CIA: Cluster-Instance Alignment for Unsupervised Day-Night Vehicle Re-Identification

Yongguo Ling[1,2], **Chen Zhang**[1], **Yiming Liu**[1], **Wenhao Shao**[1*]

[1]School of Computer, Electronics and Information, Guangxi University, China
[2]Guangxi Key Laboratory of Multimedia Communications and Network Technology, China
{ygling,wenhao.shao}@gxu.edu.cn, {2413393088,2313394024}@st.gxu.edu.cn

Abstract

Cross-time vehicle re-identification (Re-ID), especially across day and night conditions, remains a challenging problem due to drastic illumination variations that lead to significant domain shifts. While existing methods perform well under daytime scenarios, their effectiveness degrades severely in cross-domain settings, and fully supervised solutions demand costly annotations in both domains. In this paper, we introduce a new setting, Unsupervised Day-Night Vehicle Re-Identification (USL-DN-ReID), and propose a novel Cluster-Instance Alignment (CIA) framework to address it. CIA performs dual-level alignment: 1) at the cluster level, a Dictionary-Guided Graph Matching (DGM) module builds a cross-domain topological graph using soft similarities among cluster centers and solves global matching via the Hungarian algorithm; 2) at the instance level, a Multi-Factor Adaptive Alignment (MAA) module introduces a multi-factor adaptive weighting strategy that emphasizes high-confidence pairwise relations while suppressing noise. Together, these components enable robust and scalable cross-domain adaptation without requiring target-domain labels. Extensive experiments conducted on the DN-348 and DN-Wild benchmarks demonstrate the effectiveness and superiority of the proposed CIA framework, setting new state-of-the-art results on both datasets.

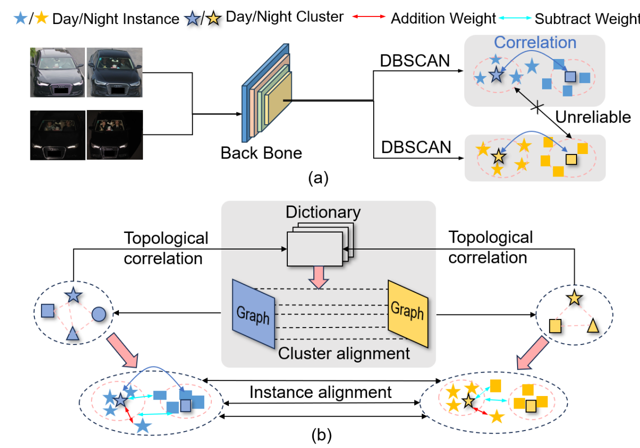

Figure 1: (a) Severe cross-domain gaps impair the ability of unsupervised models to learn consistent identity representations, resulting in unreliable cross-domain Correlation. (b) The proposed CIA framework for USL-DN-ReID jointly performs cross-domain alignment at both the cluster-level via topological structures and the instance-level with adaptive weighting, enabling reliable unsupervised cross-domain matching.

Introduction

Vehicle re-identification (Re-ID) aims to match images of the same vehicle captured by non-overlapping cameras and plays a key role in intelligent transportation systems. While existing methods (Yang et al. 2023; Shen et al. 2023; Li et al. 2022) perform well under daytime conditions, they struggle in cross-time scenarios, particularly between day and night, due to drastic illumination-induced appearance changes. This domain shift severely degrades model performance when trained only on daytime data. To address this, the Day-Night Vehicle Re-ID (DN-ReID) task (Li et al. 2024) has been proposed. However, it relies on fully supervised learning with dense annotations in both domains, leading to high labeling costs and limited scalability.

To alleviate the annotation burden, we propose an unsupervised learning paradigm for the DN-ReID task, termed Unsupervised Day-Night Vehicle Re-Identification (USL-DN-ReID). Compared with conventional unsupervised vehicle Re-ID methods (Lu et al. 2023; Wang et al. 2023; Xu et al. 2025), USL-DN-ReID faces a key challenge: the significant domain gap between daytime and nighttime images hinders consistent identity learning. Relying solely on clustering for cross-domain alignment often yields unreliable associations, as shown in Fig. 1(a). Moreover, related unsupervised cross-domain matching methods (Wu and Ye 2023; Yang et al. 2022) rely on fixed similarity metrics and thresholding strategies, failing to adapt to feature distribution shifts caused by lighting variations, which leads to accumulated matching bias and reduced robustness in cross-domain alignment.

To address these challenges, we propose a Cluster-Instance Alignment (CIA) framework for USL-DN-ReID, which jointly considers cross-domain alignment at both the cluster-level with topological structure and the instance-

*Corresponding author

level with adaptive weighting, thereby achieving reliable unsupervised cross-domain matching under day-night scenarios, as illustrated in Fig. 1 (b). CIA explores multi-dimensional correlation structures by explicitly modeling the mutual information among cluster centers and dynamically adapting to the correlation strength between instances. This joint modeling leads to more accurate and robust feature alignment, significantly improving performance in cross-domain matching. Specifically, CIA decomposes the alignment into two complementary components: **1) Dictionary-Guided Graph Matching (DGM)**: From the cluster-level perspective, we introduce a dictionary-guided graph matching algorithm that leverages mutual information between cluster centers. Specifically, DGM constructs a mutual information dictionary by estimating soft similarity scores among cluster centers, which is further used to build a cross-domain topological graph encoding high-confidence correspondences. Under the joint constraints of the dictionary and the graph, DGM employs the Hungarian algorithm (Bruff 2005) to achieve globally optimal cluster matching. In addition, DGM integrates a cross-domain mapping function to improve the robustness of the pseudo labels. **2) Multi-Factor Adaptive Alignment (MAA)**: From the instance-level perspective, we design an MAA method that explicitly captures inter-instance correlations during alignment, enabling more adaptive and effective feature alignment. Unlike conventional MMD (Jambigi, Rawal, and Chakraborty 2021), which overlooks structural relationships and treats all sample pairs equally—making it vulnerable to the influence of hard negative pairs—our MAA emphasizes the structural correlations between samples and assigns higher weights to more similar pairs by introducing a multi-factor adaptive weighting (MAW) strategy, thereby mitigating the adverse impact of hard pairs on the alignment process. By enforcing structural consistency within and across domains, this module facilitates smoother and more stable cross-domain feature alignment.

In summary, our main contributions are as follows:

- We introduce the USL-DN-ReID task and propose a novel CIA framework that performs dual-level cross-domain alignment in a fully unsupervised manner.
- We develop a Dictionary-Guided Graph Matching (DGM) module that captures mutual information among cluster centers to model their topological structure and enable reliable cross-domain cluster alignment.
- We propose the Multi-Factor Adaptive Alignment (MAA) method, which explicitly takes into account the correlation between instances during the alignment process, leading to more flexible and efficient feature adaptation.
- Extensive experiments on the DN-348 and DN-Wild datasets demonstrate that our method sets a new state-of-the-art performance for the unsupervised USL-DN-ReID task.

Related Work

Vehicle re-identification. Liu et al. (Liu et al. 2016) introduced the VeRi-776 benchmark and developed a deep relative-distance learning method. Lou et al. (Lou et al. 2019) created the VERI-Wild dataset for real-world conditions and employed an adversarial feature-distance strategy to mine hard negatives. Zhong et al. (Zhong et al. 2021) assembled the VSW unsupervised search dataset and devised the UDANet domain-adaptation framework. Li et al. (Li et al. 2024) released two day–night datasets (DN-348 and DN-Wild) and proposed the Day-Night Dual-domain Modulation (DNDM) framework to improve robustness across illumination changes.

Unsupervised vehicle re-identification. In domain adaptation and pseudo-supervised learning, Yu et al. (Yu et al. 2023) present Weakly Supervised Contrastive Learning (WSCL), which aligns cross-camera domains via Entropy Prototype Clustering (EPC) and Progressive Sub-domain Alignment (PSSA). Zhang et al. (Zhang et al. 2024) introduce MATNet—a Multi-domain Adaptive Transformer with Dual Mutual Dynamic Update (DMDU) for refined pseudo-labels. Wang et al. (Wang et al. 2023) propose a progressive few-shot UDA framework that couples target-domain clustering with pseudo-label generation through Dual-Constrained Label Smoothing Regularization (DCLSR) and Domain Difference Penalty Terms (DDPT). Qiu et al. (Qiu et al. 2024a,b) apply a two-stage Camera-aware Differential Clustering (CDC) and an Intra-/Inter-Cluster Reorganization (ICR) to suppress pseudo-label noise. Xu et al. (Xu et al. 2025) leverage CLIP in ViewCoOp, using dynamic multi-view text prompts and cross-domain mutual-information graphs to fine-tune multi-view representations. Taleb et al. (Taleb, Zhu, and Wang 2024) treat each camera as a style domain and employ StarGAN2 for cross-camera style transfer, reducing viewpoint-induced variation. Holla et al. (Holla et al. 2025) integrate single-platform, cross-platform, multi-modal, and cross-domain paradigms into a unified framework that charts vehicle re-ID progress and flags challenges like cross-view matching and extreme-environment adaptation. Lu et al. (Lu et al. 2023) mitigate pseudo-label noise in unsupervised re-ID by combining mask-aware feature extraction with adaptive threshold neighborhood consistency. Kuang et al. (Kuang et al. 2023) isolate identity features using a variational autoencoder paired with mutual information minimization to eliminate background clutter and viewpoint variation.

Method

In this section, we introduce the details of our proposed Cluster-Instance Alignment (CIA) framework, illustrated in Fig.2. CIA adopts a dual-stream architecture (Yang et al. 2022) based on ResNet-50 (He et al. 2016). Specifically, we divide the shallow layers (the initial convolution and first two residual blocks) into domain-specific branches to capture domain-aware features, and merge them into shared layers (the 3rd and 4th blocks) for domain-invariant representation learning. After feature extraction, we apply DBSCAN (Ester et al. 1996) independently to generate pseudo labels for each domain. To enable effective cross-domain alignment, CIA performs dual-level robust cross-domain: 1) at the cluster level, a Dictionary-Guided Graph Matching (DGM) module captures structural relations among clus-

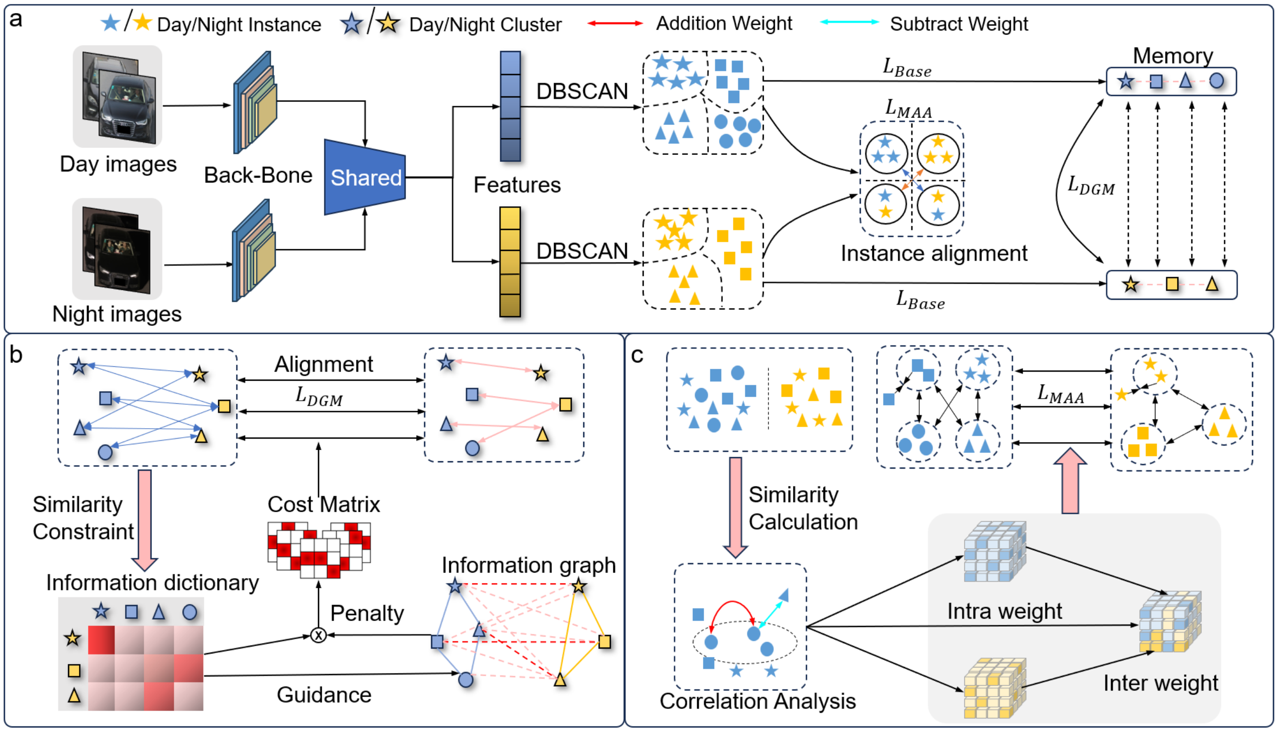

Figure 2: Overview of the proposed CIA framework. (a) Dual-branch ResNet-50 extracts domain-specific features from daytime and nighttime images, followed by DBSCAN-based pseudo label generation. CIA performs dual-level alignment to bridge domain gaps. (b) DGM module aligns cluster structures via a mutual information dictionary and cross-domain graph, solved using the Hungarian algorithm. (c) MAA module embeds the MAW mechanism into the MMD, enhancing instance-level alignment by emphasizing reliable pairs and suppressing noisy ones.

ter centers for robust topological alignment; and 2) at the instance level, a Multi-Factor Adaptive Alignment (MAA) emphasizes reliable sample pairs to guide robust feature alignment.

Base

Given a day-night vehicle dataset consisting of K^d daytime images $D = \{d_i\}_{i=1}^{K^d}$ and K^n nighttime images $N = \{n_i\}_{i=1}^{K^n}$, we extract daytime features $F^d = \{f_1^d, \ldots, f_{K^d}^d\}$ and nighttime features $F^n = \{f_1^n, \ldots, f_{K^n}^n\}$. To generate supervisory signals for unsupervised learning, we apply DBSCAN (Ester et al. 1996) clustering to each domain independently, obtaining pseudo labels. The center of the k-th cluster in domain $r \in \{d, n\}$ is computed as:

$$C_k^r = \frac{1}{|\phi_k^r|} \sum_{i \in \phi_k^r} f_i^r, \tag{1}$$

where ϕ_k^r denotes the index set of the k-th cluster.

To enforce intra-cluster compactness and inter-cluster separability, we adopt the ClusterNCE loss (Dai et al. 2022):

$$L_r = -\sum_{i=1}^{K^r} \log \frac{\exp\left(C_+^r \cdot f_i^r / \tau\right)}{\sum_{k=1}^{K^r} \exp\left(C_k^r \cdot f_i^r / \tau\right)}, \tag{2}$$

where C_+^r is the positive cluster center associated with f_i^r, and τ is a temperature hyperparameter. The final baseline objective aggregates the domain-specific losses:

$$L_{\text{base}} = L_d + L_n. \tag{3}$$

Dictionary-Guided Graph Matching

To enhance the robustness of cross-domain association and alignment, we abandon the conventional approach of directly computing cosine similarities between cluster centers. Instead, we leverage the mutual information among clusters to construct a structured topology and adaptively discover reliable correspondences across domains. Specifically, we propose a novel Dictionary-Guided Graph Matching (DGM) strategy that imposes dual topological constraints derived from a mutual information dictionary and a cross-domain affinity graph. The dictionary encodes latent semantic dependencies among clusters, while the graph captures their structural relations across domains. By jointly reasoning over both structures, DGM effectively filters noisy matches and selects high-confidence cluster pairs to guide alignment. Moreover, we integrate a cross-domain mutual mapping loss into DGM to align feature distributions across domains.

Mutual Information Dictionary. To capture nuanced correlations between clusters across domains with varying confidence levels, we construct a mutual information dictionary based on cross-domain similarity matrices of cluster centers. Given cluster centers C_i^d from the day domain and

C_j^n from the night domain, we compute their cosine similarity as:

$$S_{ij} = \frac{\langle C_i^d, C_j^n \rangle}{||C_i^d|| \cdot ||C_j^n||}, \tag{4}$$

where $\langle \cdot, \cdot \rangle$ means the inner product operation, $|| \cdot ||$ denotes the L2 norm.

To better distinguish reliable matches from ambiguous ones, we further introduce a piecewise continuous weighting function to build the mutual information dictionary M_{ij}:

$$M_{ij} = \begin{cases} 1 + \alpha(S_{ij} - S_\tau), & \text{if } S_{ij} \geq S_\tau \\ \alpha \cdot S_{ij}, & \text{otherwise} \end{cases}, \tag{5}$$

where S_τ is a similarity threshold, and α is a tunable parameter that controls the sensitivity and scaling of the mapping. This adaptive weighting emphasizes confident cross-domain associations while suppressing noisy alignments.

Cross-Domain Topological Graph. Leveraging the constructed mutual information dictionary, we build a cross-domain topological graph to capture the structural correlations between cluster centers across domains. In this graph, cluster centers from different domains are treated as nodes, and the edge weights are determined using a dynamic thresholding strategy defined as:

$$T_{ada} = T_0 + w_0 \cdot T_{epo}, \tag{6}$$

where T_0 is the initial threshold, w_0 is the difference between a final threshold and an initial threshold, and T_{epo} is the ratio of the current epoch to the total epochs, which linearly increases the threshold to dynamically adjust the filtering strategy.

Based on the dynamic threshold, the model progressively filters and preserves only the most reliable mutual information connections, resulting in a sparse yet trustworthy cross-domain topological graph. Guided by the established graph structure, we construct a weighted cost matrix as:

$$G_{ij} = \begin{cases} 1.0 - M_{ij}, & \text{if } M_{ij} \geq T_{ada} \\ G_p, & \text{otherwise} \end{cases}, \tag{7}$$

where G denotes the cost matrix and G_p is a penalty coefficient to discourage low-confidence pairings.

We use the Hungarian algorithm (Bruff 2005) on the graph-constrained cost matrix to obtain the global optimal assignment, which leads to reliable cross-domain bidirectional mapping pseudo-labels. Based on the mapping relationships of these pseudo-labels, we construct a bidirectional cross-domain pseudo-label set R, where R_{d2n} denotes the set of indexes for daytime samples aligned with nighttime samples, thus enhancing the alignment between cross-domain clustering centers.

Cross-domain mutual mapping loss. After obtaining the bi-directional pseudo-label mappings between the daytime and nighttime domains, we introduce a cross-domain mutual mapping loss to mitigate the discrepancy between features across domains. For the mapping from day to night, the loss is computed as:

$$\mathcal{L}_{d2n} = -\sum_{i \in R_{d2n}} M_{ij} \cdot \log \frac{\exp\left(f_i^d \cdot C_j^n / \tau\right)}{\sum_{k=1}^{K^n} \exp\left(f_i^d \cdot C_k^n / \tau\right)}, \tag{8}$$

where f_i^d is the features of the ith daytime sample, C_j^n represents the matched nighttime cluster center, where j indicates the index of the nighttime cluster center corresponding to the daytime cluster to which sample i belongs, and τ is the temperature hyperparameter.

Similarly, the reverse mapping loss $\mathcal{L}_{n2d}$ from night to day is computed analogously. The overall DGM is given by:

$$\mathcal{L}_{\text{DGM}} = \mathcal{L}_{d2n} + \mathcal{L}_{n2d}. \tag{9}$$

Multi-Factor Adaptive Alignment

The stark illumination contrast between day and night scenes introduces substantial domain shifts, which cannot be fully addressed by cluster-level alignment alone. To enhance instance-level consistency, we introduce a Multi-Factor Adaptive Alignment (MAA) that aligns feature distributions in a reproducing kernel Hilbert space (RKHS). Unlike conventional MMD (Jambigi, Rawal, and Chakraborty 2021), which treats all instance pairs uniformly and ignores intrinsic sample structures, our MMA adaptively assigns higher weights to cross-domain sample pairs with high alignment confidence, explicitly emphasizing their importance during distribution alignment. This design mitigates the influence of noisy or hard negative pairs induced by diurnal variation and enables dynamic adaptation during the alignment process.

Multi-Factor adaptive weighting. We use Eq. (4) to compute the intra-domain similarity S_{ij}^{intra} and inter-domain similarity S_{ij}^{inter}, which quantify the correlations between instances within and across domains. To facilitate precise alignment, we design a Multi-Factor Adaptive Weighting (MAW) mechanism that dynamically assigns weights to instance pairs based on basic confidence, similarity strength, and semantic consistency:

$$\mathbf{W}_{ij}^{qp} = \alpha_{\text{base}}^{qp}(i,j) \cdot \alpha_{\text{sim}}^{qp}(i,j) \cdot \alpha_{\text{class}}^{qp}(i,j), \tag{10}$$

where $q, p \in \{d, n\}$, $\alpha_{\text{base}}^{qp}(i,j)$ controls basic matching confidence, $\alpha_{\text{sim}}^{qp}(i,j)$ emphasizes similarity strength, and $\alpha_{\text{class}}^{qp}(i,j)$ captures semantic consistency. Initially, all weighting factors are set to 1.0. We then adaptively adjust these weighting factors for intra-domain and inter-domain cases.

1) Intra-domain weighting ($q = p$). To mitigate overfitting to easily aligned samples and encourage the model to focus on more challenging positive instances, we refine the intra-domain weights. Specifically, for a positive pair satisfying $\phi_i^q = \phi_j^q$, the base weight is defined as:

$$\alpha_{\text{base}}^{qp}(i,j) = 1.0 + w_1(1 - S_{ij}^{\text{intra}}), \tag{11}$$

where w_1 is a domain-specific scaling constant.

2) Inter-domain weighting ($q \neq p$). For the inter-domain weights, we adopt an adaptive thresholding strategy to suppress noisy or unreliable matches:

$$S_{\text{ada}} = S_0 + \frac{1}{K^d K^n} \sum_{i=1}^{K^d} \sum_{j=1}^{K^n} S_{ij}^{inter}, \tag{12}$$

where S_0 is a predefined constant.

If the sample pair satisfies $S_{ij}^{qp} > 0$, we assign $\alpha_{base}^{qp}(i,j) = \alpha_0$, where α_0 is a predefined constant.

If the cross-domain pair similarity exceeds the adaptive threshold, i.e., $S_{ij}^{inter} > S_{\text{ada}}$, we enhance the weight using a soft nonlinear modulation:

$$\alpha_{\text{sim}}^{qp}(i,j) = 1.0 + \alpha_0 \left(1 - \left(1 - S_{ij}^{inter}\right)^2\right), \tag{13}$$

which progressively amplifies the weights of more confident matches while suppressing those with low similarity.

If the pair meets the high-similarity threshold and shares the same pseudo label, we set the category weight factor $\alpha_{class}^{qp}(i,j) = 1.1$ to highlight semantic consistency.

MAW employs this hierarchical weight stacking strategy to achieve more robust adaptive weighting, which effectively enhances all positively matched sample pairs while suppressing noisy or unreliable ones.

MAA loss. MAA mitigates the forced alignment problem inherent in traditional MMD by integrating an MAW mechanism. The mechanism dynamically adjusts the weighting factors according to the training dynamics and the global distribution of the samples, thus adaptively regulating the contribution of instance pairs during cross-domain alignment. The square form of the MAA loss is defined as follows:

$$\begin{aligned}\mathcal{L}_{MAA}^2 = &\sum_{i,i'} \mathbf{W}_{ii'}^{dd} \cdot k\left(x_i, x_{i'}\right) + \sum_{j,j'} \mathbf{W}_{jj'}^{nn} \cdot k\left(y_j, y_{j'}\right) \\ &- \sum_{i,j} \mathbf{W}_{ij}^{dn} \cdot k\left(x_i, y_j\right) - \sum_{j,i} \mathbf{W}_{ji}^{nd} \cdot k\left(y_j, x_i\right),\end{aligned} \tag{14}$$

where $k(,)$ is the Gaussian kernel function, and W^{dd}, W^{nn}, W^{dn} and W^{nd} correspond to the four sub-blocks of the weight matrix.

Overall

By considering the proposed DGM and MAA modules, the overall loss function for model training is the sum of the base loss (3), DGM loss (9), and MAA loss (14):

$$\mathcal{L}_{total} = \mathcal{L}_{base} + \lambda \mathcal{L}_{DGM} + \beta \mathcal{L}_{MAA}, \tag{15}$$

where λ and β are both hyperparameters.

Experiments

Experimental Settings

Dataset. We evaluate the proposed Cluster-Instance Alignment (CIA) framework on two benchmark datasets: DN-348 and DNWild, both introduced in DN-ReID (Li et al. 2024). DNWild contains 70,981 daytime images and 35,384 nighttime images for training, covering 1,574 identities. The query and gallery sets consist of 14,964 daytime and 19,568 nighttime images from 712 identities, respectively. This dataset provides paired day-night images per identity, facilitating evaluation in cross-domain (day-to-night) scenarios. However, it suffers from significant sample imbalance between domains. In contrast, DN-348 offers a more balanced distribution. Its training set includes 9,962 daytime and 10,022 nighttime images from 200 identities, while the query and gallery sets contain 10,121 daytime and 3,972 nighttime images from 148 identities. Each vehicle in the training set has approximately 50 images captured under both daytime and nighttime conditions, making it a well-suited benchmark for evaluating model robustness across domains.

Evaluation Protocols. For both datasets, we adopt the Cumulative Matching Characteristic (CMC) and mean Average Precision (mAP) as evaluation metrics to assess the effectiveness of our proposed method. For CMC, we report the rank-1 (R1) accuracy.

Implementation Details

We adopt a two-stream network architecture (Yang et al. 2022) built upon ResNet-50 (He et al. 2016) as our baseline. All experiments are conducted using the PyTorch framework on a single NVIDIA GeForce RTX 3090 GPU. During training, each mini-batch consists of 8 identities, with 4 daytime and 4 nighttime images randomly sampled per identity, resulting in a total of 64 images. We apply random horizontal flipping and random erasing (Zhong et al. 2020) for data augmentation, and resize all input images to 288×144. We employ the Adam optimizer with an initial learning rate of 3.5×10^{-4}, which is reduced by a factor of 0.1 every 20 epochs. The training process spans 100 epochs. In the first 50 epochs, we exclusively utilize the MAA for initial cross-domain alignment, assigning it a weight of 0.5. In the subsequent 50 epochs, we introduce the DGM to further enhance cross-domain matching and reduce the weight of the MAA loss to 0.3.

Comparison to State-of-the-Art Methods

We compare our method with several widely used unsupervised cross-domain re-identification approaches on the DN-348 and DN-Wild datasets. The results are presented in Table 1.

Results on DN-348. On the DN-348 dataset, our method achieves substantial improvements over existing state-of-the-art methods. In the day-to-night setting, we achieve a Rank-1 accuracy of 67.74% and an mAP of 40.71%. Compared with PGM (Wu and Ye 2023), our approach improves Rank-1 and mAP by 2.76% and 1.06%, respectively. Compared with PCA (Yang, Hu, and Hu 2025), we observe improvements of 2.86% in Rank-1 and 2.16% in mAP. In the night-to-day setting, our method attains 78.88% Rank-1 and 42.20% mAP.

Results on DN-Wild. On the DN-Wild dataset, our method further demonstrates superior performance. In the day-to-night setting, we achieve a rank-1 accuracy of 49.12% and a mAP of 27.86%. Compared with PGM (Wu and Ye 2023), our approach improves rank-1 accuracy and mAP by 0.73% and 0.33%, respectively. Compared with the latest PCA (Yang, Hu, and Hu 2025), our method achieves improvements of 0.70% in rank-1 and 0.99% in mAP. In the night-to-day scenario, the rank-1 and mAP reach 46.86% and 27.51%, respectively.

Datasets		DN-348				DN-Wild			
Settings		Day-to-Night		Night-to-Day		Day-to-Night		Night-to-Day	
Method	Venue	R1	mAP	R1	mAP	R1	mAP	R1	mAP
AGW(Ye et al. 2021)	TPAMI'21	63.70	38.38	77.34	39.95	47.77	26.75	45.66	26.49
MMD ReID(Jambigi et al. 2021)	BMVC'21	63.22	39.46	78.75	40.84	48.29	26.65	44.67	26.20
OTLA(Wang et al. 2022)	ECCV'22	40.88	27.79	70.05	36.66	46.34	26.19	42.84	26.67
ADCA(Yang et al. 2022)	MM'22	63.52	38.90	76.71	40.63	48.14	26.87	45.55	26.53
GUR*(Yang, Chen, and Ye 2023)	ICCV'23	57.38	28.03	71.83	26.25	48.26	25.81	42.83	24.73
MBCCM(Cheng et al. 2023a)	MM'23	45.78	29.42	76.14	37.94	47.41	26.14	44.14	26.77
PGM(Wu and Ye 2023)	CVPR'23	64.98	39.65	77.52	41.48	48.63	27.61	46.13	27.18
SDCL(Yang, Chen, and Ye 2024)	CVPR'24	62.22	36.64	77.82	38.29	47.53	25.64	43.71	25.23
PCA(Yang, Hu, and Hu 2025)	TIFS'25	64.88	38.55	78.78	40.01	48.42	26.98	45.07	26.44
CIA (Ours)	AAAI'26	**67.74**	**40.71**	**79.86**	**42.27**	**49.12**	**27.86**	**46.86**	**27.51**

Table 1: Comparison with state-of-the-art methods on the DN-348 and DN-Wild datasets. All methods are evaluated using rank-1 (%) and mAP (%). GUR* denotes the results without camera information.

Base	DGM	MAA	Day→Night		Night→Day	
			R1	mAP	R1	mAP
✓			63.17	38.60	76.61	40.20
✓	✓		65.32	40.01	79.33	41.80
✓		✓	63.76	39.51	78.22	41.66
✓	✓	✓	**67.74**	**40.71**	**79.86**	**42.27**

Table 2: The effectiveness of DGM and MAA in CIA.

Evaluation

Effectiveness of each component. To assess the contribution of each component in our CIA framework, we conduct ablation studies on the DN-348 dataset. As shown in Table 2, integrating the DGM module into the baseline yields notable improvements in both the day-to-night and night-to-day settings. Similarly, introducing MAA alone also leads to performance gains. Furthermore, combining both DGM and MAA further enhances the results, demonstrating their complementary benefits. These findings confirm that both DGM and MAA individually facilitate cross-domain alignment and, when used together, play a crucial role in boosting the overall performance of our framework.

Comparison with other cross-domain approaches. We validate the superiority of our proposed CIA framework by comparing it to representative cross-modal alignment methods, including PGM (Wu and Ye 2023), CMA (Yang et al. 2022), and MMD-ReID (Jambigi, Rawal, and Chakraborty 2021). As Table 3 shows, under a common baseline on the DN-348 dataset, our CIA framework significantly outperforms each of these methods. We attribute this gain to our joint modeling of cluster-level and instance-level alignments. To offer a more fine-grained analysis, we next evaluate the two core modules of CIA individually:

1) Comparison with cluster-level alignment methods. We compare our Dictionary-Guided Graph Matching (DGM) module against cluster-level alignment methods PGM (Wu and Ye 2023) and CMA (Yang et al. 2022) to assess its effectiveness. Table 3 shows that, in the day-to-night scenario, DGM improves Rank-1 by 0.34% and mAP by 0.36% over PGM, and boosts Rank-1 by 1.8% and mAP by 1.11% over CMA. In the night-to-day scenario, DGM outperforms PGM by 1.81% in Rank-1 and 0.32% in mAP, and exceeds CMA by 2.62% in Rank-1 and 1.17% in mAP. These results demonstrate that explicitly modeling cluster topology via mutual information yields more effective and robust cross-domain alignment than approaches based on static similarity matrices.

2) Comparison with instance-level alignment methods. We further compare MAA against traditional MMD-based methods—MMD-ReID and MarginMMD (Jambigi, Rawal, and Chakraborty 2021), both built on ADCA (Yang et al. 2022)—on the DN-348 dataset to demonstrate the advantages of our approach. Table 3 presents the results. In the day-to-night scenario, MAA outperforms MMD by 1.1% in rank-1 accuracy and 0.54% in mAP, and exceeds MarginMMD by 0.22% in rank-1 and 0.89% in mAP. In the night-to-day scenario, MAA achieves gains of 0.13% and 1.06% over MMD in rank-1 and mAP, respectively, and improves upon MarginMMD by 0.13% in rank-1 and 2.35% in mAP. These results confirm that our MAA method effectively captures inter-instance correlations during alignment, yielding a more self-adaptive and efficient global cross-domain alignment.

Method	Day→Night		Night→Day	
	R1	mAP	R1	mAP
Base	63.17	38.60	76.61	40.20
Base+CMA	63.52	38.90	76.71	40.63
Base+PGM	64.98	39.65	77.52	41.48
Base + DGM	**65.32**	**40.01**	**79.33**	**41.80**
ADCA+MMD	63.22	39.46	78.75	40.84
ADCA+MarginMMD	64.10	39.11	78.75	39.55
ADCA + MAA	**64.32**	**40.00**	**78.88**	**41.90**
Base + CIA	**67.74**	**40.71**	**79.86**	**42.27**

Table 3: Comparison with other cross-modal alignment methods and MMD variants.

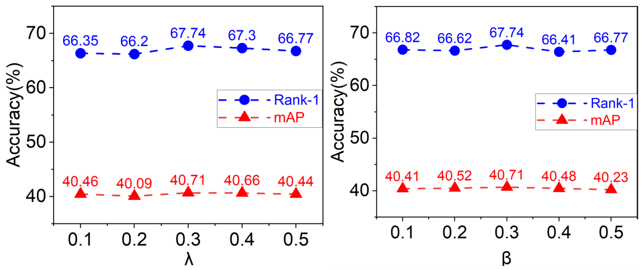

Figure 3: The effects of different hyperparameters λ, and β on the DN-348 dataset. We report the R1 and mAP.

Other analysis

Parameter analysis. We evaluated the influence of key hyperparameters on training accuracy in Fig. 3. First, we varied the loss-balance weights λ and β and found that setting both to 0.3 yielded the best final performance (Rank-1: 67.74%, mAP: 40.71 %). We then set S_τ and α in Eq. (5) to 0.5 and 0.2, respectively, so that similarity scores meeting the 0.5 threshold map into a continuous range via multiplication by 0.2, replacing the original discrete values. For the adaptive threshold in Eq. (6), after exploring a wider range, we determined that the interval between the initial threshold $T_0 = 0.7$ and the final threshold 1.15 is most conducive to model learning. We set w_1 in equation (11) to 0.35 for nighttime and 0.3 for daytime to emphasize more reliable nighttime information.

Scoring Strategies for Dictionaries. To validate the advantage of using a similarity-driven continuous scoring scheme to guide dictionary formation in the DGM matching algorithm, we compare the traditional discrete strategy (Xu et al. 2025)and the direct scoring strategy. The results in Table 4 demonstrate that the continuous scoring scheme delivers superior overall optimization.

Strategy	Day→Night		Night→Day	
	R1	mAP	R1	mAP
+Fixed	66.92	39.89	78.12	42.04
+Discrete(Xu et al. 2025)	66.96	40.41	80.36	42.05
+Continuous	67.74	40.71	79.86	42.27

Table 4: Comparison of different lexicon construction strategies, where Fixed means scoring using fixed values and Discrete means using discrete scores

Effectiveness of generalization. To further validate generalization ability of CIA, we apply it to the RegDB pedestrian re-identification dataset. Table 5 presents the experimental results. Compared to prior pedestrian ReID approaches—OTLA(Wang et al. 2022), NGLR(Cheng et al. 2023b), MBCCM(Cheng et al. 2023a), CCLNet(Chen et al. 2023), PGM(Wu and Ye 2023), GUR*(Yang, Chen, and Ye 2023), SDCL(Yang, Chen, and Ye 2024), and Base(Yang et al. 2022)—Experimental results show that our design can be effectively applied to new domains, which highlights its robustness and adaptability.

Retrieval Results. To validate the effectiveness of CIA, we compared the top ten ranked retrieval results generated by baseline and CIA on the DN-384 dataset and visualized them in Figure 4. The results show that CIA retrieved more true-positive results at higher rankings and significantly reduced the number of false matches.

Settings		V→T		T→V	
Method	Venue	Rank-1	mAP	Rank-1	mAP
OTLA	ECCV'22	32.90	29.70	32.10	28.60
NGLR	MM'23	85.60	76.70	82.90	75.00
MBCCM	MM'23	83.80	77.90	82.80	76.70
CCLNet	MM'23	69.90	65.50	70.20	66.70
PGM	CVPR'23	69.50	65.40	69.90	65.20
GUR*	ICCV'23	73.90	70.20	75.00	69.90
SDCL	CVPR'24	86.91	78.92	85.76	77.25
Base	MM'22	67.20	64.10	68.50	63.80
CIA(Ours)	AAAI'26	86.12	79.34	85.49	77.81

Table 5: Evaluate the generalization capability on the RegDB dataset (Nguyen et al. 2017), against with the state-of-the-art methods.

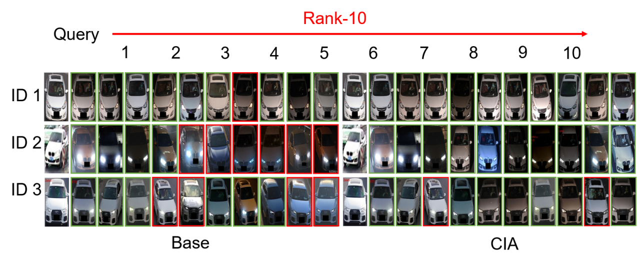

Figure 4: Top-10 retrieval results of the baseline method and our CIA method on the DN-348 dataset. The samples shown in the first column are used as query daytime images. Positive samples are marked with green rectangles, while negative samples are highlighted with red rectangles.

Conclusion

We tackle the problem of unsupervised day–night vehicle reidentification (USL-DN-ReID) by proposing a novel Cluster–Instance Alignment (CIA) framework to address the severe domain shift between daytime and nighttime conditions. At the cluster level, we introduce a dictionary-guided graph matching (DGM) module to model topological correspondences, while at the instance level, we design a multifactor adaptive alignment (MAA) mechanism to perform structure-aware feature adaptation. Together, these components enable more robust and accurate cross-domain alignment. Our CIA framework achieves state-of-the-art performance on two benchmark datasets, and extensive experiments clearly demonstrate its effectiveness. Beyond advancing unsupervised vehicle ReID under extreme illumination variations, this work also offers a generalizable paradigm for a broader range of cross-domain matching tasks.

Acknowledgments

This work is supported by the National Natural Science Foundation of China (No. 62466003), the Guangxi Natural Science Foundation of China (No. 2024GXNSFBA010245, No. 2025GXNSFAA069096), and the Guangxi Science and Technology Base and Talent Special Project (No. AD25069071).

References

Bruff, D. 2005. The assignment problem and the hungarian method. *Notes for Math*.

Chen, Z.; Zhang, Z.; Tan, X.; Qu, Y.; and Xie, Y. 2023. Unveiling the power of clip in unsupervised visible-infrared person re-identification. In *Proceedings of the 31st ACM International Conference on Multimedia*.

Cheng, D.; He, L.; Wang, N.; Zhang, S.; Wang, Z.; and Gao, X. 2023a. Efficient bilateral cross-modality cluster matching for unsupervised visible-infrared person reid. In *Proceedings of the 31st ACM international conference on multimedia*.

Cheng, D.; Huang, X.; Wang, N.; He, L.; Li, Z.; and Gao, X. 2023b. Unsupervised visible-infrared person reid by collaborative learning with neighbor-guided label refinement. In *Proceedings of the 31st ACM international conference on multimedia*.

Dai, Z.; Wang, G.; Yuan, W.; Zhu, S.; and Tan, P. 2022. Cluster contrast for unsupervised person re-identification. In *Proceedings of the Asian conference on computer vision*.

Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X.; et al. 1996. A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*.

He, K.; Zhang, X.; Ren, S.; and Sun, J. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Holla, A.; MM, M. P.; Verma, U.; and Pai, R. M. 2025. Vehicle Re-Identification and Tracking: Algorithmic Approach, Challenges and Future Directions. *IEEE Open Journal of Intelligent Transportation Systems*.

Jambigi, C.; Rawal, R.; and Chakraborty, A. 2021. MMD-ReID: A Simple but Effective solution for Visible-Thermal Person ReID. In *British Machine Vision Conference*.

Kuang, Z.; He, C.; Huang, Y.; Ding, X.; and Li, H. 2023. Joint image and feature levels disentanglement for generalizable vehicle re-identification. *IEEE Transactions on Intelligent Transportation Systems*.

Li, H.; Chen, J.; Zheng, A.; Wu, Y.; and Luo, Y. 2024. Day-Night Cross-domain Vehicle Re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Li, H.; Li, C.; Zheng, A.; Tang, J.; and Luo, B. 2022. Attribute and state guided structural embedding network for vehicle re-identification. *IEEE transactions on image processing*.

Liu, H.; Tian, Y.; Yang, Y.; Pang, L.; and Huang, T. 2016. Deep relative distance learning: Tell the difference between similar vehicles. In *Proceedings of the IEEE conference on computer vision and pattern recognition*.

Lou, Y.; Bai, Y.; Liu, J.; Wang, S.; and Duan, L. 2019. Veri-wild: A large dataset and a new method for vehicle re-identification in the wild. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*.

Lu, Z.; Lin, R.; He, Q.; and Hu, H. 2023. Mask-aware pseudo label denoising for unsupervised vehicle re-identification. *IEEE Transactions on Intelligent Transportation Systems*.

Nguyen, D. T.; Hong, H. G.; Kim, K. W.; and Park, K. R. 2017. Person recognition system based on a combination of body images from visible light and thermal cameras. *Sensors*.

Qiu, M.; Lu, Y.; Li, X.; and Lu, Q. 2024a. Camera-aware Differentiated Clustering with Focal Contrastive Learning for Unsupervised Vehicle Re-Identification. *IEEE Transactions on Circuits and Systems for Video Technology*.

Qiu, M.; Lu, Y.; Li, X.; and Lu, Q. 2024b. Inter-Intra Cluster Reorganization for Unsupervised Vehicle Re-Identification. *IEEE Transactions on Intelligent Transportation Systems*.

Shen, F.; Xie, Y.; Zhu, J.; Zhu, X.; and Zeng, H. 2023. Git: Graph interactive transformer for vehicle re-identification. *IEEE Transactions on Image Processing*.

Taleb, H.; Zhu, X.; and Wang, C. 2024. Unsupervised Vehicle Re-Identification with Clustering and Appearance Style Exploration. In *Proceedings of the 2024 3rd International Conference on Algorithms, Data Mining, and Information Technology*.

Wang, J.; Zhang, Z.; Chen, M.; Zhang, Y.; Wang, C.; Sheng, B.; Qu, Y.; and Xie, Y. 2022. Optimal transport for label-efficient visible-infrared person re-identification. In *European Conference on Computer Vision*. Springer.

Wang, Q.; Zhong, Y.; Min, W.; Zhao, H.; Gai, D.; and Han, Q. 2023. Dual similarity pre-training and domain difference encouragement learning for vehicle re-identification in the wild. *Pattern Recognition*.

Wu, Z.; and Ye, M. 2023. 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*.

Xu, J.; Wang, Q.; Xiong, X.; Gai, D.; Zhou, R.; and Wang, D. 2025. Clip-driven view-aware prompt learning for unsupervised vehicle re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*.

Yang, B.; Chen, J.; and Ye, M. 2023. Towards grand unified representation learning for unsupervised visible-infrared person re-identification. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*.

Yang, B.; Chen, J.; and Ye, M. 2024. Shallow-deep collaborative learning for unsupervised visible-infrared person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*.

Yang, B.; Ye, M.; Chen, J.; and Wu, Z. 2022. Augmented dual-contrastive aggregation learning for unsupervised visible-infrared person re-identification. In *Proceed-*

ings of the 30th ACM International Conference on Multimedia.

Yang, H. F.; Cai, J.; Liu, C.; Ke, R.; and Wang, Y. 2023. Cooperative multi-camera vehicle tracking and traffic surveillance with edge artificial intelligence and representation learning. *Transportation research part C: emerging technologies*.

Yang, Y.; Hu, W.; and Hu, H. 2025. Progressive Cross-modal Association Learning for Unsupervised Visible-Infrared Person Re-Identification. *IEEE Transactions on Information Forensics and Security*.

Ye, M.; Shen, J.; Lin, G.; Xiang, T.; Shao, L.; and Hoi, S. C. 2021. Deep learning for person re-identification: A survey and outlook. *IEEE transactions on pattern analysis and machine intelligence*.

Yu, J.; Oh, H.; Kim, M.; and Kim, J. 2023. Weakly supervised contrastive learning for unsupervised vehicle reidentification. *IEEE transactions on neural networks and learning systems*.

Zhang, X.; Ling, Y.; Li, K.; Shi, W.; and Zhou, Z. 2024. Multimodality Adaptive Transformer and Mutual Learning for Unsupervised Domain Adaptation Vehicle Re-Identification. *IEEE Transactions on Intelligent Transportation Systems*.

Zhong, X.; Zhao, S.; Wang, X.; Jiang, K.; Liu, W.; Huang, W.; and Wang, Z. 2021. Unsupervised vehicle search in the wild: A new benchmark. In *Proceedings of the 29th ACM International Conference on Multimedia*.

Zhong, Z.; Zheng, L.; Kang, G.; Li, S.; and Yang, Y. 2020. Random Erasing Data Augmentation. In *AAAI*.