Towards Cross-Modal Preservation, Consistency and Alignment for Privacy-Preserving Visible-Infrared Person Re-Identification

Supplementary Material

Table 5. **Ablation study on SYSU-MM01**. "KPR" refers to Keypoint-Preserving Regularization. "DCMA" means Differential Consistency-guided Modality Alignment. Rank1 accuracy(%), mAP (%) and mINP (%) are reported.

Components			Query		Gallery		All Search			Indoor Search		
Baseline	KPR	DCMA	Raw	Anon	Raw	Anon	Rank1	mAP	mINP	Rank1	mAP	mINP
✓			✓		✓		58.0	56.4	42.7	66.6	72.8	69.3
			✓			✓	48.6	44.7	27.7	59.2	67.4	64.1
				✓	✓		47.9	47.3	33.5	59.1	66.7	63.6
				✓		✓	43.6	40.7	24.7	50.3	60.1	56.2
✓	✓		✓		✓		59.4	57.4	43.3	66.6	73.4	70.1
			✓			✓	50.8	45.1	27.5	58.3	66.4	62.4
				✓	✓		51.6	51.3	37.6	60.6	68.3	64.8
				✓		✓	46.1	42.8	26.5	51.2	60.9	57.1
✓		✓	✓		✓		58.0	55.9	40.9	68.0	73.1	69.4
			✓			✓	50.9	47.0	31.1	61.8	67.9	63.3
				✓	✓		49.9	48.9	35.0	60.5	68.0	64.6
				✓		✓	44.7	42.2	27.4	55.0	62.9	58.7
✓	✓	✓	✓		✓		**60.7**	**58.8**	**45.0**	**70.3**	**75.6**	**72.2**
			✓			✓	50.9	46.4	29.4	61.9	68.7	64.4
				✓	✓		53.3	51.9	37.9	62.9	70.6	67.6
				✓		✓	46.2	43.4	27.8	56.6	65.1	61.0

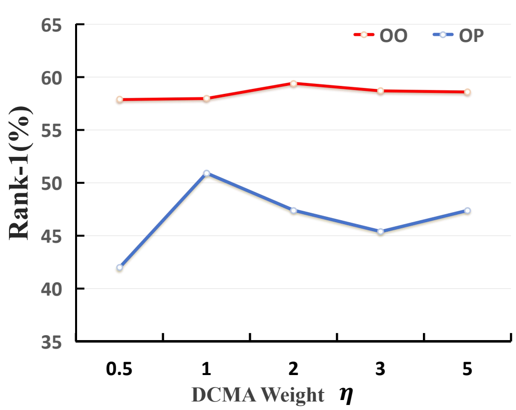

Figure 6. The impact of DCMA weight value on SYSU-MM01 dataset. OO refers to raw query/gallery, OP refers to raw query, anonymized gallery.

6. Additional Parameter Experiments

We conduct additional parameter analysis on the DCMA loss weight η in our framework. All experiments are performed on the SYSU-MM01 dataset with mosaic initialization.

To analyze the DCMA module, we set the KPR loss weight δ to 0 and varied the DCMA loss weight η. As shown in Figure 6, the model's performance peaks when η is set to 1.0. This demonstrates the effectiveness of DCMA when acting as the primary regularizer.

When both modules are combined, we found through further tuning that the optimal balance for our full PPA model was achieved with $\delta = 1.0$ and $\eta = 0.1$ as stated in Section 4.2. We found KPR ($\delta = 1.0$) is crucial for preserving structural integrity. With this strong structural prior established, only a small DCMA weight ($\eta = 0.1$) is needed to align the differential consistency.

7. Complete Ablation Experiments

To comprehensively evaluate the effectiveness of our proposed KPR and DCMA modules, this section details a complete ablation study. While the main paper table 4 presented results exclusively for the Anon-to-Raw setting, here we provide the full results across all four testing scenarios: Raw-to-Raw, Raw-to-Anon, Anon-to-Raw, and Anon-to-Anon. All experiments were conducted on the SYSU-MM01 dataset with mosaic initialization. The complete results in table 5 below clearly demonstrate the contribution of each component in every setting.

8. Visualization of Pose Estimation on Infrared Images

To address potential concerns regarding the domain gap between the RGB-trained YOLOv8n-pose model and infrared (IR) images, we provide qualitative visualizations of the pose estimation results on the IR modality.

As shown in Figure 7, the pre-trained model accurately captures human body structures even in challenging low-light IR scenarios. This visually confirms that thermal contours provide distinct and sufficient structural cues for accurate pose estimation, without requiring any IR-specific retraining or fine-tuning.

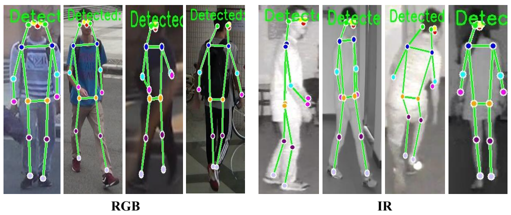

Figure 7. Qualitative visualization of pose estimation on infrared (IR) images. The RGB-trained pose estimator successfully captures accurate body structures despite the low-light conditions and modality differences.

9. Disentangling the Mixed Gap

To explicitly quantify the individual impact of the modality gap and the anonymization gap, as well as their combined effect, we conducted a controlled experiment. Table 6 presents the performance degradation across different query and gallery settings.

Table 6. Impact analysis of individual and mixed gaps on retrieval performance (Rank-1 accuracy). “Raw” refers to original images, “Anon” refers to anonymized images, “Vis” is the visible (RGB) modality, and “IR” is the infrared modality.

Setting (Query / Gallery)	Gap Type	Rank-1	Drop
Raw-Vis / Raw-Vis	None (Intra-modal)	84.5%	-
Raw-IR / Raw-Vis	Modality Only	76.1%	-8.4%
Raw-Vis / Anon-Vis	Anonymization Only	68.5%	-16.0%
Raw-IR / Anon-Vis	**Mixed Gap**	**61.3%**	**-23.2%**

As shown in Table 6, introducing solely the modality gap or the anonymization gap results in noticeable performance drops of 8.4% and 16.0%, respectively. However, when both challenges are present simultaneously (Raw-IR / Anon-Vis), we observe a severe compound drop of 23.2%. This significant degradation confirms that the Mixed Gap is not merely a simple linear addition of the two individual gaps, but rather a complex, compounded challenge that necessitates our proposed alignment framework.