Bridging Day and Night: Unsupervised Cross-Domain Re-Identification with Synergistic Prompt and Prototype Learning

Supplementary Material

In this supplementary material, we present further empirical evidence to substantiate the performance and advantages of the proposed method.

1. Supplementary Experiments

Experiments Details. Table 1 summarizes the data configurations of the DN-348 and DN-Wild benchmarks. DN-348 offers a balanced distribution of daytime and nighttime samples, whereas DN-Wild is substantially larger in scale and naturally exhibits a pronounced day–night imbalance. In addition, we report the computational characteristics of the proposed framework in Table 2. The overall training pipeline is outlined in Algorithm 1.

Datasets	Train(Day)	Train(Night)	Test(Day)	Test(Night)
DN-348	200/9962	200/10022	148/10121	148/3972
DN-Wild	1574/70981	1574/35384	712/14964	712/19568

Table 1. Details of the dataset configuration, where */* denotes ID/Sample Size.

Modules	FLOPs	Params
Image-Encoder	14.70(G)	85.65M
Text-Encoder	1.94(G)	38.13M
Dynamic-bias Net	143.36(K)	0.139M

Table 2. Computational complexity of the model components.

Generalization Ability of Individual Modules. Table 3 presents the cross-domain generalization performance of individual modules when transferring between DN-Wild and DN-348. The baseline already exhibits strong bidirectional transferability, suggesting that prototype-based representation learning inherently contributes to stable cross-domain generalization. Incorporating IPL yields further improvements in Rank-1 and mAP across all transfer settings by providing more reliable semantic alignment between visual instances and their dynamically adapted prompts. CPML alone also offers substantial gains by enforcing cross-domain prototype consistency, achieving more reliable cross-domain identity alignment. When combined, IPL and CPML deliver the most consistent performance enhancement, demonstrating their complementary roles in strengthening the model's generalization ability under diverse day–night domain configurations.

Algorithm 1 The training procedure of proposed method.

Input: Unlabeled day-night dataset D, image encoder E_{img}, text encoder E_{txt}, Dynamic-Bias Net f_θ

Stage-1:

Freeze E_{img} and E_{txt}; initialize learnable prompt tokens $\{V_1, ..., V_M\}$
for image I_i in D with epoch [0, epochs] **do**
  Extract visual features $f_i = E_{img}(I_i)$ and compute instance-specific bias $f_\theta(f_i)$.
  Construct textual prompts according to Eq. 1 and obtain textual embeddings t_i via E_{txt}.
  Compute bidirectional contrastive loss using Eq. 2-3.
end for
Save learned prompts and dynamic-bias net.

Stage-2:

Perform DBSCAN clustering and initializing empty prototype memory banks $\{\phi_d, \phi_n\}$.
Initialize domain-specific classifier parameters $\{\psi_d, \psi_n\}$ using the corresponding prototypes.
for image I in D with epoch [0, iterations] **do**
  Update prototypes via momentum using Eq. 5.
  Calculate intra-domain loss: prototype contrastive loss $\mathcal{L}_{pcl}$ (Eq. 6) and pseudo-label classification loss $\mathcal{L}_{id}$ (Eq. 7-8).
  Recompute textual centroids within batch; Compute $\mathcal{L}_{i2tce}$ by Eq. 9.
  Construct cross-domain k-NN sets $\mathcal{R}(\phi_d), \mathcal{R}(\phi_n)$.
  Identify mutual nearest neighbors as positive pairs, with remaining neighbors treated as negatives.
  Compute cross-domain prototype matching loss $\mathcal{L}_{cpml}$ following Eq. 13.
  Update E_{img} by optimizing the composite objective: $\mathcal{L} = \mathcal{L}_{id} + \mathcal{L}_{pcl} + \mathcal{L}_{i2tce} + \mathcal{L}_{cpml}$.
end for
return E_{img}

Pseudo-label Accuracy. As illustrated in Figure 1, we conduct a comprehensive comparison of clustering performance with several state-of-the-art methods on the DN-348 dataset. To ensure a fair and controlled evaluation, all approaches adopt the DBSCAN clustering algorithm with same hyperparameter settings; in particular, the maximum distance threshold eps is fixed to 0.7 for both domains. The specific clustering metrics are explained as follows:

- ARI (Adjusted Rand Index) measures the similarity between two data partitionings.

Index	Components			DN-Wild→DN-348						DN-348→DN-Wild					
				Day-to-Night			Night-to-Day			Day-to-Night			Night-to-Day		
	Baseline	IPL	CPML	Rank1	Rank5	mAP	Rank1	Rank5	mAP	Rank1	Rank5	mAP	Rank1	Rank5	mAP
1	✓			0.637	0.799	0.347	0.689	0.829	0.352	0.462	0.930	0.234	0.423	0.859	0.242
2	✓	✓		0.655	0.810	0.373	0.723	0.857	0.362	0.470	0.935	**0.244**	0.430	**0.884**	**0.256**
3	✓		✓	0.663	0.804	0.365	0.689	0.839	0.363	0.474	0.935	0.241	**0.432**	0.878	0.252
4	✓	✓	✓	**0.676**	**0.814**	**0.375**	**0.748**	**0.874**	**0.375**	**0.476**	**0.942**	0.240	0.431	0.881	0.254

Table 3. Cross-domain generalization performance of individual modules, evaluating the impact of IPL and CPML when transferring between DN-Wild and DN-348.

- AMI (Adjusted Mutual Information) quantifies the agreement between clusters by normalizing shared information.
- FMI (Fowlkes-Mallows Index) evaluates the geometric mean of pairwise precision and recall.
- V-Measure computes the harmonic mean of completeness and homogeneity.

The proposed method (Ours) demonstrates superior performance across multiple clustering evaluation metrics, consistently outperforming recent approaches including RPNR, NULC, PCLHD, and PCA. This performance gain can be attributed to our synergistic integration of instance-aware prompt learning and cross-domain prototype matching, which effectively models the complex many-to-many correspondences between day and night clusters. The results validate that our framework establishes more robust identity associations under significant illumination shifts, bridging the domain gap more effectively than existing unsupervised methods.

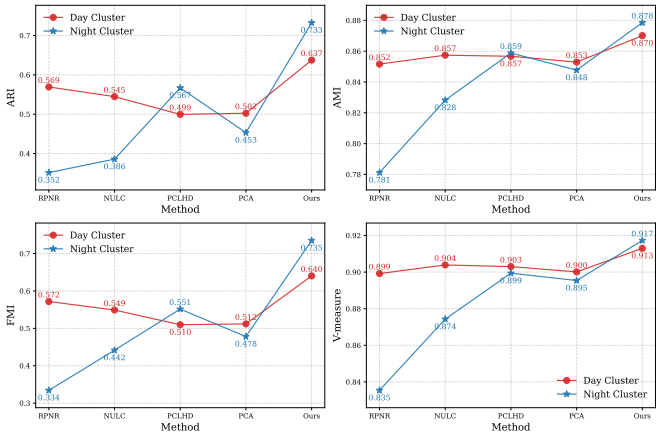

Figure 1. Performance comparison of clustering metrics against state-of-the-art methods on DN-348.

Visualization of Grad-CAM Activation Map. Figure 2 presents a comparison of Grad-CAM activation maps between the baseline and our method. The baseline model exhibits scattered and unstable attention, frequently responding to background clutter or illumination artifacts, particularly in nighttime scenes where glare and low-light noise dominate the visual field. In contrast, our method consistently concentrates on semantically meaningful and identity-relevant structures—such as headlights, grilles, and brand contours—across both day and night conditions. This stable focus indicates that the proposed framework alignment effectively suppress illumination bias and guide the network toward domain-invariant cues.

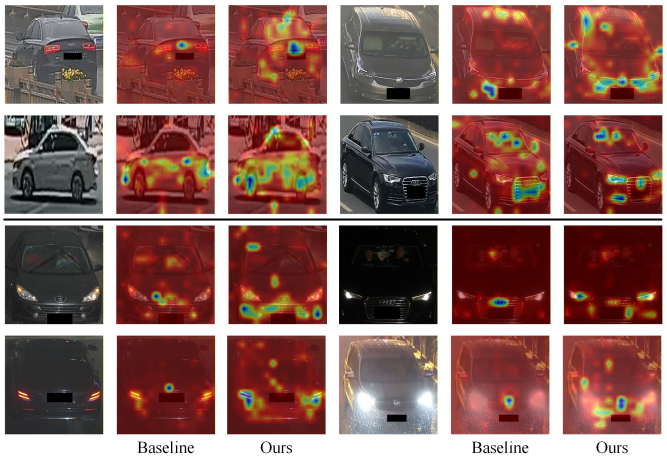

Figure 2. Comparative visualization of Grad-CAM activation maps for the baseline model and the proposed method.

Cross-domain Retrieval Ranklist. Figure 3 illustrates the top-5 cross-domain retrieval results for both day-to-night and night-to-day settings. Compared with the baseline, our method retrieves significantly more correct matches, especially under severe illumination changes where the baseline frequently confuses vehicles with similar colors or headlight patterns. The proposed approach consistently ranks the true identity at higher positions and maintains stable retrieval across varying viewpoints and lighting conditions. In challenging nighttime queries with glare or low-visibility regions, the baseline often fails to capture discriminative structure and produces visually similar but incorrect candidates, whereas our method preserves identity-relevant cues and suppresses illumination bias, leading to more accurate cross-domain association.

Figure 3. A comparison of the top-5 retrieval results under cross-domain day-night conditions is presented for the proposed method and the baseline. A green bounding box indicates a match with the ground truth identity, whereas a red box denotes a mismatch.

2. Discussion

Our work reveals several insights regarding the design choices and remaining challenges in unsupervised day–night vehicle re-identification. First, although the text encoder and prompt tokens are frozen in Stage-2, the textual representations remain dynamically updated through the instance-specific bias generated by the Dynamic-Bias Net. This mechanism effectively adapts the text features to the evolving visual representations, preventing semantic drift and preserving the benefits of Stage-1 alignment throughout the entire training pipeline. Second, CPML constructs negative pairs from mutually non-matching cross-domain prototype neighbors, which naturally form the hardest negative samples in the latent space. These hard negatives enforce sharper prototype boundaries and encourage the model to refine cross-domain identity discrimination.