Supplementary Material for "Weakly Supervised Visible-Infrared Person Re-Identification via Heterogeneous Expert Collaborative Consistency Learning"

Yafei Zhang[1], Lingqi Kong[1], Huafeng Li[1]*, Jie Wen[2]
[1]Faculty of Information Engineering and Automation, Kunming University of Science and Technology
[2]School of Computer Science and Technology, Harbin Institute of Technology, Shenzhen
zyfeimail@163.com, lingqikong2333@gmail.com, hfchina99@163.com, jiewen_pr@126.com

1. Algorithm

To provide a clear understanding of the model, the complete training procedure of the proposed method is presented in Algorithm 1.

Algorithm 1 Training process of the proposed method

Input: dataset $\mathcal{X}$, individual labels $\mathcal{Y}$
Output: updated $\boldsymbol{E}^r$, $\boldsymbol{E}^v$ and shared classifier $\boldsymbol{W}^c$
Initialize: Initialize $\boldsymbol{E}^r$, $\boldsymbol{E}^v$, $\boldsymbol{W}^v$, $\boldsymbol{W}^r$, and $\boldsymbol{W}^c$.
Phase I: Heterogeneous Expert Construction
for $iter = 1$ to T_1 **do**
    Optimize $\boldsymbol{E}^r$, $\boldsymbol{E}^v$, $\boldsymbol{W}^v$, and $\boldsymbol{W}^r$ by minimizing the loss in Eq.(14).
end for
Phase II: Collaborative Consistency Learning
for $iter = 1$ to T_2 **do**
    Building $\mathcal{P}^v$ and $\mathcal{P}^r$.
    Predict cross-modal identity by using $\boldsymbol{W}^v$ and $\boldsymbol{W}^r$.
    Generate $\boldsymbol{M}_c$, $\boldsymbol{M}_s$, $\boldsymbol{M}_w$ by Eq.(4), Eq.(5), Eq.(6).
    Optimize $\boldsymbol{E}^r$, $\boldsymbol{E}^v$ and $\boldsymbol{W}^c$ by minimizing $\mathcal{L}_{id}^{stro}$, $\mathcal{L}_{id}^{weak}$, $\mathcal{L}_{wrt}^{cros}$.
    Optimize $\boldsymbol{W}^v$, $\boldsymbol{W}^r$ by minimizing $\mathcal{L}_{id}^{exp}$, $\mathcal{L}_{homo}$.
    Update $\mathcal{P}^v$ and $\mathcal{P}^r$ via Eq.(10).
end for

In addition, Figure 1 compares the recall rate of cross-modal pseudo-labels before and after incorporating CLAE, showing a significant improvement in pseudo-label quality.

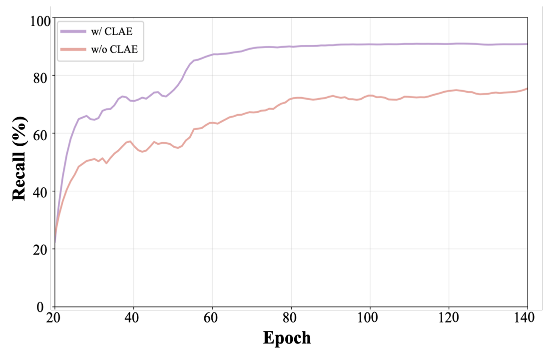

Figure 1. Comparison of cross-modal pseudo-label quality on the SYSU-MM01 dataset before and after adding CLAE in a weakly supervised setting, with the x-axis representing training epochs and the y-axis representing recall rate.

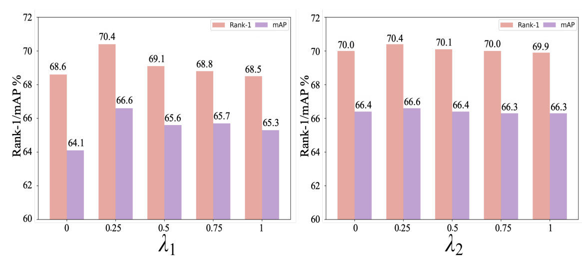

Figure 2. Effect of different values of λ_1 and λ_2 on model performance in the all search mode of the SYSU-MM01 dataset.

2. Selection of Hyperparameters

In the loss function of this paper, two hyperparameters, λ_1 and λ_2, are involved. To achieve better model performance, we conducted a search for the optimal values of λ_1 and λ_2 within the interval [0, 1]. As shown in Figure 2, when λ_1 and λ_2 are increased from 0 to 0.25, the model performance reaches its optimum; however, when they exceed 0.25, the model performance starts to decline. Therefore, in this paper, we set both λ_1 and λ_2 to 0.25.

3. Analysis of Parameter Settings in Eq. (12)

In the total collaborative consistency loss

$$\mathcal{L}_{homo} = w_v \mathcal{L}_{homo}^{v} + w_r \mathcal{L}_{homo}^{r}, \tag{1}$$

we use

$$w_v = \frac{H(\boldsymbol{p}^{r\to v})}{H(\boldsymbol{p}^{r\to v}) + H(\boldsymbol{p}^{v\to r})}, \quad w_r = \frac{H(\boldsymbol{p}^{v\to r})}{H(\boldsymbol{p}^{r\to v}) + H(\boldsymbol{p}^{v\to r})} \tag{2}$$

*Corresponding author: Huafeng Li (hfchina99@163.com).

to replace manually chosen hyperparameters. To verify its effectiveness, we conduct an experimental analysis, as shown in Table 1. Table 1 summarizes the following cases:

- Case 0 corresponds to the scenario where CLAE is not used.
- Case 1 sets both w_v and w_r to 0.5.
- Case 2 exchanges the roles of w_v and w_r in Eq. (1).
- Case 3 applies the Softmax operation to w_v and w_r following Eq. (2).
- Case 4 determines w_v and w_r based on Eq. (2).

Table 1. Influence of different weight settings in $\mathcal{L}_{homo}$.

Cases	SYSU-MM01			
	Rank-1	Rank-10	Rank-20	mAP
0	68.0	95.3	98.4	64.6
1	69.0	95.4	98.4	65.4
2	68.7	95.4	98.6	64.9
3	69.8	95.1	98.3	65.8
4	**70.4**	**95.8**	**98.8**	**66.6**

The results listed in Table 1 show that the model's recognition performance is relatively low in the absence of CLAE (Case 0), highlighting its importance. In Case 1, although assigning equal weights to all experts led to some performance improvement, the gain was limited as it failed to dynamically adjust weights based on input results. In Case 2, performance was only marginally better than Case 0 due to excessive constraints on low-credibility classifiers. While Case 3 employed the Softmax operation to mitigate the impact of credibility on weights, it still yielded suboptimal performance. In contrast, Case 4 achieved the best performance by adopting an adaptive adjustment strategy. These results clearly validate the effectiveness of our adaptive weight selection strategy in optimizing strength adjustment.

4. Visualization of Retrieval Results

To further demonstrate the effectiveness of our method, we present retrieval results on the SYSU-MM01[3] dataset in Figure 3. In each retrieval result, green boxes denote correctly retrieved images matching the query sample, whereas red boxes indicate incorrect matches. These results clearly demonstrate that our method effectively enhances the ranking of correct matches and increases their presence in the top ranks.

5. Experiments on RegDB

We also evaluate our proposed method on the RegDB [1] dataset, a small-scale visible-infrared person re-identification benchmark captured using a dual-camera system (visible and infrared). It contains 8,240 images from 412 identities, each with 10 pairs of visible and infrared images. As shown in Table 2, our method achieves excellent performance under both the visible-to-infrared and infrared-to-visible evaluation settings.

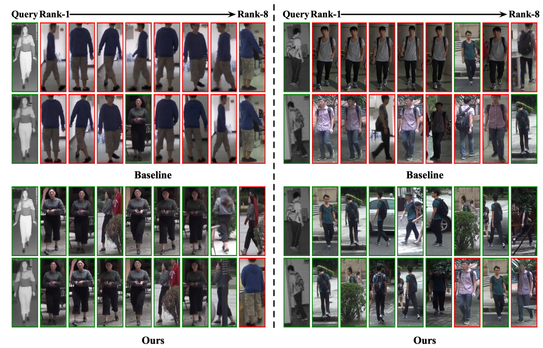

Figure 3. Visualization of retrieval results on the SYSU-MM01 dataset: baseline vs. our Method

Table 2. Comparison of Rank-1 (%) and mAP (%) performances with the state-of-the-art methods on **RegDB** dataset.

Methods	VIS to IR		IR to VIS	
	Rank-1	mAP	Rank-1	mAP
DPIS [2]	62.3	53.2	61.5	52.7
AGW [6]	70.1	66.4	70.5	65.9
GUR [4]	73.9	70.2	75.0	69.9
CAJ [5]	85.0	79.1	84.8	77.8
MUCG [7]	86.9	76.7	83.7	74.1
Ours	**86.9**	**80.1**	**86.4**	**80.6**

6. Limitations of the Method and Future Work

Despite the effectiveness of the proposed weakly supervised cross-modal person ReID method, several limitations remain. The reliance on modality-specific classification experts introduces a potential sensitivity to the quality of single-modal labels. In scenarios where single-modal identity annotations are noisy or inconsistent, the accuracy of cross-modal identity correspondence establishment may be affected. Future work could explore robust learning strategies to mitigate the impact of label noise, such as self-correction mechanisms or label refinement techniques.

References

[1] Dat Tien Nguyen, Hyung Gil Hong, Ki Wan Kim, and Kang Ryoung Park. Person recognition system based on a combination of body images from visible light and thermal cameras. *Sensors*, 17(3), 2017. 2

[2] Jiangming Shi, Yachao Zhang, Xiangbo Yin, Yuan Xie, Zhizhong Zhang, Jianping Fan, Zhongchao Shi, and Yanyun Qu. Dual pseudo-labels interactive self-training for semi-supervised visible-infrared person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11218–11228, 2023. 2

[3] Ancong Wu, Wei-Shi Zheng, Hong-Xing Yu, Shaogang Gong, and Jianhuang Lai. Rgb-infrared cross-modality person re-identification. In *Proceedings of the IEEE international Conference on Computer Vision (ICCV)*, pages 5380–5389, 2017. 2

[4] Bin Yang, Jun Chen, and Mang Ye. Towards grand unified representation learning for unsupervised visible-infrared person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 11069–11079, 2023. 2

[5] Mang Ye, Weijian Ruan, Bo Du, and Mike Zheng Shou. Channel augmented joint learning for visible-infrared recognition. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 13547–13556, 2021. 2

[6] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven C. H. Hoi. Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(6):2872–2893, 2022. 2

[7] Xiying Zheng, Yukang Zhang, Yang Lu, and Hanzi Wang. Semi-supervised visible-infrared person re-identification via modality unification and confidence guidance. In *Proceedings of the 32nd ACM International Conference on Multimedia (ACMMM)*, pages 5761–5770, 2024. 2