D: DukeMTMC-reID. All methods use ResNet-50 as the backbone.

Methods	Venue	D → M		M → D	
		mAP	rank-1	mAP	rank-1
SPGAN [2]	CVPR'18	22.8	51.5	22.3	44.1
HHL [48]	ECCV'18	31.4	62.2	27.2	46.9
ECN [49]	CVPR'19	43.0	75.1	40.4	63.3
SSG [6]	ICCV'19	58.3	80.0	53.4	73.0
UCDA-CCE [24]	ICCV'19	34.5	64.3	36.7	55.4
MMCL [34]	CVPR'20	60.4	84.4	51.4	72.4
DG-Net++ [52]	ECCV'20	61.7	82.1	63.8	78.9
GDS [14]	ECCV'20	61.2	81.1	55.1	73.1
ECN+ [50]	TPAMI'20	63.8	84.1	54.4	74.0
MMT-500 [7]	ICLR'20	71.2	87.7	63.1	76.8
MMT-500+Ours	This Work	**76.5**	**90.1**	**65.0**	**79.5**

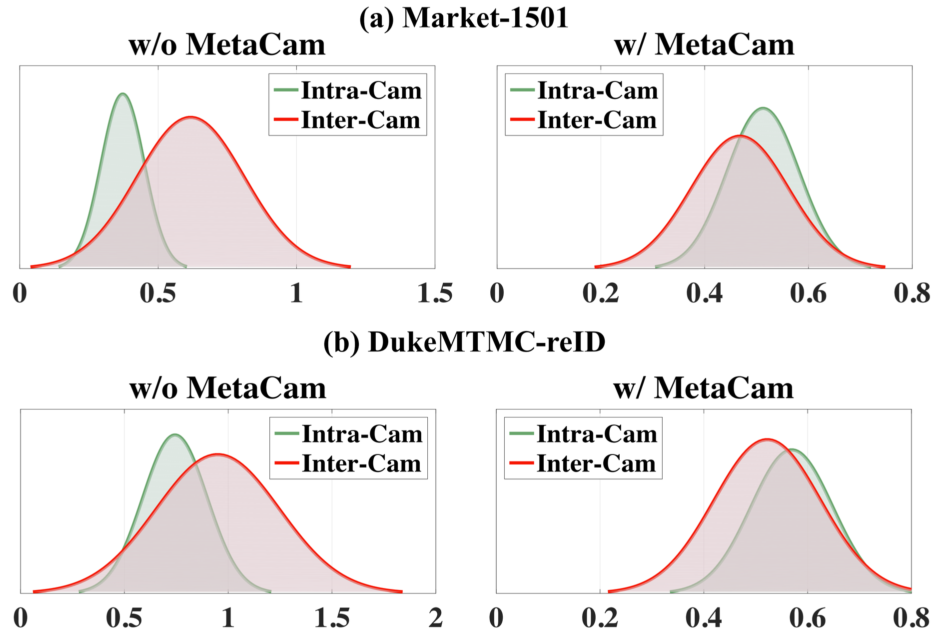


Fig 2. Distance distributions of positive pairs for intra-camera (green curve) and inter-camera (red curve).

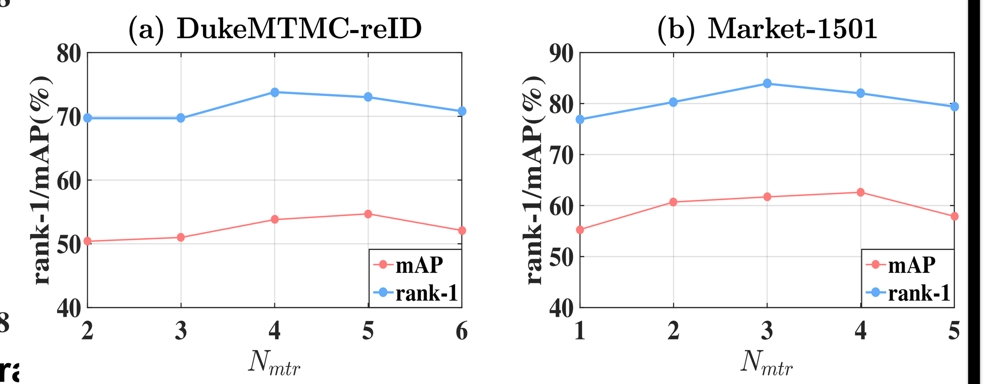


Fig 3. Sensitivity analysis of N_{mtr}.

References

[1] Lin, et al. A bottom-up clustering approach to unsupervised person re-identification. In AAAI'19.

[2] Zhong, et al. Generalizing a person retrieval model hetero-and homogeneously. In ECCV'18.

[3] Ghosh et al. Robust loss functions under label noise for deep neural networks. In AAAI'17.

[4] Yu, and Zheng. Weakly supervised discriminative feature learning with state information for person identification. In CVPR'20.