Supplementary Material
ICE: Inter-instance Contrastive Encoding for Unsupervised Person Re-identification

Appendices

A. Algorithm Details

The ICE algorithm details are provided in Algorithm 1.

Algorithm 1: Inter-instance Contrastive Encoding (ICE) for fully unsupervised ReID.

```
Input  : Unlabeled dataset X, ImageNet pre-trained
         online encoder θ_o, ImageNet pre-trained
         momentum encoder θ_m, maximal epoch E_max
         and maximal iteration I_max.
Output: Momentum encoder θ_m after training.
for epoch = 1 to E_max do
    Encode X to momentum representations M with the
      momentum encoder θ_m;
    Rerank and Generate clustering pseudo labels Y on
      momentum representations M with DBSCAN;
    Calculate cluster proxies in Eq. (4) and camera
      proxies in Eq. (6) based on Y;
    for iter = 1 to I_max do
        Calculate inter-instance similarities in a
          mini-batch;
        Train θ_o with the total loss in Eq. (3) which
          combines proxy contrastive loss in Eq. (8), hard
          instance contrastive loss in Eq. (9) and soft
          instance consistency loss in Eq. (12);
        Update θ_m by Eq. (2);
    end
end
```

B. Backbone Network

Instance-batch normalization (IBN) [4] has shown better performance than regular batch normalization in unsupervised domain adaptation [4, 2] and domain generalization [3]. We compare the performance of ICE with ResNet50 and IBN-ResNet50 backbones in Tab. 1. The performance of our proposed ICE can be further improved with an IBN-ResNet50 backbone network.

C. Threshold in clustering

In DBSCAN [1], the distance threshold is the maximum distance between two samples for one to be considered as in the neighborhood of the other. A smaller distance threshold is likely to make DBSCAN mark more hard positives as different classes. On the contrary, a larger distance threshold makes DBSCAN mark more hard negatives as same class.

In the main paper, the distance threshold for DBSCAN between same cluster neighbors is set to 0.55, which is a trade-off number for Market1501, DukeMTMC-reID and MSMT17 datasets. To get a better understanding of how ICE is sensitive to the distance threshold, we vary the threshold from 0.45 to 0.6. As shown in Tab. 2, a smaller threshold 0.5 is more appreciate for the relatively smaller dataset Market1501, while a larger threshold 0.6 is more appreciate for the relatively larger dataset MSMT17. State-of-the-art unsupervised ReID methods SpCL [2] and CAP [5] respectively used 0.6 and 0.5 as their distance threshold. Our proposed ICE can always outperform SpCL and CAP on the three datasets with a threshold between 0.5 and 0.6.

D. Camera-agnostic scenario

As mentioned in the main paper, we provide the dynamic cluster numbers of camera-agnostic ICE during the training in Fig. 1. The red curve is trained without the hard instance contrastive loss $\mathcal{L}_{h\_ins}$ as intra-class variance constraint. In this case, the soft instance consistency loss $\mathcal{L}_{s\_ins}$ maintains high intra-class variance, *e.g.*, $AA \not\approx AP_1 \not\approx AP_2 \not\approx AP_3 \not\approx 1$, which leads to less compact clusters. The orange curve is trained without $\mathcal{L}_{s\_ins}$, which has less clusters at the beginning but more clusters at last epochs than the blue curve. The blue curve is trained with both $\mathcal{L}_{h\_ins}$ and $\mathcal{L}_{s\_ins}$, whose cluster number is most accurate among the three curves at last epochs. Fig. 1 confirms that combining $\mathcal{L}_{h\_ins}$ and $\mathcal{L}_{s\_ins}$ reduces naturally captured and artificially augmented view variance at the same time, which gives optimal ReID performance.

E. Future work

Our proposed method is designed for traditional short-term person ReID, in which persons do not change their

Backbone	Market1501				DukeMTMC-reID				MSMT17			
	mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10
ResNet50	82.3	93.8	**97.6**	98.4	69.9	83.3	91.5	94.1	38.9	70.2	80.5	84.4
IBN-ResNet50	**82.5**	**94.2**	**97.6**	**98.5**	**70.7**	**83.6**	**91.9**	**93.9**	**40.6**	**70.7**	**81.0**	**84.6**

Table 1. Comparison of ResNet50 and IBN-ResNet50 backbones on Market1501, DukeMTMC-reID and MSMT17 datasets.

Threshold	Market1501				DukeMTMC-reID				MSMT17			
	mAP	R1	R5	R10	mAP	R1	R5	R10	mAP	R1	R5	R10
0.45	82.5	93.4	97.5	98.3	68.0	82.8	**91.5**	93.4	36.6	69.2	79.3	82.7
0.5	**83.0**	**94.1**	**97.7**	98.3	69.2	82.9	91.2	93.2	38.4	69.9	80.2	83.8
0.55	82.3	93.8	97.6	98.4	**69.9**	83.3	**91.5**	**94.1**	38.9	70.2	80.5	84.4
0.6	81.2	93.0	97.3	**98.5**	69.4	**83.5**	91.4	94.0	**39.4**	**70.9**	**81.0**	**84.5**

Table 2. Comparison of different distance thresholds on Market1501, DukeMTMC-reID and MSMT17 datasets.

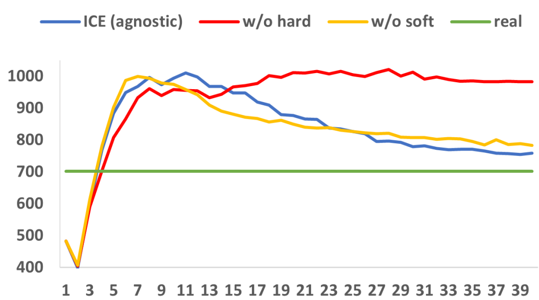

Figure 1. Dynamic cluster numbers of **ICE(agnostic)** during 40 training epochs on DukeMTMC-reID. A lower number denotes that clusters are more compact (less intra-cluster variance).

clothes. For long-term person ReID, when persons take off or change their clothes, our method is prone to generate less robust pseudo labels, which relies on visual similarity (mainly based on cloth color). For future work, an interesting direction is to consider how to generate robust pseudo labels to tackle the cloth changing problem for long-term person ReID.

References

[1] Martin Ester, Hans-Peter Kriegel, Jörg Sander, and Xiaowei Xu. A density-based algorithm for discovering clusters in large spatial databases with noise. In *KDD*, 1996. 1

[2] Yixiao Ge, Feng Zhu, Dapeng Chen, Rui Zhao, and Hongsheng Li. Self-paced contrastive learning with hybrid memory for domain adaptive object re-id. In *NeurIPS*, 2020. 1

[3] Jieru Jia, Q. Ruan, and Timothy M. Hospedales. Frustratingly easy person re-identification: Generalizing person re-id in practice. In *BMVC*, 2019. 1

[4] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *ECCV*, 2018. 1

[5] Menglin Wang, Baisheng Lai, Jianqiang Huang, Xiaojin Gong, and Xian-Sheng Hua. Camera-aware proxies for unsupervised person re-identification. In *AAAI*, 2021. 1