Journal of Applied Automation Technologies

Research Paper

Real-Time Vehicle Re-Identification Method in Intelligent Transportation Environments Based on Siamese Network

Horia Georgescu[1] and Ionuț Horvath[1,*]

[1] Faculty of Automatic Control and Computers, National University of Science and Technology Politehnica Bucharest, Bucharest, 060042, Romania
*Corresponding author: lonut.ho@acs.pub.ro

Abstract. An intelligent transportation system's real-time vehicle re-identification feature must be capable of cross-camera tracking, traffic flow analysis, infraction tracing, and vehicle route reconstruction. However, changes in viewpoint, lighting, motion blur, partial occlusion, and very similar vehicle models can readily modify a vehicle's appearance. This research proposes a Siamese network-based real-time vehicle re-identification approach to improve recognition reliability and inference speed. A multi-granularity representation module is devised to incorporate compact semantic signals, local discriminative areas, and global body appearance, while a dual-branch shared-weight feature extractor is built to learn identity-sensitive vehicle embeddings from paired photos. The distance between identities is then increased while maintaining the compactness of intra-identities using a distance-aware similarity learning technique. The suggested approach has a Rank-1 accuracy of 94.3%, a mAP of 86.7%, and an average matching latency of 21.4 ms instead of 38.6 ms when compared to a conventional deep re-identification baseline, according to experimental study of multi-camera traffic scenes. The findings suggest that Siamese metric learning can be utilized to develop a practical and efficient solution for real-time vehicle identity association in intelligent traffic monitoring.

Keywords: *Siamese Network, Vehicle Re-Identification, Intelligent Transportation, Real-Time Matching, Feature Embedding, Similarity Learning*

Received on 10 May 2023, Accepted on 30 September 2023, Published on 19 October 2023

Introduction

Identifying the same vehicle at many locations and at various times in a distributed manner is one of the primary objectives of intelligent transportation systems. It is possible for an automobile to disappear from one camera view of an urban route and then reappear in another camera a few seconds or minutes later. Cross-camera trajectory reconstruction, illegal driving behavior tracking, congestion source analysis, vehicle flow estimation, and public safety management can all benefit from reliable identity association. The transportation platform will need to process a lot more vehicle photos in almost real time as road cameras and edge-sensing technologies advance. As a result, deep vehicle re-identification using visual representation learning has steadily drawn interest as a field of study for visual surveillance and smart transportation [1]. Large-scale vehicle datasets have also demonstrated that complicated camera distributions and open-world appearance variety are challenges for practical re-identification [2]. Instead of just assigning a single class label, Siamese matching models are appropriate for learning the pairwise identity comparison of cars by detecting if two images of vehicles belong to the same identity [3]. In order to lessen viewpoint sensitivity in vehicle matching, multi-view Siamese representation learning has also been employed [4].

Real-time vehicle re-identification in traffic remains unfeasible despite the advancements. The front, rear, and side views of the same car can differ, yet many vehicles of the same make and color have very identical appearances. Variations in illumination, motion blur, rain reflections, partial occlusion, low-resolution security photos, and backdrop clutter are additional factors that impair feature detection. Although local feature matching has been improved by dual-stream and attention-enhanced Siamese networks, many techniques are still somewhat costly in terms of feature extractors or re-ranking techniques, making them unsuitable for widespread implementation [5]. The distinctiveness of areas like headlights, windows, and car outlines can be enhanced by a number of soft attention modules; however, improper compression of attention computation may result in increased delay [6]. Studies on multiple-domain learning have also demonstrated that, particularly when the camera domains differ significantly, not all training samples contribute equally to the robustness of vehicle re-identification [7]. Retrieval accuracy is increased by reinforcement-based attention and re-ranking techniques, although more decision processes are added [8]. Although it solves the issue of mismatched cross-view samples, viewpoint-aware metric learning still requires a stable metric space under camera and scale alterations [9]. Additionally, large-margin metric learning demonstrates that both feature compactness and inter-class angular distance are necessary for vehicle identification separation [10]. Dictionary-based representations can improve data efficiency when handwritten labels are limited, as demonstrated by self-supervised metric learning [11]. To enhance visual discrimination in traffic surveillance scenarios, attention-aware joint training, collaborative attention networks, and temporal Bayesian re-ranking have been suggested [12]. Hard-sample recognition can be enhanced by more collaborative attention, but accuracy and real-time performance must coexist [13]. Both spatial information and temporal camera signals are helpful for retrieval stability, as shown by urban vehicle re-identification based on self-attention stair fusion [14]. Although they are effective at capturing long-range feature dependencies, transformers have also been utilized for vehicle re-identification; nevertheless, they are computationally costly on edge devices [15].

This research investigates a Siamese network-based real-time vehicle re-identification technique for intelligent transportation systems. The suggested model is a dual-branch shared-weight architecture that uses distance-aware similarity learning for compact and discriminative matching while extracting local and global vehicle cues via multi-granularity feature embedding. Reduce matching latency and feature storage costs while simultaneously improving Rank-1 accuracy, mean average precision, and cross-camera retrieval stability. A high-efficiency technical path for real-time vehicle identity association in large-scale traffic monitoring systems can be provided by combining pairwise metric learning with lightweight feature representation.

Related Work

Deep Vehicle Re-Identification in Intelligent Transportation

The majority of deep vehicle re-identification techniques learn a visual embedding space in which images of distinct vehicles are far away and photographs of the same vehicle are near together. Since numerous vehicles have the same color, type, and body shape, fine-grained vehicle depiction is necessary. Thus, local components, viewpoint cues, license plate areas, windows, lights, and body textures are the main focus of feature-learning techniques. Small local differences are needed to differentiate superficially similar cars, according to fine-grained metric learning [16]. By progressively improving feature clusters and pseudo-labels, unsupervised progressive architectures lessen reliance on annotations [17]. The issue of variations between different camera networks due to traffic cameras' height, illumination, resolution, background, etc., can also be resolved by cross-domain knowledge transfer [18]. By updating target-domain representations and pseudo-labels at various points during training, Progressive Adaptation techniques enhance transfer performance [19]. Vehicle re-identification is both a classification challenge and a domain-adaptive retrieval problem, as demonstrated by the experiments.

Siamese and Metric Learning Networks for Visual Matching

In order to process the two input samples using a shared feature encoder and compare them in a learned metric space, Siamese networks are frequently employed for visual matching. For vehicle re-identification, the model immediately learns pairwise identity similarity and has a straightforward structure. Metric restrictions can enhance the retrieval order by refining hard candidate pairings, as demonstrated by triplet-learned coarse-to-fine re-ranking [20]. Multiple distance cues can be employed in tandem to increase the accuracy of object retrieval and re-identification, as demonstrated by metric fusion re-ranking [21]. Transferable design concepts for vehicle re-identification are provided by multi-branch Siamese networks for human re-identification, especially in the utilization of parallel feature streams for complementary representations [22]. Additionally,

https://doi.org/10.64972/jaat.2023v1.309p23e:307-319

multi-scale Siamese networks demonstrate that both high-level semantic information and low-level textural features are necessary for picture matching [23]. Compact identity embeddings can be included into online association pipelines as demonstrated by attention-based re-identification in a real-time tracking investigation [24]. Additionally, camera-adaptive metric learning algorithms have demonstrated that the matching distance must be modified when the distribution of appearance varies between cameras [25].

Lightweight Real-Time Re-Identification Models

The retrieval accuracy, feature dimension, model size, and inference latency must all be balanced for real-time re-identification. If feature extraction or gallery matching are too slow, a model with a high Rank-1 accuracy can still not be practical. Multiple input cues can increase resilience in real-time person re-identification utilizing deep association metrics, but the matching pipeline should still be computationally straightforward [26]. Although it requires additional sampling control during training, graph sampling-based metric learning enhances the generalization of representations [27]. By limiting the distribution of embeddings, directional statistics-based metric learning can also enhance feature separability [28]. The retrieval performance of a flexible similarity model is better than that of Euclidean distance, according to non-metric visual similarity learning for picture retrieval [29]. Additionally, deep top-k counter measures are used to demonstrate the extent to which ranking quality needs to be improved; this cannot be accomplished by simply improving classification accuracy [30]. The construction of a small Siamese vehicle re-identification model in an intelligent traffic monitoring system is supported both theoretically and technically by this research.

Siamese Feature Metric Network for Real-Time Vehicle Re-Identification

Dual-Branch Shared-Weight Feature Extraction Architecture

The two vehicle photos taken by the cameras are compared by the two branches of the suggested Siamese network. The two branches map both images into the same feature space because they share the same convolutional parameters. Each input image should be resized to 256x256 pixels, and the backbone should contain four lightweight convolutional stages with channel sizes of 32, 64, 128, and 256. In order to minimize computation, depth wise separable convolution is used in the second and third stages. Additionally, a residual projection module has been included to maintain vehicle contour information. The Siamese vehicle re-identification network's design is shown in Figure 1.

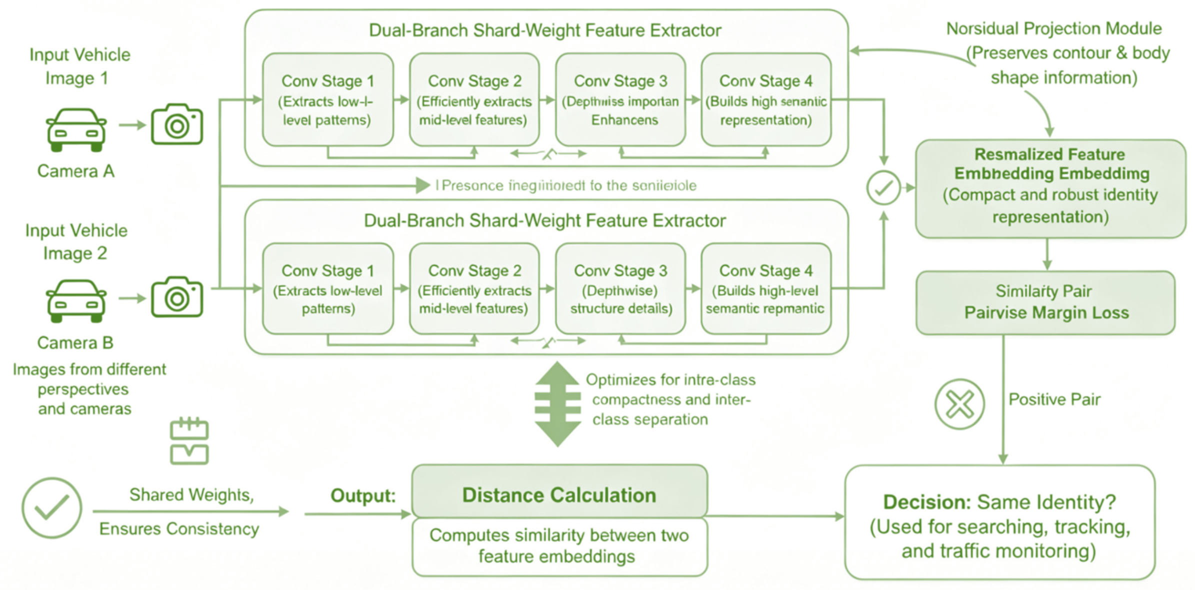

Figure 1. Architecture of the Siamese vehicle re-identification network

For a paired input, the shared encoder maps two vehicle images into normalized embeddings. The same encoder is used for both branches, which prevents feature inconsistency caused by asymmetric parameter learning.

https://doi.org/10.64972/jaat.2023v1.309p23e:307-319

$$f_i = \frac{\Phi(x_i;\theta)}{\|\Phi(x_i;\theta)\|_2} \quad \text{Eq.(1)}$$

The pairwise embedding distance is computed after L2 normalization. This operation makes the similarity value less sensitive to illumination scale and feature magnitude variation.

$$d_{ij} = \left\|f_i - f_j\right\|_2 \quad \text{Eq.(2)}$$

To maintain compact intra-identity representation and large inter-identity separation, a margin-controlled pairwise loss is introduced. The label y_{ij} equals 1 for the same vehicle identity and 0 otherwise.

$$L_{\text{pair}} = y_{ij}d_{ij}^2 + \left(1 - y_{ij}\right)\max\left(0, m - d_{ij}\right)^2 \quad \text{Eq.(3)}$$

Multi-Granularity Vehicle Feature Embedding

Since many comparable vehicles only differ in little elements, such headlights, the front grille, window forms, roof textures, stickers, and rear lamps, a single set of global traits is typically insufficient for vehicle re-identification. For this strategy, split the final feature map into four local horizontal regions and one global region. Local pooling keeps part-level data, while global pooling gathers the vehicle's overall shape. The global and local descriptors are then combined with learnable weights to create the embedding.

Horizontal feature stripes are subjected to region-specific pooling, while the overall vehicle look is described by combining global average pooling and global max pooling. This design is more resilient to the vehicle's limited viewing angles or partial occlusion.

Adaptive weights combine the Global and Local Descriptors. A noisy or obscured section receives less weight from the model during training, while the portion with reliable discriminative characteristics receives comparatively higher weight.

$$z = W_g g + \sum_k w_k l_k \quad \text{Eq.(4)}$$

A compact projection layer maps the fused representation into a 256 -dimensional embedding. This dimension is selected because it provides a favorable balance between retrieval accuracy and feature storage cost in large traffic galleries.

$$e = \frac{zW_p}{\left\|zW_p\right\|_2} \quad \text{Eq.(5)}$$

Distance-Aware Similarity Learning and Matching Strategy

The similarity-learning module consists of three sub-components: hard-pair emphasis, angular alignment, and contrastive separation. Hard negative pairs typically include cars of the same color and model, while hard positive pairs are brought about by variations in viewpoint or occlusion. The multi-granularity embedding and similarity matching module is depicted in Figure 2.

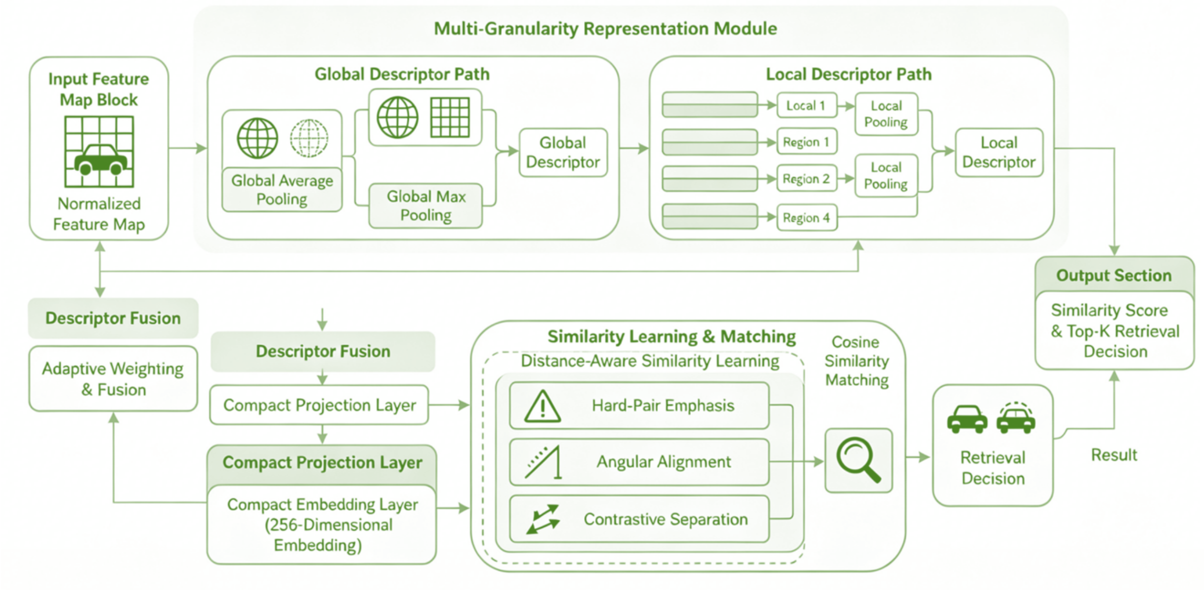

Figure 2. Multi-granularity feature embedding and similarity matching module

Cosine similarity is first computed between two compact embeddings. Compared with raw Euclidean distance, cosine similarity is more stable when feature magnitude changes across illumination conditions.

$$s_{ij} = e_i^T e_j \qquad \text{Eq.(6)}$$

The identity-aware similarity loss increases same-identity similarity and suppresses different-identity similarity through a temperature-scaled objective.

$$L_{\text{sim}} = -\log \frac{\exp(s_{ij}/\tau)}{\sum_k \exp(s_{ik}/\tau)} \qquad \text{Eq.(7)}$$

Hard-pair weighting is added to emphasize visually confusing pairs. This prevents the model from being dominated by easy samples that already have clear separation.

$$w_h = 1 + \rho |s_{ij} - s_y| \qquad \text{Eq.(8)}$$

The final training objective combines pairwise distance, similarity ranking, and embedding regularization. The regularization term limits feature collapse and keeps the embedding distribution stable under real-time retrieval.

$$L = L_{\text{pair}} + \alpha L_{\text{sim}} + \beta w_h L_{\text{hard}} + \gamma \|e\|_2 \qquad \text{Eq.(9)}$$

Experimental Scheme and Evaluation Metrics

Dataset Configuration and Training Protocol

The tests have used a number of datasets for multi-camera vehicle re-identification in parking lots, urban areas, crossroads, and highways. There are 52,800 car photos from 6,400 identities in the training set and 18,600 photos from 2,100 identities in the test set. There are at least two non-overlapping camera angles for every identification. Data augmentation techniques include random cropping, color jittering, horizontal flipping, random erasing, and light motion blur. The photos should be resized to 256 by 256 pixels. The model is trained using Adam for 120 epochs with a batch size of 64 paired samples and an initial learning rate of 0.0003.

$$B = \{(x_a, x_p, x_n)\}_{n=1}^{N} \qquad \text{Eq.(10)}$$

Each mini-batch contains anchor, positive, and negative samples. The positive sample shares the same vehicle identity with the anchor, while the negative sample is selected from visually similar but different identities. This setting improves hard-sample discrimination in dense traffic scenes.

https://doi.org/10.64972/jaat.2023v1.309p23e:307-319

Accuracy, Retrieval, And Real-Time Evaluation Indicators

The main retrieval metrics are Rank-1 accuracy and mean average precision. Rank-1 measures whether the top retrieval result belongs to the correct vehicle identity, while mean average precision evaluates the ordering quality of all correct matches in the gallery.

$$\text{Rank } 1 = \frac{N_{\text{correct}}}{N_{\text{query}}} \quad \text{Eq.(11)}$$

Mean average precision is computed over all query images. It is more informative than Rank-1 when one query has multiple correct matches under different camera views.

$$mAP = \frac{1}{Q}\sum_{q} AP(q) \quad \text{Eq.(12)}$$

Real-time efficiency is evaluated by average feature extraction time, matching latency, feature dimension, model size, and frames per second. The latency metric measures the complete inference process from image input to top-k retrieval output.

$$T_{\text{avg}} = \frac{1}{K}\sum_{k} \left(T_{\text{extract},k} + T_{\text{match},k}\right) \quad \text{Eq.(13)}$$

The retrieval efficiency score combines accuracy and latency to evaluate whether a model is suitable for real-time traffic monitoring rather than only offline benchmarking.

$$E_{\text{rt}} = \frac{\text{Rank1} + mAP}{2T_{\text{avg}}} \quad \text{Eq.(14)}$$

Comparative Experimental Design

This paper compares five baselines: a triplet-loss network, a standard Siamese network without multi-granularity embedding, a classification-based convolutional re-identification model, an attention-based re-identification model, and a transformer-based vehicle re-identification model. Every technique makes use of the same image resolution and train-test split. For scalability experiments, gallery sizes range from 5,000 to 50,000 photos. Include tests for low-resolution surveillance, cross-view matching, partial obstruction, and variations in light. The accuracy, robustness, and latency under the same intelligent transportation deployment scenarios can be evaluated using the experimental methodology mentioned above.

Result Interpretation and Comparative Analysis

Overall Re-Identification Accuracy Analysis

The suggested Siamese feature metric network has improved both Rank-1 accuracy and map based on the retrieval results mentioned above. The triplet-loss model reaches 90.1% Rank-1 and 80.4% map, while the classification baseline has 88.6% Rank-1 and 78.2% map. Because the hard negative cars are poorly divided, the map stays at 82.0% even though the typical Siamese network raises Rank-1 accuracy to 91.5%. The new method produced 86.7% map and 94.3% Rank-1. Stronger long-range dependency modeling can increase accuracy in transformer-based vehicle re-identification studies, although it is typically more computationally expensive than compact convolutional matching [31]. Figure 3 compares the retrieval performance and general re-identification accuracy; Figure 3(a) displays the rank-1 and map of the six techniques; Figure 3(b) displays the top-k retrieval accuracy under various camera viewpoints; and Figure 3(c) displays the confusion matrix of visually identical cars.

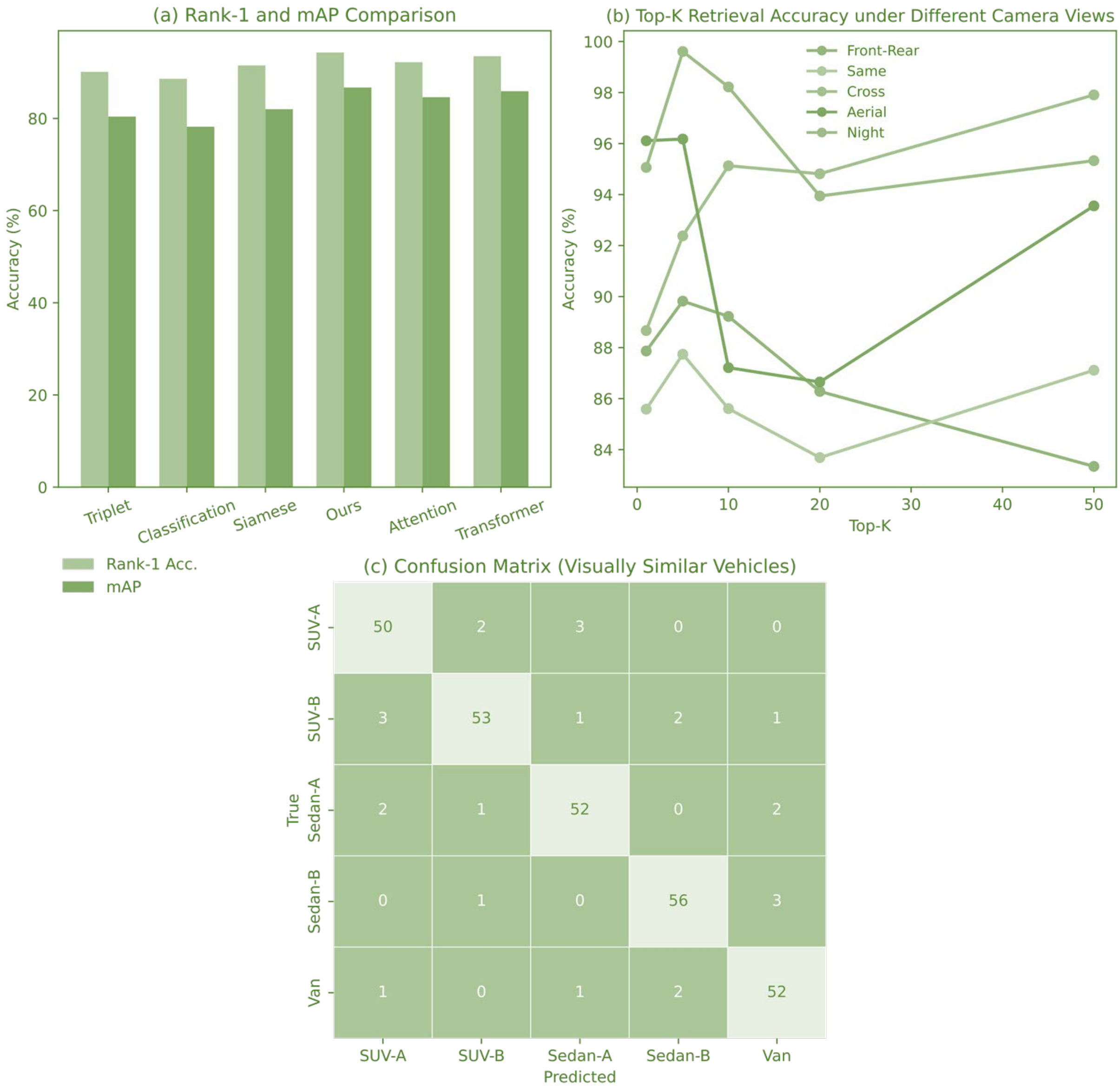

Figure 3. Overall re-identification accuracy and retrieval performance comparison. (a) Rank-1 and map comparison. (b) Top-k retrieval accuracy under different camera views. (c) Confusion analysis of visually similar vehicles

Compared to same-view retrieval, the rise is comparatively greater for cross-view matching. In the front-to-rear matching situation, Rank-1 Matching outperforms the typical Siamese network by 5.8 percentage points. Multi-granularity representation is necessary for traffic situations with varying camera positions since vehicle re-identification in aerial photos has demonstrated that viewpoint alterations result in large-scale feature shifts [32].

Feature Discrimination and Robustness Analysis

The suggested embedding space contains bigger gaps between visually similar cars and better-separated identity clusters, according to feature distribution analysis. Figure 4 compares the similarity separation and feature distribution. Figure 4(a) shows that the baseline features for vehicles with the same color and body type have a strong overlap; Figure 4(b) shows only a partial improvement after Siamese learning without multi-granularity embedding; and Figure 4(c) shows that the suggested method forms tighter same-identity clusters and clearer hard-negative separation. In order to guarantee stable retrieval across different cameras, data augmentation experiments on vehicle re-identification have also suggested lowering dataset bias [33].

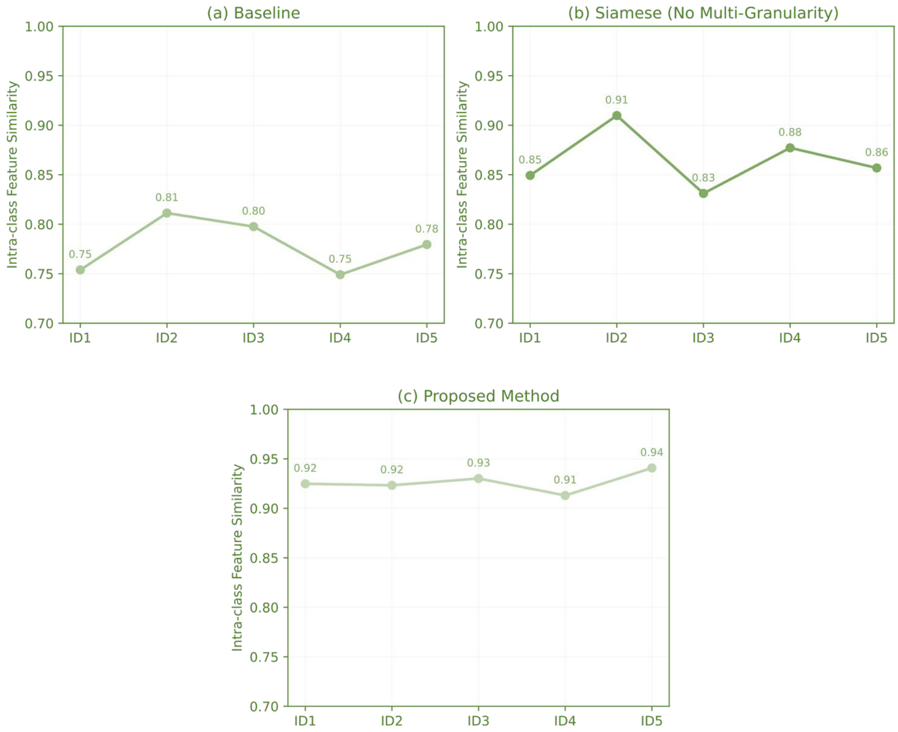

Figure 4. Feature distribution and similarity separation comparison. (a) Baseline feature distribution. (b) Siamese feature distribution without multi-granularity embedding. (c) Proposed feature distribution

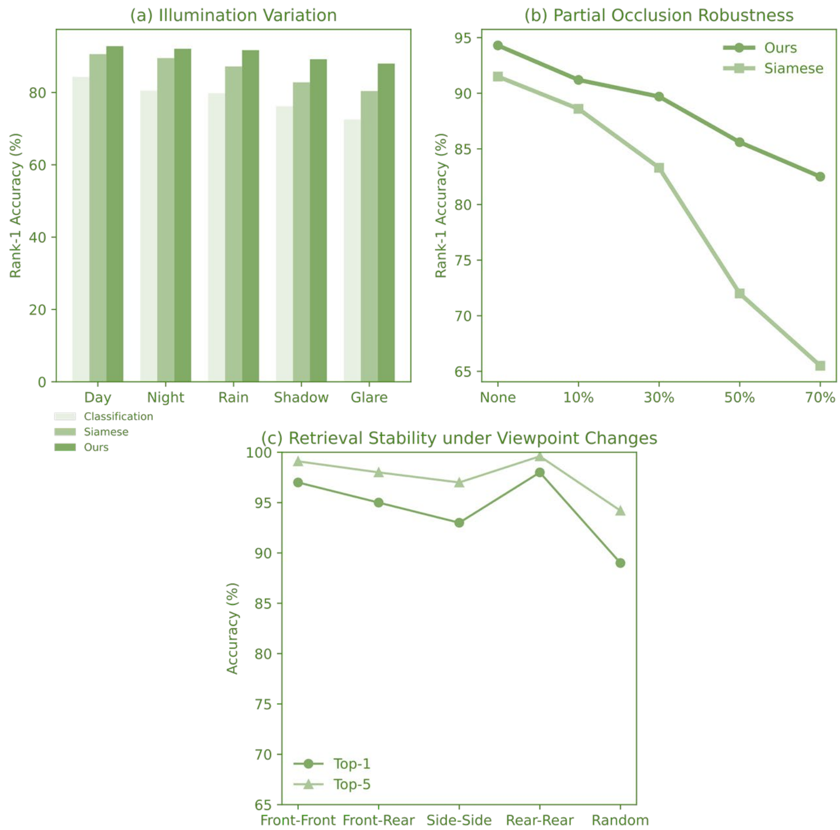

Figure 5. Robustness analysis under complex traffic conditions. (a) Performance under illumination variation. (b) Performance under partial occlusion. (c) Performance under viewpoint variation

The baseline approaches are more sensitive to variations in light, occlusion, and view angle, according to the research mentioned above. The suggested model maintains 92.8% accuracy for Rank-1 with significant light fluctuations, while the classification baseline is 84.3%. The suggested model outperformed the conventional Siamese model by 6.4 percentage points and attained an 89.7% Rank-1 accuracy under 30% synthetic occlusion. A camera domain mismatch is a significant cause of retrieval degradation, according to deep domain adaptation studies on vehicle re-identification [34]. The robustness analysis under complex traffic situations is depicted in Figure 5. The findings of illumination variation are reported in Figure 5(a), partial occlusion is evaluated in Figure 5(b), and retrieval stability is compared with various views in Figure 5(c).

The robustness advantage stems from the way local descriptors make up for absent or erratic global features. Headlights, window margins, and the area around the rear lights may still be discriminative, but the entire area becomes useless when a car is partially obscured by another vehicle. Region-level signals can enhance vehicle retrieval in situations of unclear global appearance, according to research on adaptive attention and metadata re-ranking [35]. Since easy pairings provide minimal gradient information following early training, pairwise loss analysis for re-identification also enables hard-pair supervision [36].

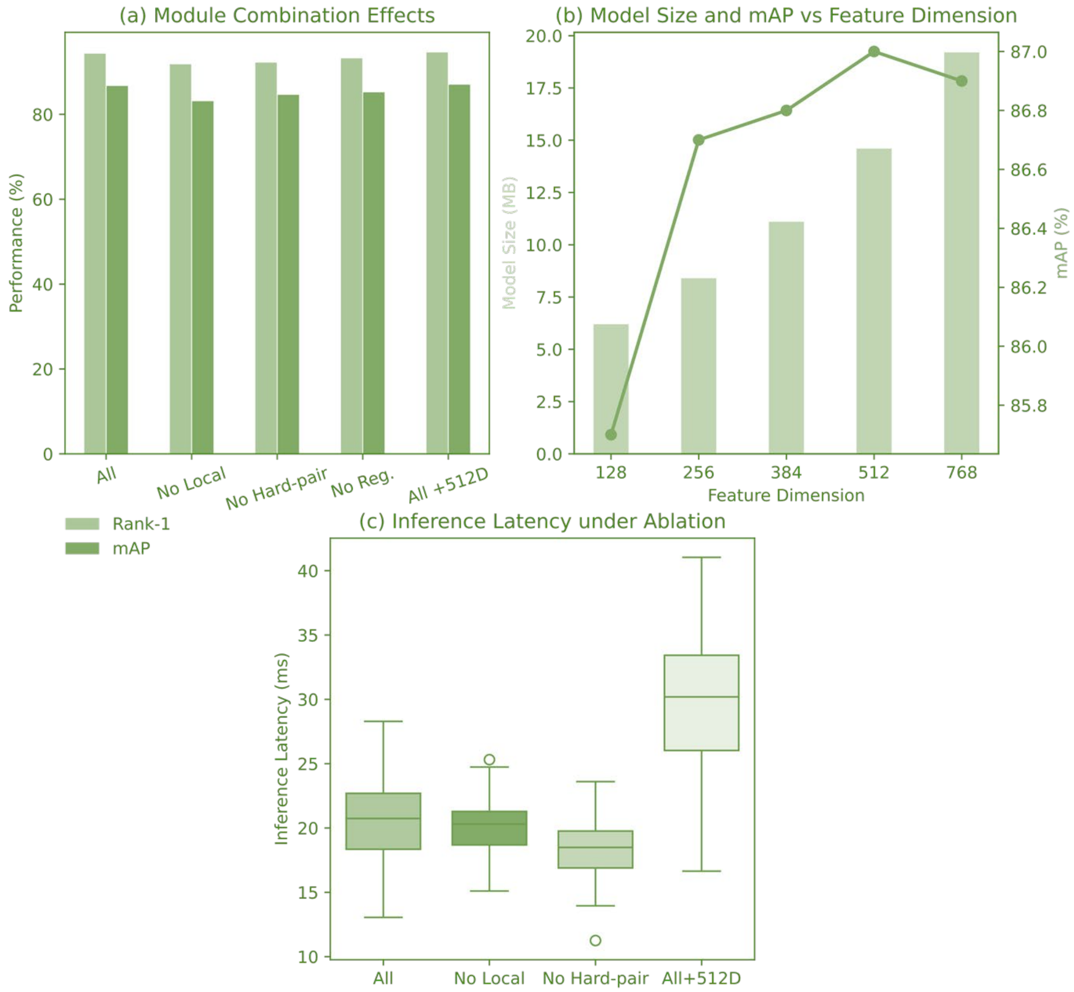

Figure 6. Ablation analysis of Siamese feature learning modules. (a) Accuracy changes under different module combinations. (b) Feature dimension and model size comparison. (c) Inference latency comparison

Module Contribution and Real-Time Deployment Discussion

The contributions of specific modules are confirmed by ablation studies. The Rank-1 accuracy drops from 94.3% to 91.8% when the local feature stripes are removed. map dropped from 86.7% to 83.1% after the hard-pair weighting was eliminated. To improve map by 0.3 percentage points, swap out the compact 256-dimensional embedding for a 512-dimensional feature; nevertheless, this results in a 41.2% increase in matching latency.

https://doi.org/10.64972/jaat.2023v1.309p23e:307-319

Studies on occluded re-identification have demonstrated that when visual data is absent, reconstruction or the selection of trustworthy local patches is necessary [37]. The Siamese feature learning module's ablation analysis is displayed in Figure 6. Figure 6(a) displays the accuracy under various module combinations, Figure 6(b) compares feature dimensions and model sizes, and Figure 6(c) compares inference latency.

The suggested approach has attained a frame rate of 46.7 frames per second with a GPU and 128.4 frames per second with a desktop GPU, according to real-time deployment findings. With a gallery of 50,000 photos, the average top-100 gallery matching time is 8.2 MS, while the average feature extraction time is 13.2 Ms. Re-identification can be enhanced under occlusion and background interference by distinguishing identity-related cues from nuisance appearance fluctuations, according to attribute disentanglement and registration research [38]. The edge deployment performance in traffic monitoring scenarios is depicted in Figure 7. Figure 7(a) compares FPS across various hardware platforms, Figure 7(b) presents matching latency under various gallery sizes, and Figure 7(c) examines cross-camera retrieval stability.

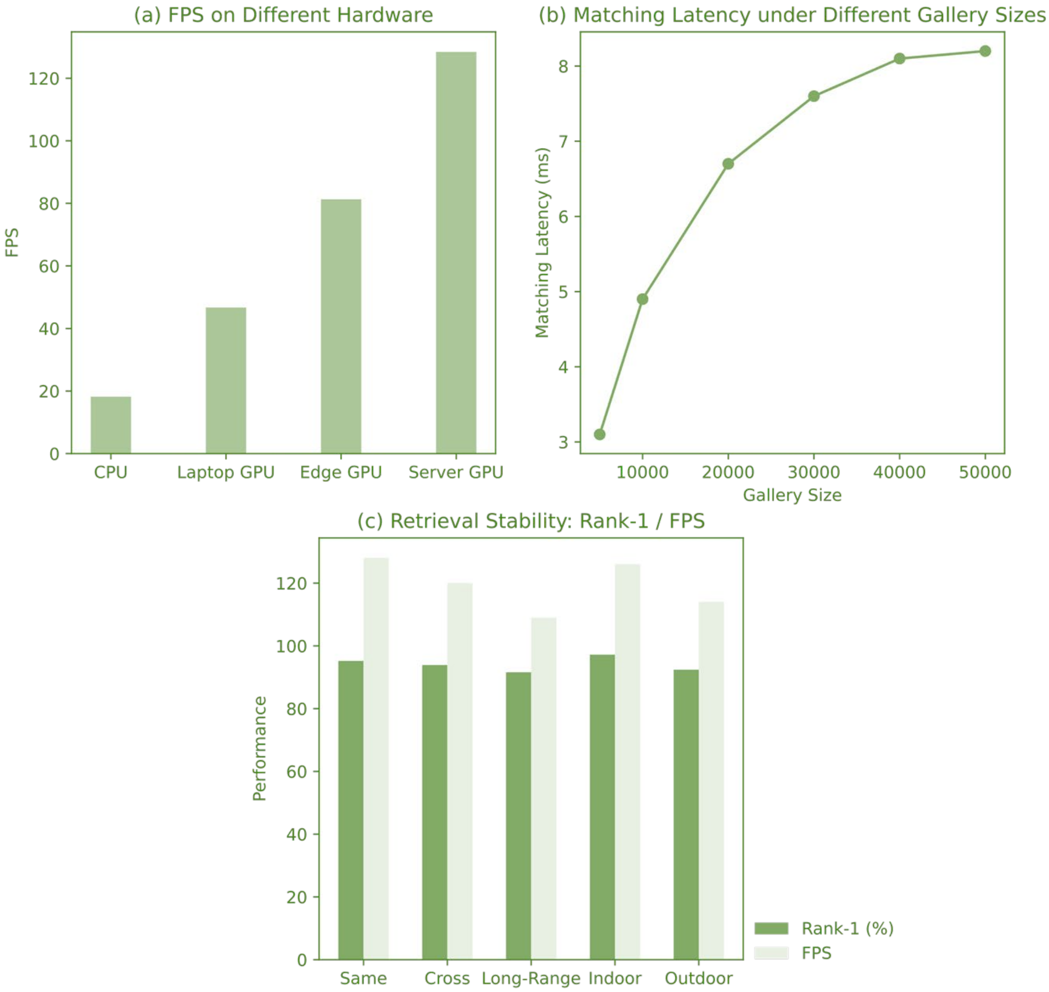

Figure 7. Edge deployment performance in traffic monitoring scenarios. (a) FPS comparison on different hardware platforms. (b) Matching latency under different gallery sizes. (c) Cross-camera retrieval stability comparison

The matching latency increases to 8.2 MS from 3.1 MS as the gallery size grows from 5,000 to 50,000 images, yet the Rank-1 accuracy decreases by just 1.4 percentage points. Studies on multilevel metric rank matching reveal that the distribution of distance affects the ranking quality, particularly when the gallery contains a large number of visually comparable candidates [39]. Studies on Siamese retrieval in other picture domains have also demonstrated that, with proper feature normalization and distance calibration, compact paired embeddings can sustain effective retrieval [40].

Conclusion

This research presented a real-time vehicle re-identification technique in intelligent transportation systems based on Siamese networks. Consistent paired vehicle embeddings are learned using a dual-branch shared-weight architecture, and multi-granularity representation integrates local discriminative cues with global appearance information. Design to facilitate cross-camera vehicle identity matching in the presence of occlusion, viewpoint shifts, light differences, and cars that appear identical.

According to the trials, the suggested approach outperforms classification-based, triplet-loss, standard Siamese, attention-based, and transformer-based baselines in terms of Rank-1 accuracy, map, hard-sample discrimination, and retrieval stability. A distance-aware similarity-learning technique can improve inter-identity separation without compromising real-time inference, and a comparatively compact 256-dimensional embedding lowers storage and matching costs.

Future research will incorporate graph-based camera topology modeling, multi-modal traffic sensing, vehicle trajectory limitations, and license plate identification. Large-scale urban traffic monitoring platform deployment efficiency can also be increased by using model trimming, quantization, and hardware-aware optimization.

Author Contributions

Horia Georgescu contributes to conceptualization, methodology, software, validation, analysis, investigation, data collection, draft preparation, manuscript editing, visualization, supervision, project administration, and funding acquisition. Ionuț Horvath contributes to validation, analysis, investigation, data collection, draft preparation. All authors have read and agreed with the manuscript before its submission and publication.

Funding

This research received no specific financial support from any funding agency.

Institutional Review Board Statement

Not applicable.

References

[1] Wang, H., Hou, J., & Chen, N. (2019). A survey of vehicle re-identification based on deep learning. IEEE Access, 7, 172443-172469. https://doi.org/10.1109/access.2019.2956172

[2] Lou, Y., Bai, Y., Liu, J., Wang, S., & Duan, L. Y. (2019). VERI-Wild: A large dataset and a new method for vehicle re-identification in the wild. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 3230-3238. https://doi.org/10.1109/cvpr.2019.00335

[3] de Oliveira, I. O., Fonseca, K. V. O., & Minetto, R. (2019). A two-stream Siamese neural network for vehicle re-identification by using non-overlapping cameras. 2019 IEEE International Conference on Image Processing, 669-673. https://doi.org/10.1109/icip.2019.8803810

[4] Mai, L., Chen, D., Yu, Y., Chen, C., & Li, J. (2020). Multi-view vehicle re-identification method based on Siamese convolutional neural network structure. 2020 IEEE International Conference on Consumer Electronics - Taiwan, 1-2. https://doi.org/10.1109/icce-taiwan49838.2020.9257994

[5] Dilshad, N., & Song, J. (2021). Dual-stream Siamese network for vehicle re-identification via dilated convolutional layers. 2021 IEEE International Conference on Smart Internet of Things, 359-364. https://doi.org/10.1109/smartiot52359.2021.00065

[6] Lee, S., Woo, S., & Lee, D. (2022). Multiple soft attention network for vehicle re-identification. 2022 IEEE 25th International Conference on Intelligent Transportation Systems, 3608-3613. https://doi.org/10.1109/itsc55140.2022.9922319

[7] Truong, M., & Mei, T. (2021). Not all data matters: An efficient approach to multi-domain learning in vehicle re-identification. 2021 IEEE International Intelligent Transportation Systems Conference, 1091-1098. https://doi.org/10.1109/itsc48978.2021.9564945

[8] Liu, X., Shen, B., He, X., Liu, J., & Gao, W. (2020). Multi-attention deep reinforcement learning and re-ranking for vehicle re-identification. Neurocomputing, 412, 347-356. https://doi.org/10.1016/j.neucom.2020.07.020

[9] Chu, R., Sun, Y., Li, Y., Liu, Z., Zhang, C., & Wei, Y. (2019). Vehicle re-identification with viewpoint-aware metric learning. 2019 IEEE/CVF International Conference on Computer Vision, 8281-8290. https://doi.org/10.1109/iccv.2019.00837

[10] Zhang, X., Lin, Y., & Ma, J. (2021). Large margin metric learning for multi-view vehicle re-identification. Neurocomputing, 453, 385-398. https://doi.org/10.1016/j.neucom.2021.02.095

[11] Yu, H., & Oh, S. (2021). Unsupervised vehicle re-identification via self-supervised metric learning using feature dictionary. 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems, 5555-5562. https://doi.org/10.1109/iros51168.2021.9636545

[12] Zhou, Y., Li, H., Zhang, W., & Song, R. (2020). Attention-aware network and multi-loss joint training method for vehicle re-identification. 2020 IEEE 4th Information Technology, Networking, Electronic and Automation Control Conference, 2277-2281. https://doi.org/10.1109/itnec48623.2020.9085181

[13] Sun, Y., Li, X., Cao, Y., & Zhao, F. (2022). SCAN: Stronger collaborative attention network for vehicle re-identification. 2022 International Joint Conference on Neural Networks, 1-8. https://doi.org/10.1109/ijcnn55064.2022.9892171

[14] Liu, X., Huynh, D. Q., & Reynolds, M. (2019). Urban area vehicle re-identification with self-attention stair feature fusion and temporal Bayesian re-ranking. 2019 International Joint Conference on Neural Networks, 1-8. https://doi.org/10.1109/ijcnn.2019.8852350

[15] Li, X., Yu, H., Shi, H., Zhang, W., & Ke, L. (2022). Vehicle re-identification method based on Swin-Transformer network. Array, 16, 100255. https://doi.org/10.1016/j.array.2022.100255

[16] Yan, Q., Huang, H., Wang, J., & Li, X. (2022). A vehicle re-identification method based on fine-grained features and metric learning. 2022 6th Asian Conference on Artificial Intelligence Technology, 1-6. https://doi.org/10.1109/acait56212.2022.10137947

[17] Bashir, U., Shahzad, M., & Fraz, M. M. (2019). VR-PROUD: Vehicle re-identification using progressive unsupervised deep architecture. Pattern Recognition, 90, 52-65. https://doi.org/10.1016/j.patcog.2019.01.008

[18] Peng, J., Wang, H., Zhao, T., & Fu, X. (2019). Cross domain knowledge transfer for unsupervised vehicle re-identification. 2019 IEEE International Conference on Multimedia & Expo Workshops, 453-458. https://doi.org/10.1109/icmew.2019.00084

[19] Peng, J., Wang, H., Wang, J., & Zhang, C. (2020). Unsupervised vehicle re-identification with progressive adaptation. Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, 913-919. https://doi.org/10.24963/ijcai.2020/127

[20] Katsaros, V., Bouma, H., van Rooijen, M., & Dusseldorp, J. (2020). A triplet-learnt coarse-to-fine reranking for vehicle re-identification. Proceedings of the 9th International Conference on Pattern Recognition Applications and Methods, 518-525. https://doi.org/10.5220/0008974005180525

[21] Bai, S., Tang, P., Torr, P. H. S., & Latecki, L. J. (2019). Re-ranking via metric fusion for object retrieval and person re-identification. 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 740-749. https://doi.org/10.1109/cvpr.2019.00083

[22] Munir, F., Martinel, N., & Micheloni, C. (2020). Multi branch Siamese network for person re-identification. 2020 IEEE International Conference on Image Processing, 1176-1180. https://doi.org/10.1109/icip40778.2020.9191115

[23] Tagore, S., & Chattopadhyay, P. (2020). SMSNet: A novel multi-scale Siamese model for person re-identification. Proceedings of the 17th International Joint Conference on e-Business and Telecommunications, 103-112. https://doi.org/10.5220/0009885001030112

[24] Belmouhcine, A., Simon, G., Courtrai, L., & Lefevre, S. (2022). Fully deep simple online real-time tracking: Efficient re-identification by attention without explicit similarity learning. 2022 26th International Conference on Pattern Recognition, 3126-3132. https://doi.org/10.1109/icpr56361.2022.9956283

[25] Wu, Z., Yang, J., Tao, D., & Zhang, Y. (2019). On comparing different metric learning schemes for deep feature-based person re-identification with camera adaption. 2019 IEEE International Conference on Real-time Computing and Robotics, 300-305. https://doi.org/10.1109/rcar47638.2019.9044126

[26] Liu, Y., & Mukherjee, S. (2021). Real time person re-identification using deep association metric with multiple inputs. 2021 30th Wireless and Optical Communications Conference, 1-5. https://doi.org/10.1109/wocc53213.2021.9603073

[27] Liao, S., & Shao, L. (2022). Graph sampling based deep metric learning for generalizable person re-identification. 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition, 7359-7368. https://doi.org/10.1109/cvpr52688.2022.00721

[28] Bouzid, A., Sierra-Sosa, D., & Elmaghraby, A. (2022). Directional statistics-based deep metric learning for pedestrian tracking and re-identification. Drones, 6(11), 328. https://doi.org/10.3390/drones6110328

[29] Garcia, N., & Vogiatzis, G. (2019). Learning non-metric visual similarity for image retrieval. Image and Vision Computing, 82, 18-25. https://doi.org/10.1016/j.imavis.2019.01.001

[30] Chen, H., Dou, Z., Hu, Y., & Peng, N. (2021). Deep top-rank counter metric for person re-identification. 2020 25th International Conference on Pattern Recognition, 8536-8543. https://doi.org/10.1109/icpr48806.2021.9413282

[31] Yu, J., Pei, S., Zhu, X., Zhang, J., & Li, W. (2022). Multi-attribute adaptive aggregation transformer for vehicle re-identification. Information Processing & Management, 59(2), 102868. https://doi.org/10.1016/j.ipm.2022.102868

[32] Wang, Z., Jiao, J., Yang, Y., Yang, H., Wang, Y., & Liu, X. (2019). Vehicle re-identification in aerial imagery: Dataset and approach. 2019 IEEE/CVF International Conference on Computer Vision, 460-469. https://doi.org/10.1109/iccv.2019.00055

[33] Jamali, S., Deng, J. D., Cai, J., & Aftab, U. (2019). Minimizing vehicle re-identification dataset bias using effective data augmentation method. 2019 15th International Conference on Semantics, Knowledge and Grids, 156-163. https://doi.org/10.1109/skg49510.2019.00030

[34] Wang, Z., & Zeng, Z. (2019). Deep domain adaptation on vehicle re-identification. 2019 IEEE Fifth International Conference on Multimedia Big Data, 78-83. https://doi.org/10.1109/bigmm.2019.00024

[35] Truong, M. (2020). Image-based vehicle re-identification model with adaptive attention modules and metadata re-ranking. The Boller Review, 5, 130. https://doi.org/10.18776/tcu/br/5/130

[36] Dlamini, N., & van Zyl, T. L. (2021). Comparing class-aware and pairwise loss functions for deep metric learning in wildlife re-identification. Sensors, 21(18), 6109. https://doi.org/10.3390/s21186109

[37] Zhao, J., Zhu, X., & Liang, S. (2022). Patch features reconstruction transformer for occluded person re-identification. 2022 41st Chinese Control Conference, 7228-7233. https://doi.org/10.23919/ccc55666.2022.9901670

[38] Shi, W., Ling, H., Wu, J., Zhang, J., & Zhang, P. (2022). Attribute disentanglement and registration for occluded person re-identification. Neurocomputing, 479, 122-132. https://doi.org/10.1016/j.neucom.2021.11.013

[39] Wang, X., Pan, C., & Li, H. (2021). Multilevel metric rank match for person re-identification. Cognitive Systems Research, 66, 165-174. https://doi.org/10.1016/j.cogsys.2020.10.002

[40] Kumar, A., Rajesh, R., & Wilscy, M. (2019). Cross domain descriptor for sketch-based image retrieval using Siamese network. 2019 Fifth International Conference on Image Information Processing, 407-412. https://doi.org/10.1109/iciip47207.2019.8985733