
Supplementary Appendix

Learning Source-free Domain Adaptation for Visible-Infrared Person Re-Identification

Anonymous Author(s)

Affiliation

Address

email

1 In this supplementary appendix, we provide extended discussions and empirical analyses that com-
2 plement the main paper. Appendix A details the characteristics and construction of the experimental
3 datasets, which form the foundation for our evaluation. Appendix B presents comprehensive data-
4 level visualizations, revealing the domain discrepancies under both the *Basic Setting* and the *Weather*
5 *Setting*, and highlighting the challenges inherent to source-free VI-ReID. Appendix C elaborates on
6 the training procedure, offering clarity on the practical implementation of our framework. Appendix D
7 includes additional experiments that further validate our findings and shed light on specific design
8 choices. Lastly, Appendix E reflects on the limitations of our approach and considers its broader
9 impact in real-world applications.

10 A Details about Experimental Datasets

11 Our research is conducted on three widely adopted VI-ReID datasets: SYSU-MM01 [1], RegDB [2],
12 and LLCM [3]. These datasets span a range of application scenarios, environmental conditions, and
13 modality characteristics, providing a comprehensive foundation for evaluating the generalizability of
14 cross-modal person re-identification methods. Detailed dataset statistics are provided in Table 1.

- 15 • **SYSU-MM01 [1]:** This is a large-scale benchmark for VI-ReID. It is collected using four
16 visible-light and two near-infrared cameras, capturing images across both indoor and outdoor
17 environments. The training set comprises 22,257 visible and 11,909 infrared images from
18 395 identities. For single-shot evaluation, 3,803 infrared images from 96 identities are used
19 as queries, and the gallery consists of 301 visible images, randomly selected to represent
20 the same set of identities. The dataset introduces moderate challenges such as cross-camera
21 variations and background clutter, making it suitable for evaluating baseline performance.
- 22 • **RegDB [2]:** RegDB is a relatively smaller dataset that contains image pairs captured by
23 two well-aligned cameras: one visible and one thermal. Each identity is associated with 10
24 visible and 10 infrared images, captured under similar viewpoints and distances. Owing to
25 the high alignment and pose consistency between modalities, RegDB presents fewer cross-
26 modal variations, making it a suitable testbed for verifying the basic matching capability of
27 VI-ReID methods under controlled conditions.
- 28 • **LLCM [3]:** LLCM is a recently released and more challenging VI-ReID dataset designed
29 to reflect real-world scenarios under adverse lighting. It is collected using nine cameras
30 in low-light conditions, resulting in 46,767 annotated bounding boxes from 1,064 iden-
31 tities. In contrast to SYSU-MM01 and RegDB, LLCM contains severe low-light noise,
32 including motion blur, extreme illumination variations, and substantial pose differences, all
33 of which exacerbate the domain gap across modalities. This dataset serves as a rigorous
34 benchmark for assessing the robustness and adaptability of cross-modal models in complex
35 and unconstrained environments.

Table 1: Detailed statistics of SYSU-MM01, RegDB, and LLCM datasets across training, query, and gallery subsets for both Visible (RGB) and Infrared (IR) modalities.

Dataset	Train		Query		Gallery		Cameras	
	IDs	Images (RGB / IR)	IDs	Images (IR / RGB)	IDs	Images (RGB / IR)	RGB	IR
SYSU-MM01	395	11,659 / 7,486	96	384 / 384	96	2,972 / 3,419	4	2
RegDB	206	2,060 / 2,060	206	206 / 206	206	2,060 / 2,060	1	1
LLCM	572	14,726 / 16,835	384	3,394 / 3,394	880	35,309 / 36,938	6	6

B Data Visualization for Domain Discrepancy

B.1 Domain Adaptation Settings

To comprehensively evaluate the effectiveness and robustness of our method in SFDA scenario, where no source domain data is accessible during adaptation, we define two settings that reflect realistic deployment challenges under diverse domain discrepancies. These settings are carefully designed to capture both structural dataset-level shifts and environmental perturbations, as detailed in Table 2.

Table 2: Domain adaptation settings with emphasis on domain discrepancy sources.

Setting	Scenario	Source Domain	Target Domain
Basic Setting	SYSU-MM01→RegDB	SYSU-MM01	RegDB
	SYSU-MM01→LLCM	SYSU-MM01	LLCM
	LLCM→RegDB	LLCM	RegDB
	LLCM→SYSU-MM01	LLCM	SYSU-MM01
Weather Setting	SYSU-MM01→RegDB-W	SYSU-MM01	RegDB-W (☀ Sunny)
			RegDB-W (☁ Fog)
			RegDB-W (❄ Frost)
			RegDB-W (🌧 Rain)
			RegDB-W (❄ Snow)
	SYSU-MM01→LLCM-W	SYSU-MM01	LLCM-W (☀ Sunny)
			LLCM-W (☁ Fog)
			LLCM-W (❄ Frost)
			LLCM-W (🌧 Rain)
			LLCM-W (❄ Snow)

1) Basic Setting: This setting evaluates the model’s ability to generalize across datasets exhibiting common domain shifts, such as variations in illumination, background clutter, pedestrian density, and camera viewpoints. As shown in Table 2, we consider four cross-dataset adaptation scenarios: (i) SYSU-MM01→RegDB, (ii) SYSU-MM01→LLCM, (iii) LLCM→RegDB, and (iv) LLCM→SYSU-MM01.

In line with the SFDA setting, the model is pre-trained on a labeled source dataset and then adapted to the unlabeled target domain without accessing the source data. Among these datasets, SYSU-MM01 and LLCM are selected as source domains due to their larger scale and more diverse visual conditions, while RegDB, LLCM, and SYSU-MM01 serve as target domains depending on the scenario. RegDB is not used as a source due to its relatively limited identity scale and simpler cross-modal settings. This protocol captures typical inter-domain discrepancies caused by dataset-specific sensor characteristics, scene layouts, and modality distributions.

2) Weather Setting: To further assess the model’s robustness under environmental domain shifts in the SFDA setting, we introduce synthetic weather corruptions to simulate real-world conditions. Specifically, we augment both RegDB and LLCM with five common weather types: ☀ Sunny, ☁ Fog, ❄ Frost, 🌧 Rain, and ❄ Snow, each simulated at three severity levels. The resulting corrupted datasets are denoted as RegDB-W and LLCM-W.

As summarized in Table 2, the model is pre-trained on SYSU-MM01 and adapted to each weather-corrupted target domain without access to source data. This scenario introduces more severe domain discrepancies, combining cross-dataset distribution gaps with intra-dataset weather-induced distor-



Figure 1: Data-level visualization of various visual domain shift types in *Basic Setting*.

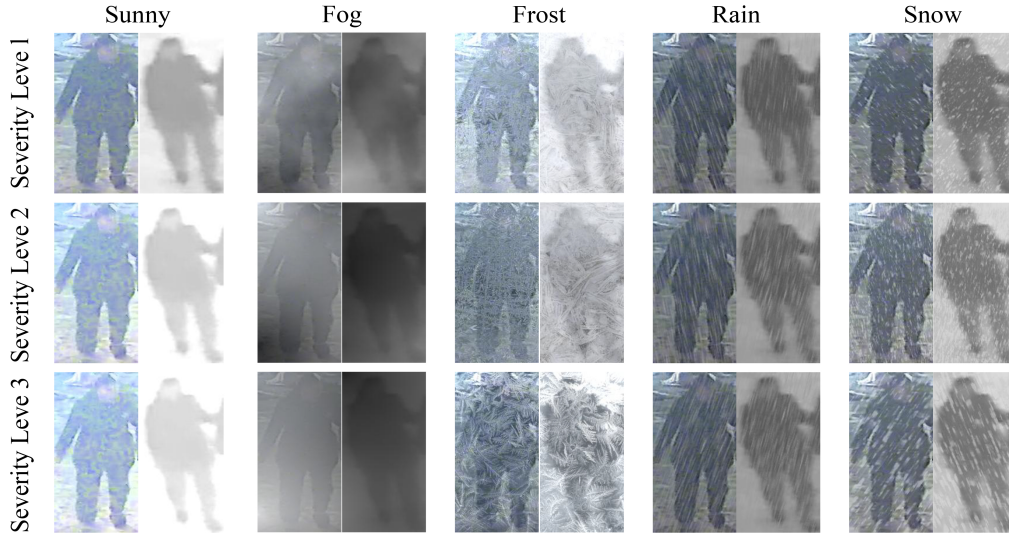


Figure 2: Data-level visualization of various visual domain shift types in *Weather Setting*.

62 tions. It reflects practical SFDA challenges where adverse environmental conditions significantly
63 impact both image quality and modality consistency during deployment.

64 B.2 Data-level Visualization of *Basic Setting*

65 To provide an intuitive understanding of the domain discrepancies encountered in SFDA, we present
66 data-level visualizations for the *Basic Setting* in Figure 1, where SYSU-MM01 is used as the source
67 domain and RegDB and LLCM serve as target domains. These visualizations illustrate typical cross-
68 domain differences in illumination, background context, and modality characteristics. Specifically,
69 SYSU-MM01 includes a mix of indoor and outdoor scenes under controlled lighting conditions,
70 along with rich visible-infrared associations, making it a suitable source dataset. In contrast, RegDB
71 provides aligned visible-thermal image pairs with relatively uniform poses, but its thermal infrared
72 images differ significantly from the near-infrared counterparts of SYSU-MM01. LLCM poses further
73 adaptation challenges due to its low-light conditions and more severe image degradation.

74 B.3 Data-level Visualization of Weather Setting

75 To simulate more realistic and dynamic environmental shifts under the SFDA paradigm, we introduce
 76 five synthetic weather perturbations: Sunny, Fog, Frost, Rain, and Snow to the RegDB and LLCM
 77 datasets. Each weather condition is applied at three severity levels, forming weather-augmented
 78 target datasets, denoted as RegDB-W and LLCM-W. This design facilitates a systematic evaluation
 79 of model robustness under varying environmental conditions. By introducing progressively more
 80 severe perturbations, we aim to assess the model’s resilience to both cross-dataset distribution gaps
 81 and intra-dataset visual distortions. Representative visualizations of these weather-induced domain
 82 shifts are shown in Figure 2, emphasizing the added complexity in this setting.

83 C Training Procedure

84 Algorithm 1 provides a comprehensive overview of the optimization process for our SVIP learning
 85 paradigm.

Algorithm 1 Training Procedure of SVIP.

Input: Target domain dataset D_T , source model Φ_S , hyper-parameters $\eta, \tau, T, \lambda_1, \lambda_2$, batch size N_b , total epochs N_e .
 1: Initialize the target model Φ_T using random initialization;
 2: Initialize source and target memory banks M_S, M_T ;
 3: **for** epoch $n_e = 1$ to N_e **do**
 4: Extract features $\Phi_S(\mathbf{x}_i^P)$ and $\Phi_T(\mathbf{x}_i^P)$ for all $\mathbf{x}_i^P \in D_T$;
 5: Cluster features via DBSCAN to get centers C_S, C_T (*i.e.*, Equation (1));
 6: Update memory banks M_S, M_T via EMA (*i.e.*, Equation (2-3));
 7: **while** not all batches processed **do**
 8: Sample batch $\{\mathbf{x}_i^P\}_{i=1}^{N_b}$ from D_T ;
 9: Compute similarity distributions $\mathbf{p}_S, \mathbf{p}_T$ (*i.e.*, Equation (4-5));
 10: Calculate confidence weights γ via entropy (*i.e.*, Equation (6));
 11: Compute SGCL loss \mathcal{L}_{dis} (*i.e.*, Equation (7));
 12: Compute RTL loss \mathcal{L}_{res} (*i.e.*, Equation (8));
 13: Compute cross-modal similarities S_{ij} (*i.e.*, Equation (9));
 14: Filter reliable pairs via C_1, C_2 (*i.e.*, Equation (10-11));
 15: Compute SCCA loss \mathcal{L}_{cre} (*i.e.*, Equation (12));
 16: Update Φ_T via total loss $\mathcal{L}_{all} = \mathcal{L}_{dis} + \lambda_1 \mathcal{L}_{res} + \lambda_2 \mathcal{L}_{cre}$ (*i.e.*, Equation (13));
 17: **end while**
 18: **end for**
Output: Target model Φ_T .

86 D More Experimental Results

87 Due to space limitations in the main text, we present additional experimental results in this section.
 88 To ensure a fair and comprehensive evaluation, we include both UVI-ReID and SFDA methods as
 89 baselines. The UVI-ReID method is implemented strictly in accordance with its original design. In
 90 contrast, most existing SFDA approaches are developed for unimodal scenarios and do not account for
 91 cross-modal heterogeneity. Following the strategy in [4], we incorporate the Hungarian algorithm and
 92 contrastive learning into these methods to enable effective cross-modal association in the VI-ReID.

93 D.1 More Comparisons on the Weather Setting

94 To further assess the adaptability of our method under complex environmental perturbations, we
 95 conduct detailed comparisons across Severity Levels 1 and 2, as illustrated in Table 3 and Table 4.
 96 The results demonstrate clear performance degradation trends as the severity of weather-induced
 97 corruption increases, providing a more comprehensive perspective on the sensitivity of different
 98 methods to varying degrees of domain shift. Despite the increasing difficulty, our method consistently
 99 achieves superior performance across diverse weather conditions, indicating its robustness and

strong generalization ability. Notably, the proposed framework exhibits effective adaptation to target domains affected by substantial visual degradation, even in the absence of source domain data. This capability is particularly important in source-free domain adaptation scenarios, where models must rely solely on unlabeled target data to overcome both inter-domain and intra-domain discrepancies.

Table 3: Comparisons with state-of-the-art methods under the *Weather Setting* (Severity Level 2), reporting R-1 (%) accuracy.

	Methods	LLCM-W (Visible to Infrared)						RegDB-W (Visible to Thermal)					
		Sunny	Fog	Frost	Rain	Snow	Avg.	Sunny	Fog	Frost	Rain	Snow	Avg.
	Source Only	12.00	18.16	9.72	12.21	10.70	12.56	15.33	23.34	12.51	14.55	13.60	15.87
UVI-ReID	OTAL [5]	14.29	18.44	12.92	11.42	15.16	14.45	32.67	36.14	27.80	36.02	21.23	23.40
	CCLNet [6]	27.09	31.98	25.66	31.77	23.29	27.96	47.01	54.65	44.15	51.59	30.82	23.40
	PGM [4]	26.51	32.49	28.58	26.69	24.21	27.70	56.63	61.81	48.71	63.36	38.32	53.76
	GUR [7]	27.41	31.32	26.58	30.30	23.76	27.87	49.23	64.28	49.77	59.68	37.22	52.04
	SDCL [8]	29.92	34.73	31.13	31.82	25.10	30.54	58.01	70.16	54.78	70.31	44.41	59.53
	MMM [9]	30.70	36.21	28.76	30.05	28.59	30.86	54.09	74.20	55.06	70.78	46.53	60.13
SFDA	CMT [10]	23.48	33.84	24.26	30.31	24.25	27.23	55.40	67.07	49.50	63.00	37.67	54.53
	LEAD [11]	24.37	32.16	23.62	21.90	22.93	25.00	54.04	73.50	49.74	70.21	40.22	57.54
	DRU [12]	24.47	33.37	27.47	30.74	22.43	27.69	53.70	71.30	48.69	69.37	46.45	57.90
	SVIP (Ours)	31.12	40.65	35.52	33.91	28.83	34.01	58.11	75.81	55.10	73.88	50.73	62.73

Table 4: Comparisons with state-of-the-art methods under the *Weather Setting* (Severity Level 1), reporting R-1 (%) accuracy.

	Methods	LLCM-W (Visible to Infrared)						RegDB-W (Visible to Thermal)					
		Sunny	Fog	Frost	Rain	Snow	Avg.	Sunny	Fog	Frost	Rain	Snow	Avg.
	Source Only	15.85	20.20	9.91	13.10	11.16	14.05	20.71	26.96	12.90	14.77	13.51	17.77
UVI-ReID	OTAL [5]	15.21	20.26	14.11	12.15	16.13	15.57	34.57	37.64	30.21	37.95	21.70	32.41
	CCLNet [6]	28.82	37.27	27.18	33.80	24.77	30.37	51.12	56.40	44.22	54.01	33.63	47.88
	PGM [4]	29.81	32.83	31.16	30.43	22.20	29.28	60.25	64.30	53.43	65.18	39.38	56.51
	GUR [7]	29.76	35.15	28.98	31.30	23.25	29.69	51.74	68.67	51.68	65.34	40.44	55.57
	SDCL [8]	35.29	32.73	33.15	34.13	25.18	32.10	60.67	74.23	57.02	78.85	46.62	63.48
	MMM [9]	33.60	36.74	30.95	35.95	28.42	33.13	59.16	77.33	57.94	77.03	48.47	63.99
SFDA	CMT [10]	27.26	35.71	26.01	32.14	23.92	29.01	59.35	72.47	51.96	74.95	39.54	59.66
	LEAD [11]	30.92	37.09	25.72	26.49	26.39	29.32	55.47	75.48	51.06	73.58	40.80	59.28
	DRU [12]	36.30	36.62	29.39	32.96	27.42	32.54	55.08	78.06	54.14	73.66	48.36	61.86
	SVIP (Ours)	37.66	42.04	38.58	37.12	31.41	37.36	61.00	78.59	58.40	80.84	54.02	66.57

103

104 D.2 Influence of the Hyper-parameter T

Figure 3 illustrates the impact of the hyper-parameter threshold T in SCCA during the training process within a single batch. A noticeable reduction is observed in the number of aligned pairs when T is applied, compared to the case without thresholding. This reduction is not merely a quantitative change, but reflects a qualitative refinement. Specifically, the introduction of T imposes a discriminative constraint that prioritizes pairs with relatively high similarity scores, thereby filtering out potentially ambiguous or noisy associations. As a result, only those cross-modal pairs with strong mutual affinity are retained, which enhances the reliability of the alignment process and provides a purer supervisory signal for model optimization.

As shown in Figure 4, the aligned pairs selected using the threshold T exhibit markedly higher accuracy compared to those obtained without thresholding, demonstrating the effectiveness of T in enhancing alignment quality. By enforcing a similarity-based constraint, T enables the model to focus on more reliable cross-modal associations. Moreover, as training advances, the number of high-confidence aligned pairs steadily increases, indicating that the model not only becomes more adept at identifying true correspondences but also progressively suppresses the influence of noisy or ambiguous pairings. This dynamic evolution underscores the role of T in promoting both stability and discriminability in the learning process.

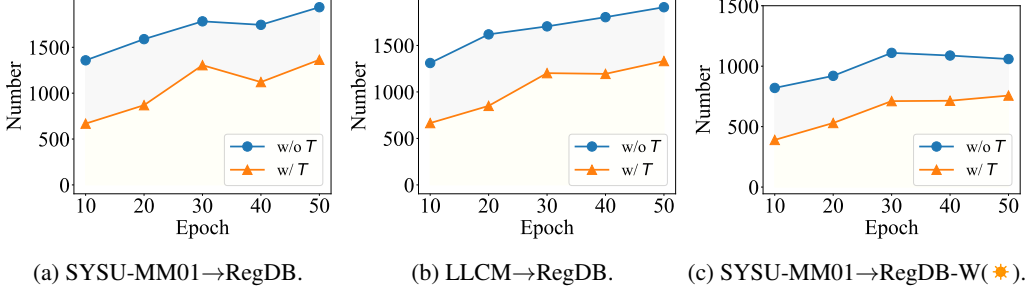


Figure 3: Effect of the threshold T in SCCA on the number of aligned pairs.

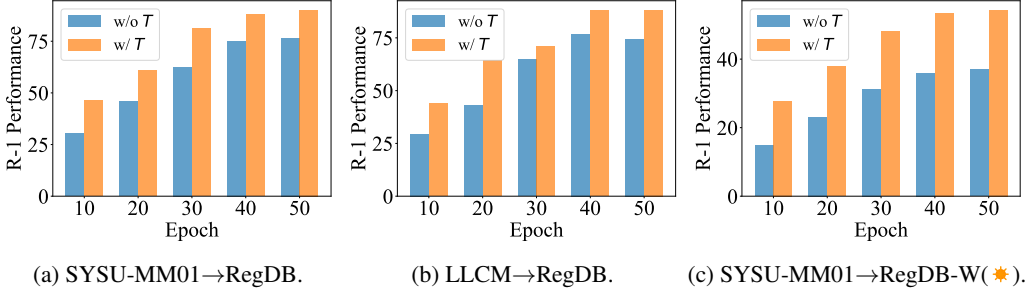


Figure 4: Accuracy analysis of aligned pairs under the threshold T in SCCA.

121 D.3 Sensitivity of Hyper-parameters

122 To evaluate the sensitivity of SVIP to the trade-off hyperparameters λ_1 and λ_2 , which balance the
 123 contributions of RTL and SCCA, we conduct an independent grid search for each and present the
 124 results in Figure 5. The analysis reveals that the best performance is achieved when $\lambda_1 = 0.2$ and
 125 $\lambda_2 = 0.8$, indicating that both components are indispensable. Notably, setting either $\lambda_1 = 0$ or
 126 $\lambda_2 = 0$ leads to a significant performance drop, underscoring the necessity of jointly optimizing both
 127 objectives. These results validate our design choice of integrating feature-level residual guidance
 128 with structure-aware cross-modal alignment, which together enable SVIP to mitigate domain shifts
 129 and modality gaps more effectively.

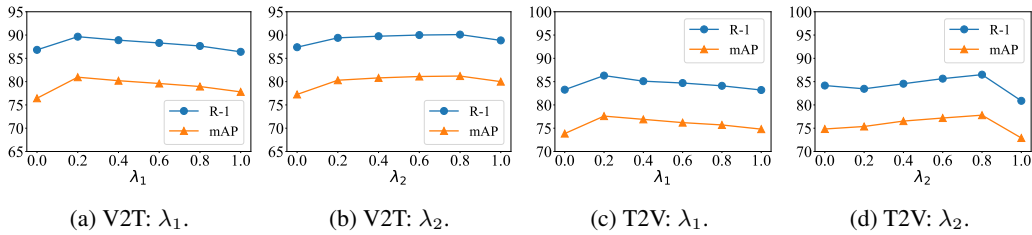


Figure 5: Performance on *Basic Setting* (SYSU-MM01→RegDB) under varied hyper-parameters λ_1 and λ_2 .

130 D.4 SCCA v.s. Hungarian Algorithm

131 To evaluate the effectiveness of the proposed SCCA mechanism, we conduct a detailed comparison
 132 with the widely adopted Hungarian algorithm [4] under the *Weather Setting* in VI-ReID. As shown
 133 in Table 5, SCCA consistently outperforms the Hungarian algorithm, demonstrating the following
 134 key advantages:

- The Hungarian algorithm relies solely on global feature matching and lacks explicit strategies to address modality heterogeneity. In contrast, SCCA incorporates *bidirectional consistency* via reciprocity and lower-bound constraints, which explicitly alleviate modality discrepancies and help filter out unreliable cross-modal associations.

- SCCA achieves consistently superior performance across all severity levels. Specifically, on the LLCM-W dataset, it obtains average Rank-1 accuracies of 37.36%, 34.01%, and 31.18% at Severity Levels 1, 2, and 3, respectively, whereas the Hungarian algorithm achieves only 35.25%, 32.37%, and 29.69% under the same conditions. These results highlight the robustness of SCCA in increasingly adverse weather scenarios.

Table 5: Comparison of Hungarian and SCCA under the *Weather Setting*, reporting Rank-1 (%) accuracy on LLCM-W (Visible to Infrared) and RegDB-W (Visible to Thermal) datasets across different weather-induced severity levels.

			Method		LLCM-W (Visible to Infrared)						RegDB-W (Visible to Thermal)					
					Sunny	Fog	Frost	Rain	Snow	Avg.	Sunny	Fog	Frost	Rain	Snow	Avg.
Severity Level 1	CMT	Hungarian	27.16	35.28	27.89	33.77	25.48	29.91	55.31	74.30	54.57	67.66	43.18	59.00		
		SCCA	29.04	36.55	29.41	34.58	26.04	31.13	60.81	75.57	52.79	75.57	39.56	60.86		
	LEAD	Hungarian	31.18	37.74	23.57	27.00	28.01	29.50	56.28	75.69	53.94	77.12	41.07	60.82		
		SCCA	32.56	38.26	26.55	27.76	28.06	30.64	56.64	78.09	53.50	77.02	43.24	61.70		
	DRU	Hungarian	36.90	36.61	31.29	33.51	27.82	33.23	58.20	78.75	52.04	75.20	50.36	62.91		
		SCCA	37.87	37.91	32.74	33.04	28.33	33.98	59.06	79.51	55.11	76.37	50.78	64.16		
	SVIP	Hungarian	35.42	39.87	36.12	35.08	29.78	35.25	58.34	75.26	55.73	77.11	51.39	63.57		
		SCCA	37.66	42.04	38.58	37.12	31.41	37.36	61.00	78.59	58.40	80.84	54.02	66.57		
Severity Level 2	CMT	Hungarian	27.16	35.28	27.89	33.77	25.48	29.91	55.31	74.30	54.57	67.66	43.18	59.00		
		SCCA	29.04	36.55	29.41	34.58	26.04	31.13	60.81	75.57	52.79	75.57	39.56	60.86		
	LEAD	Hungarian	31.18	37.74	23.57	27.00	28.01	29.50	56.28	75.69	53.94	77.12	41.07	60.82		
		SCCA	32.56	38.26	26.55	27.76	28.06	30.64	56.64	78.09	53.50	77.02	43.24	61.70		
	DRU	Hungarian	36.90	36.61	31.29	33.51	27.82	33.23	58.20	78.75	52.04	75.20	50.36	62.91		
		SCCA	37.87	37.91	32.74	33.04	28.33	33.98	59.06	79.51	55.11	76.37	50.78	64.16		
	SVIP	Hungarian	29.87	38.54	33.71	32.22	27.49	32.37	55.93	72.64	52.86	70.45	48.31	60.04		
		SCCA	31.12	40.65	35.52	33.91	28.83	34.01	58.11	75.81	55.10	73.88	50.73	62.73		
Severity Level 3	CMT	Hungarian	23.74	32.23	22.05	22.71	17.57	23.66	45.05	60.38	44.82	57.36	36.30	48.78		
		SCCA	24.89	33.04	23.23	23.62	18.25	24.61	50.82	61.78	44.79	63.86	33.51	50.95		
	LEAD	Hungarian	23.73	30.74	18.68	28.05	18.59	23.96	46.60	64.52	43.77	63.39	34.34	50.52		
		SCCA	25.73	33.75	20.54	28.23	19.95	25.64	47.18	65.21	44.06	63.95	35.29	51.14		
	DRU	Hungarian	21.03	31.99	24.84	24.00	19.99	24.37	48.78	64.99	44.05	62.69	40.68	52.24		
		SCCA	23.47	32.90	26.57	26.39	20.38	25.94	48.48	65.78	44.40	64.05	41.38	52.82		
	SVIP	Hungarian	26.45	34.27	30.68	30.91	26.12	29.69	51.86	63.03	47.20	64.81	44.18	54.22		
		SCCA	28.02	36.13	32.10	32.19	27.48	31.18	54.32	65.92	49.54	67.46	46.05	56.66		

D.5 Pluggability of Source Model

To evaluate the pluggability of the source model, we investigate the effectiveness of our approach in adapting to various pre-trained models without requiring direct access to source data. Specifically, we conduct experiments using multiple source models and assess their adaptation performance on the target domain, as shown in Figure 6. We select state-of-the-art VI-ReID methods as source models, including Vanilla (i.e., the method described in Equation (14 in the main text), DEEN [3], PMT [13], DMA [14], PGM [4], and MMM [9], to verify the generalizability of our approach. The results show that our framework consistently achieves strong performance across different pre-trained models. Moreover, source models with better source-domain performance generally achieve superior adaptation results on the target domain, underscoring the robustness and adaptability of our method across diverse source models.

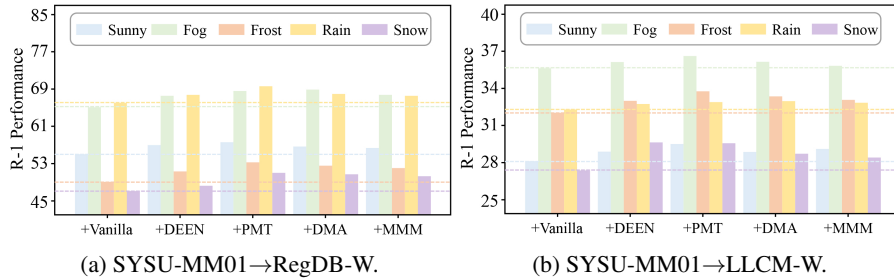


Figure 6: Adaptation performance of SVIP with different pre-trained source models.

155 D.6 Qualitative Results

156 We conduct a qualitative evaluation of SVIP by comparing it with state-of-the-art methods (e.g.,
 157 SDCL [8] and DRU [12]). To evaluate performance, we retrieve the top-5 gallery images with
 158 the highest similarity scores for each query image. As illustrated in Figure 7, the results under
 159 the *Basic Setting* demonstrate retrieval performance across two target domains: the first two rows
 correspond to RegDB, while the last two rows correspond to LLCM. Similarly, Figure 8 illustrates

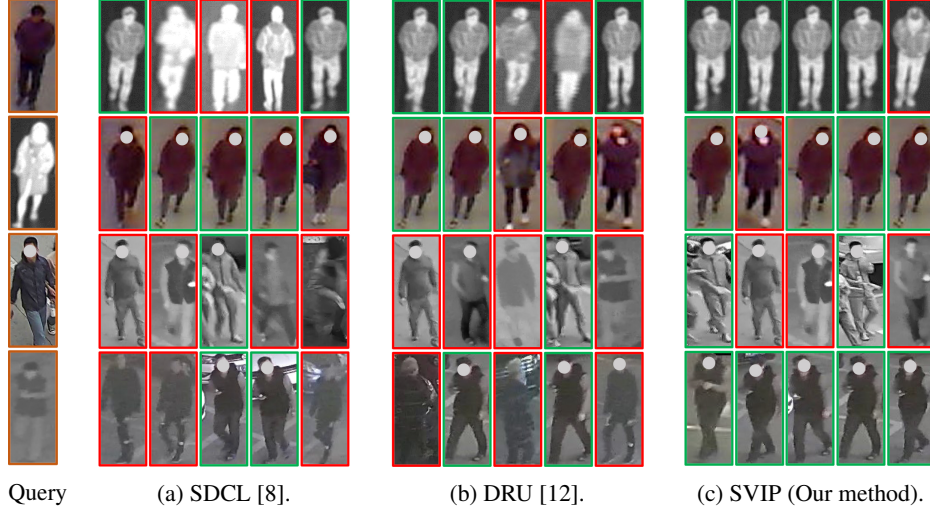


Figure 7: Some person re-identification results of (a) SDCL [8], (b) DRU [12], and (c) SVIP (Our method) on *Basic Setting*. Each row presents a query image of a person (marked with orange) on the left, with the retrieved images highlighted in green denoting correct matches and those in red indicating incorrect matches. The results are arranged in descending order.

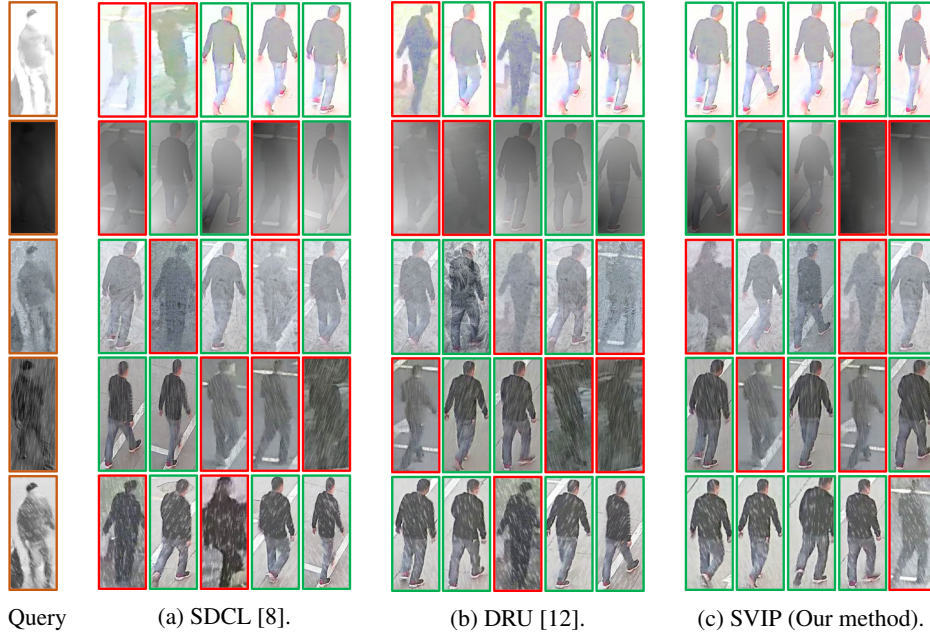


Figure 8: Some person re-identification results of (a) SDCL [8], (b) DRU [12], and (c) SVIP (Our method) on *Weather Setting*. Each row presents a query image of a person (marked with orange) on the left, with the retrieved images highlighted in green denoting correct matches and those in red indicating incorrect matches. The results are arranged in descending order.

retrieval outcomes under the *Weather Setting*, where each row corresponds to a distinct weather condition, i.e., *Sunny*, *Fog*, *Frost*, *Rain*, and *Snow*. SVIP consistently achieves more stable and accurate matches than other methods, demonstrating its effectiveness in mitigating domain shifts and improving re-identification robustness across diverse environments.

E Limitations and Potential Impact Statement

While the proposed SVIP framework has demonstrated strong performance across diverse domain adaptation scenarios in VI-ReID, there are several aspects that deserve further reflection.

One consideration lies in the reliance on a pretrained source model to guide adaptation in the absence of source data. Although this approach aligns with the source-free setting and addresses practical concerns such as privacy and efficiency, the effectiveness of knowledge transfer is inherently influenced by the generalizability of the source model. In cases where the source domain significantly differs from the target domain in terms of visual characteristics or identity distribution, the adaptation process may become less effective due to the limited capacity of model to bridge substantial domain gaps. Another aspect concerns the assumption of availability of a reliable source model. In certain real-world deployments, especially in resource-constrained or privacy-sensitive environments, obtaining a sufficiently strong source model may not always be trivial. Extending the framework to jointly estimate and refine source knowledge in a self-contained manner could further enhance its autonomy and scalability.

From a broader perspective, the development of SFDA techniques such as SVIP carries both promising and cautionary implications. On the positive side, our method facilitates privacy-preserving deployment in surveillance and smart security applications, reducing reliance on centralized data collection and annotation. However, as with any re-identification system, care must be taken to ensure appropriate safeguards are in place to prevent misuse, particularly in scenarios involving surveillance, profiling, or other sensitive contexts.

References

- [1] 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*, pages 5380–5389, 2017.
- [2] 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):605, 2017.
- [3] Yukang Zhang and Hanzi Wang. Diverse embedding expansion network and low-light cross-modality benchmark for visible-infrared person re-identification. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 2153–2162, 2023.
- [4] Zesen Wu and Mang Ye. Unsupervised visible-infrared person re-identification via progressive graph matching and alternate learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 9548–9558, 2023.
- [5] 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 re-identification. In *European Conference on Computer Vision*, pages 93–109. Springer, 2022.
- [6] Zhong Chen, Zhizhong Zhang, Xin Tan, Yanyun Qu, and Yuan Xie. Unveiling the power of clip in unsupervised visible-infrared person re-identification. In *Proceedings of the 31st ACM International Conference on Multimedia*, pages 3667–3675, 2023.
- [7] 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*, pages 11069–11079, 2023.
- [8] Bin Yang, Jun Chen, and Mang Ye. Shallow-deep collaborative learning for unsupervised visible-infrared person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16870–16879, 2024.

- 210 [9] Jiangming Shi, Xiangbo Yin, Yeyun Chen, Yachao Zhang, Zhizhong Zhang, Yuan Xie, and
211 Yanyun Qu. Multi-memory matching for unsupervised visible-infrared person re-identification.
212 In *European Conference on Computer Vision*, pages 456–474. Springer, 2025.
- 213 [10] Shengcao Cao, Dhiraj Joshi, Liang-Yan Gui, and Yu-Xiong Wang. Contrastive mean teacher
214 for domain adaptive object detectors. In *Proceedings of the IEEE/CVF conference on computer
215 vision and pattern recognition*, pages 23839–23848, 2023.
- 216 [11] Sanqing Qu, Tianpei Zou, Lianghua He, Florian Röhrbein, Alois Knoll, Guang Chen, and
217 Changjun Jiang. Lead: Learning decomposition for source-free universal domain adaptation. In
218 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages
219 23334–23343, 2024.
- 220 [12] Trinh Le Ba Khanh, Huy-Hung Nguyen, Long Hoang Pham, Duong Nguyen-Ngoc Tran, and
221 Jae Wook Jeon. Dynamic retraining-updating mean teacher for source-free object detection. In
222 *European Conference on Computer Vision*, pages 328–344. Springer, 2024.
- 223 [13] Hu Lu, Xuezhong Zou, and Pingping Zhang. Learning progressive modality-shared transformers
224 for effective visible-infrared person re-identification. In *Proceedings of the AAAI conference on
225 artificial intelligence*, volume 37, pages 1835–1843, 2023.
- 226 [14] Zhenyu Cui, Jiahuan Zhou, and Yuxin Peng. Dma: Dual modality-aware alignment for visible-
227 infrared person re-identification. *IEEE Transactions on Information Forensics and Security*,
228 2024.