Supplementary Material for Dual Pseudo-Labels Interactive Self-training for Semi-Supervised Visible-Infrared Person Re-Identification

A. Overview

In this document, we first introduce our method under the bi-semi-supervised setting to make it easier to follow. Secondly, we show the labeled sample division under the bi-semi-supervised setting. Furthermore, we supplement the comparison of the proposed method with unsupervised methods. Finally, we present the visualization of the retrieval results.

A.1. The Proposed Method Under The Bi-semi-supervised Setting

Compared with the uni-semi-supervised setting, the bi-semi-supervised setting needs not only to predict pseudo-label for infrared images but also to predict pseudo-label for visible images. We split the dataset into a labeled set and an unlabeled set. In the train phase, we first train the model using the ground-truth label in the labeled set. Then we use the trained model to predict hybrid pseudo-labels for the unlabeled data. Finally, we train the model using the pseudo label by the proposed method.

It is worth noting that the NLP module is slightly different. The confidence of visible $\hat{w}^v$ and confidence of infrared $\hat{w}^r$ are both set to 1 when we train the model in the labeled set. In the unlabeled set, same as the uni-semi-supervised setting, we obtain the confidence of visible $\hat{w}^v$ and confidence of infrared $\hat{w}^r$ by GMM, respectively. The loss $\mathcal{L}_{nlp}$ is modified as follows:

$$\begin{aligned}\mathcal{L}_{nlp} = &-\frac{1}{N_r}\sum_{j=1}^{N_r}\hat{w}_j^r \log P\left(\hat{y}_j^r \mid C\left(f(r_j)\right)\right) \\ &-\frac{1}{N_v}\sum_{i=1}^{N_v}\hat{w}_i^v \log P\left(\hat{y}_i^v \mid C\left(f(v_i)\right)\right),\end{aligned} \tag{I}$$

where N_v and N_r are the number of visible images and infrared images, respectively. $f(\cdot)$ and $C(\cdot)$ are a function for extracting the features of images from different modalities and an identity classifier, respectively.

A.2. Labeled Sample Division Under The Bi-semi-supervised Setting

We describe the labeled sample division under the bi-semi-supervised setting in detail. We conduct different labeled ratios to train the model, in which the labeled ratio varies from 10%, 25% to 50% on SYSU-MM01 and RegDB, and the number of identities and samples at different labeled ratios is shown in Table I. In detail, we employ 22,258 visible images and 11,909 infrared images with 395 identities on SYSU-MM01. For each identity, we randomly sampled from the dataset in which the labeled ratio varies from 10%, 25% to 50%. For example, we sample 5,565 images in the visible modality and 2,977 images in the infrared modality as a labeled set when the labeled ratio is 25%. Similarly, we employ 2,060 visible images and 2,060 infrared images with 206 identities on RegDB, under the same setting, 515 visible images and 515 infrared images are regarded as a labeled set when the labeled ratio is 25%.

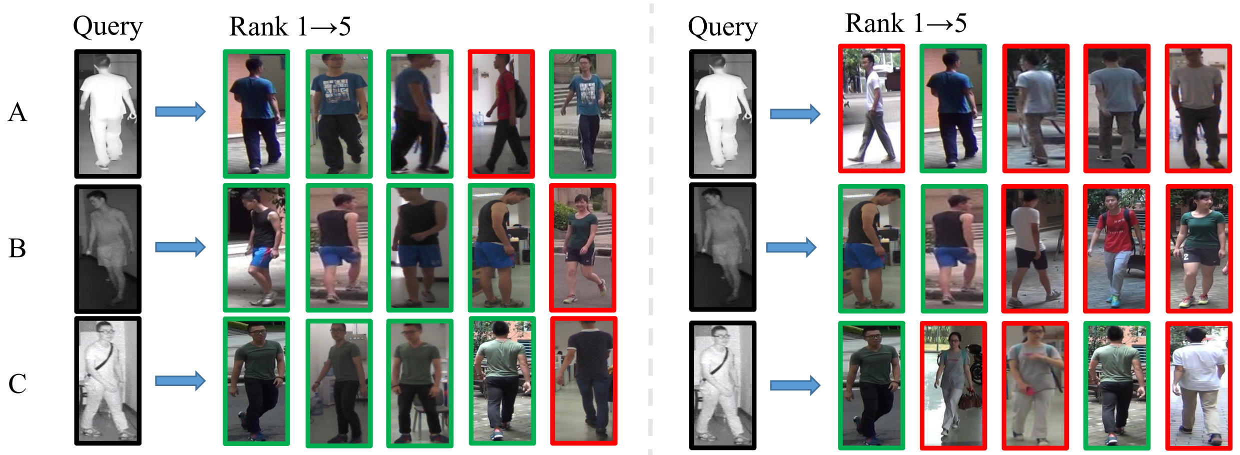
A.3. Comparison with Unsupervised Methods

We compare the proposed method in the uni-semi-supervised setting with six state-of-the-art USL methods. The quantitative results are shown in Table II. These methods only use a single pseudo-label generation approach (*e.g.*, DBSCAN), which does not take cross-modality information into account, resulting in their performances being limited. Experimental results show that our hybrid pseudo-label strategy is more conducive to semi-supervised VI-ReID.

A.4. Visualization of The Retrieval Results

We compare the top-5 retrieved results of 3 selected infrared images between the proposed method and OTLA on SYSU-MM01 (all-search mode) under the uni-semi-supervised setting, respectively. The visualization of the retrieval results is shown in Fig. I. The images in the first column are the query images and the rest of the column shows the corresponding top-5 gallery images. The retrieval results are ranked from left to right according to similarity score.

As we can see, there are a large number of false matches in OTLA. However, the proposed method achieves most of



(a) Our method (b) OTLA

Figure I. Comparison of the top-5 retrieved results of the proposed method and OTLA on the SYSU-MM01 dataset. The green bounding boxes denote the right matches and the red bounding boxes denote the false matches. (a) shows the qualitative results of OTLA. (b) shows our method. In both (a) and (b), the first column shows the query images and the rest column shows the corresponding top-5 gallery images.

Table I. The number of identities and labels with the different labeled ratios on SYSU-MM01 and RegDB under the bi-semi-supervised setting.

Labeled	SYSU-MM01				RegDB			
	Visible		Infrared		Visible		Infrared	
	ID#	Labels#	ID#	Labels#	ID#	Labels#	ID#	Labels#
10%	395	2226	395	1191	206	206	206	206
25%	395	5565	395	2977	206	515	206	515
50%	395	11129	395	5955	206	1030	206	1030

Table II. Comparisons on six advanced methods on SYSU-MM01 and RegDB under the unsupervised setting. All methods are measured by Rank-1 (%) and mAP (%). Methods marked by † denote re-implementations based on public code.

Settings			SYSU-MM01				RegDB			
			All Search		Indoor Search		Visible2Thermal		Thermal2Visible	
Type	Method	Venue	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
USL	BUC† [3]	AAAI’19	8.2	3.2	12.5	6.0	4.7	4.5	8.8	6.0
	SpCL(USL)† [1]	NIPS’20	18.7	11.4	27.1	20.9	20.6	17.3	19.0	16.6
	MetaCam† [5]	CVPR’21	14.7	9.3	23.9	17.1	23.1	17.5	20.9	16.5
	HCD† [6]	ICCV’21	18.0	17.9	24.4	28.8	10.8	12.3	12.4	13.7
	H2H [2]	TIP’21	25.5	25.2	-	-	14.1	12.3	13.9	12.7
	OTLA [4]	ECCV’22	29.9	27.1	29.8	38.8	32.9	29.7	32.1	28.6
USSL	**DPIS(ours)**	-	**58.4**	**55.6**	**63.0**	**70.0**	**62.3**	**53.2**	**61.5**	**52.7**

the right matches even though some cases are difficult for humans to recognize (*e.g.*, query A and B). This demonstrates the effectiveness of the proposed method in improving feature discriminability. In addition, we discover that in some cases (*e.g.*, query C), even if people change their belongings, the proposed method is still capable of achieving accurate image retrieval.

References

[1] 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. 2

[2] Wenqi Liang, Guangcong Wang, Jianhuang Lai, and Xiaohua Xie. Homogeneous-to-heterogeneous: Unsupervised learning for rgb-infrared person re-identification. *IEEE Trans. Image Process.*, pages 6392–6407, 2021. 2

[3] Yutian Lin, Xuanyi Dong, Liang Zheng, Yan Yan, and Yi Yang. A bottom-up clustering approach to unsupervised person re-identification. In *AAAI*, pages 8738–8745, 2019. 2

[4] Jiangming Wang, Zhizhong Zhang, Mingang Chen, Yi Zhang, Cong Wang, Bin Sheng, Yanyun Qu, and Yuan Xie. Optimal transport for label-efficient visible-infrared person reidentification. In *ECCV*, pages 93–109, 2022. 2

[5] Fengxiang Yang, Zhun Zhong, Zhiming Luo, Yuanzheng Cai, Yaojin Lin, Shaozi Li, and Nicu Sebe. Joint noise-tolerant learning and meta camera shift adaptation for unsupervised person re-identification. In *CVPR*, pages 4855–4864, 2021. 2

[6] Yi Zheng, Shixiang Tang, Guolong Teng, Yixiao Ge, Kaijian Liu, Jing Qin, Donglian Qi, and Dapeng Chen. Online pseudo label generation by hierarchical cluster dynamics for adaptive person re-identification. In *ICCV*, pages 8351–8361, 2021. 2