Grenze International Journal of Engineering and Technology, June Issue

# Vehicle Re-Identification Spanning Ground, Aerial, and Cross-View Domains

Amrutha Krishna[1] and H R Chennamma[2]
[1-2]Department of Computer Applications JSS Science and Technology University Mysuru, India
Email: amruthakrishna@jssstuniv.in, hrchennamma@jssstuniv.in

***Abstract*— Vehicle re-identification (Re-ID) is important for intelligent transportation systems and large-scale surveillance. In this paper, we briefly review vehicle Re-ID methods for ground, aerial, and cross-view scenarios. We organise current approaches into appearance-based models, part-aware methods, transformer-based architectures, and cross-view learning strategies. Our comparison shows that existing methods still struggle with large viewpoint changes and domain shifts between aerial and ground images.**

***Index Terms*—Vehicle Re-Identification, Cross-View Matching, UAV Surveillance.**

## I. INTRODUCTION

Vehicle re-identification (Re-ID) is the task of matching the same vehicle across different cameras that do not share overlapping views in large surveillance systems. This technol-ogy has become an important part of intelligent transporta-tion systems. It helps with traffic monitoring, urban security analysis, and tracking vehicle movements. Vehicle Re-ID is more challenging than general object recognition. The main difficulties are that vehicles of the same make and model often look alike, camera angles can vary a lot, lighting conditions can change, views can be blocked, and background clutter.

Early studies mainly looked at ground-based surveillance, which led to the creation of several important bench-mark datasets, such as VeRi-776, VehicleID, VeRi-Wild, and CityFlow [1], [2], [3], [4]. These datasets made it possible to systematically test appearance modelling, metric learning, and spatio-temporal reasoning techniques in environments where camera angles are stable and the surveillance setup is well-organised.

As Unmanned Aerial Vehicles (UAVs) become more com-mon, vehicle re-identification is now being used in more than just ground-based camera setups. Aerial datasets like VRAI, UAV-VeID, DroneVehicle, and UAVDT bring new challenges, such as extreme changes in camera angles, differences in scale, motion blur, and less visual detail [5], [6], [7], [8]. Studies have shown that models trained on ground-based images often perform much worse when used on aerial views. This shows the limits of traditional appearance-based methods.

Recently, aerial and ground vehicle re-identification has be-come a challenging cross-domain problem. The AGID dataset was created to study vehicle matching between UAV images and ground surveillance cameras [9]. In these cases, vehicles can look very different due to changes in viewpoint, scale, and background, creating a large gap between aerial and ground observations.

Because of these challenges, this paper provides a clear overview of vehicle re-identification methods across the ground, aerial, and cross-view domains. We grouped the cur-rent approaches into appearance-based models, part-aware representations, transformer-based architectures, and cross-view learning strategies.

*Grenze ID: 01.GIJET.12.2.37_25*

By comparing key methods and benchmark datasets, we identify the main limitations of current approaches and propose promising directions for future research in cross-view vehicle re-identification.

## II. Vehicle Re-Identification Pipeline

Generally, a vehicle re-identification system uses several steps: vehicle detection, tracking, feature learning, feature ag-gregation, and identity matching, as shown in Fig. 1. First, ve-hicles are found in surveillance images using object detection models. Next, feature-extraction networks process the detected vehicles to learn unique features based on how they appear and their structure. If there are several frames of the same vehicle, combining information over time can make the system more reliable. Finally, the system uses similarity matching and ranking to find and retrieve vehicles. This process works across different camera views. Although deep learning methods have improved a lot, large changes in viewpoint and differences between aerial and ground imagery remain challenging for current vehicle re-identification systems.

## III. Method Taxonomy

As shown in Fig. 2, vehicle re-identification methods can be categorized into feature representation learning, feature aggre-gation, cross-domain generalization, and matching strategies.

### *A. Feature Representation Learning*

Feature representation learning is a key part of vehicle re-identification systems. Its goal is to extract features that help tell apart vehicles that look similar across different cameras. Early methods mainly used global appearance features learned with convolutional neural networks. For example, deep relative distance learning [2] and strong baseline models [10] demon-strated that CNN-based features can capture vehicle color, shape, and texture patterns well.

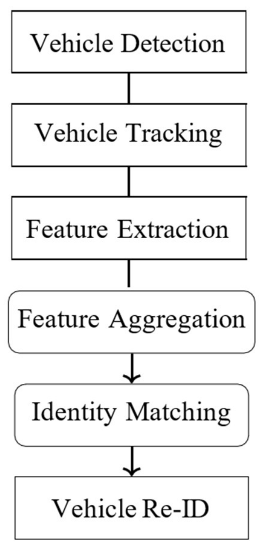



Fig. 1: General pipeline of vehicle re-identification systems.

Global appearance features often struggle to tell apart ve-hicles that look similar. To address this, some researchers use part-based feature learning. Part-regularised models [11] and attention-based alignment networks [12], [13] help by focusing on details such as headlights and logos.

Recent methods use attention-based techniques to highlight the important areas. Viewpoint-aware attention models [14],

[15] adapt to different camera angles, while global and local attention modules, such as GLAMOR [16], combine overall appearance with the detailed features. Parsing-based methods such as PVEN [17] use semantic segmentation of vehicle parts to improve learning. These approaches help the network focus on useful regions instead of just global features.

Attribute-guided learning [18], [19] also helps by using details such as vehicle type or color to make features easier to distinguish.

Recently, researchers have started using transformer-based models for vehicle re-identification. TransReID [20] and other transformer models [21], [22] use self-attention to find long-range relationships within vehicle images. These models work well on large datasets, but they often require substantial training data and computing power.

### *B. Feature Aggregation and Temporal Modeling*

In many surveillance situations, vehicles show up in multi-ple frames or video sequences. Temporal modelling techniques use this information to make the representation more robust. Early research used visual and spatio-temporal path proposals to link vehicles across camera networks [23]. By modelling temporal constraints, these methods reduce false matches caused by vehicles that look similar. Later research explored tracklet-level feature aggregation using multi-camera tracking frameworks [24], [25]. These methods combine features from consecutive frames to get more stable vehicle representations. Recent approaches use contrastive learning on camera tracklets [26], enabling models to learn temporal consistency without requiring explicit identity labels. While temporal aggregation improves robustness against occlusion and motion blur, it relies heavily on reliable tracking results.

TABLE I: COMPARISON OF REPRESENTATIVE VEHICLE RE-ID METHODS

| Method | Category | Strength | Limitation |
|---|---|---|---|
| FACT [2] | Global appearance | Efficient feature learning | Sensitive to viewpoint change |
| PVEN [17] | Parsing-based | Fine-grained modeling | Sensitive to segmentation errors and occlusion |
| GLAMOR [16] | Attention-based | Global-local fusion | High computational cost |
| TransReID [20] | Transformer-based | Context modeling | High computation |

### *C. Cross-Domain Generalization*

Vehicle re-identification systems trained on ground-based surveillance data often do not perform well when applied to aerial images. This drop in performance happens be-cause there are major differences between the viewpoints of aerial and ground cameras. To help study this problem, new datasets like aerial vehicle re-identification benchmarks [27] and aerial–ground matching datasets [9] have been created. To address these differences across domains, domain adaptation methods [28] aim to learn features that generalise across different camera types. Even with these improvements, cross-view vehicle re-identification remains difficult due to large viewpoint changes, scale differences, and insufficient aerial training data.

### *D. Matching and Post-Processing*

Once feature representations are extracted, similarity match-ing techniques are applied to find vehicles across different camera views. Metric learning methods try to make features from the same vehicle as close as possible, while keeping features from different vehicles apart. Re-ranking strategies, such as k-reciprocal encoding [29], improve retrieval by leveraging relationships between feature embeddings. Spatio-temporal constraints [30] and graph-based association methods
[31] also leverage context from surveillance networks to improve matching accuracy. These post-processing steps are very helpful, especially when working with many cameras.

## IV. COMPARATIVE ANALYSIS OF REPRESENTATIVE METHODS

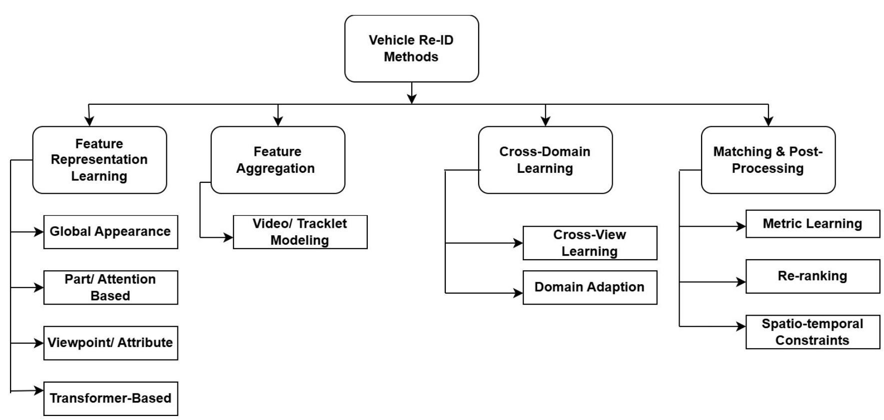
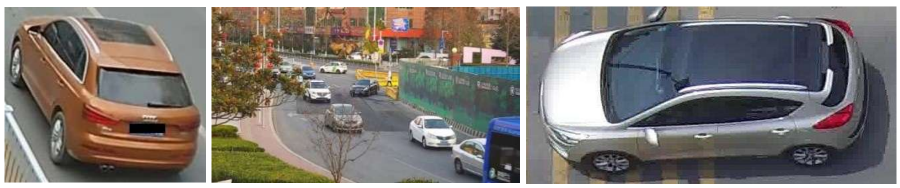



Fig. 2: Taxonomy of vehicle re-identification methods

Table I shows that, how vehicle re-identification research has been steadily improved. Early CNN-based methods mainly learned global appearance features. While these methods performed well on the benchmark datasets, they struggled to distinguish between the similar vehicles and to handle large viewpoint changes. Later methods used part-based and attention-based designs to capture the detailed vehicle parts and connect global and local features. By focusing on the vehicle's key areas, these approaches made it easier to tell vehicles apart. More recently, transformer-based models have improved their feature learning by capturing the long-range relationships in vehicle images. However, these models usually need more training data and greater computing power. Even with these advances, cross-domain vehicle re-identification is still difficult because aerial and ground surveillance images are very different. Researchers have tried domain adaptation and cross-view learning to address this, but reliable cross-view vehicle re-identification has not yet been fully achieved.

## V. BENCHMARK DATASETS AND PERFORMANCE ANALYSIS



(a) VeRi-776 (b) VeRi-Wild (c) VRAI

Fig. 3: Representative samples from widely used vehicle re-identification datasets. (a) VeRi-776 images captured from urban surveillance cameras, (b) VeRi-Wild containing large-scale city surveillance data, (c) VRAI dataset captured from UAV aerial viewpoints.

Fig 3. shows examples from popular vehicle re-identification datasets. Ground-based datasets like VeRi-776 tend to have high recognition accuracy because the viewpoints and imaging conditions are usually consistent. On the other hand, aerial and cross-view datasets such as VRAI and AGID are more difficult due to bigger viewpoint changes, differences in scale, and the gap between UAV and ground images. Researchers use several benchmark datasets to study the vehi-cle re-identification in different surveillance settings. Ground-based datasets, such as VeRi-776, include images from the fixed urban cameras and are standard for testing appearance-based methods. Larger datasets, such as VeRi-Wild, include more vehicle identities and camera angles, which makes large-scale retrieval harder. Aerial datasets like VRAI bring in much greater viewpoint and scale differences because they use images from UAVs. Together, these datasets help test vehicle re-identification algorithms across a wide range of scenarios, from regular surveillance to cross-view aerial settings.

TABLE II: VEHICLE RE-IDENTIFICATION DATASETS AND REPRESENTATIVE BENCHMARK RESULTS REPORTED IN THE LITERATURE

| Dataset | IDs | Images | Description | Method | R1 (%) | mAP (%) |
|---|---|---|---|---|---|---|
| VeRi-776 [1] | 776 | ∼49k | Ground-based urban vehicle dataset | TransReID [20] | 97.1 | 82.3 |
| VehicleID [2] | ∼26k | ∼221k | Large-scale vehicle retrieval dataset | VAMI [15] | 63.1 | – |
| VeRi-Wild [3] | ∼40k | ∼416k | Large-scale city surveillance dataset | FDA-Net [?] | 49.4 | 22.8 |
| CityFlow [4] | 666 | ∼229k | Multi-camera traffic tracking dataset | MoVI-BW [4] | 50.1 | 30.8 |
| VRAI [5] | ∼13k | ∼137k | UAV aerial ve-hicle dataset | UFDN [?] | 82.6 | 83.5 |
| AGID [9] | 834 | 20,785 | Aerial-ground cross-view vehicle dataset | CLIP-ReID + ESFC [9] | 69.0 | 54.4 |

## VI. DISCUSSION AND INSIGHTS

The benchmark results summarised in Table 2 reveal several important trends in vehicle re-identification research. Ground-based datasets, such as VeRi-776, generally achieve higher recognition accuracy because the vehicles are captured from relatively consistent surveillance viewpoints and with stable imaging conditions. As a result, modern deep learning models can effectively learn discriminative appearance features for identity matching.

In contrast, aerial and cross-view datasets such as VRAI and AGID introduce significant challenges due to the large viewpoint variations, scale differences, and changes in object orientation between aerial and ground imagery. These factors often lead to reduced matching performance even for the advanced models.

Large-scale datasets such as VeRi-Wild and CityFlow fur-ther increase the difficulty of the task by introducing thousands of identities and multiple camera views. The increased dataset scale requires models to learn more robust and generalised representations to maintain the retrieval accuracy.

Overall, the comparison indicates that while recent methods such as transformer-based architectures and attention-based feature learning have significantly improved the performance on traditional benchmarks, cross-view vehicle re-identification in aerial surveillance remains an open research problem.

## VII. Challenges

Vehicle re-identification research has advanced a lot, but some challenges still exist, especially in large-scale and cross-view surveillance.

Viewpoint and Scale Variations: Vehicles that are captured from the aerial platforms often show bigger viewpoint differ-ences than those that are captured by ground cameras. These differences cause noticeable changes in appearance, making it harder for standard feature-matching methods to work well.

Visually Similar Vehicles: Many vehicles generally look alike in colour, shape, and structure, especially when they are from the same make or model. This similarity is making it harder to distinguish between the vehicles and also increasing the risk of confusion over their identities.

Cross-Domain Generalization: Models that are trained on ground-based datasets often perform worse when they are applied to aerial images due to the differences in view-point, resolution, and in the environmental conditions. This gap between domains is still a challenge for real-world use. Occlusion and Environmental Conditions: In real-world surveillance, vehicles can be partially hidden by other objects, shadows, or buildings. Changes occur due to lighting, weather, and motion blur, also makes it harder to reliably match the identities.

Large-Scale Retrieval Complexity: Modern datasets have thousands of vehicle identities and millions of images. Finding information quickly and learning scalable representations still remain difficult when we use vehicle re-identification systems in real-world transportation applications.

## VIII. Future Research Directions

Future research in vehicle re-identification should focus on making systems stronger, capable of handling more data, and flexible enough to work in different scenarios.

Geometry-Aware Representation Learning: Using geo-metric information like vehicle orientation, camera viewpoint, and spatial relationships can help narrow the appearance gap between aerial and ground views. Adding depth estimation and bird's-eye-view transformations may also improve cross-view performance. matching.

Cross-Domain and Cross-View Learning: Developing methods for domain adaptation and generalisation is impor-tant for better performance in different surveillance settings. Learning representations that work across domains can also make the models more robust.

Transformer and Foundation Models: Recent transformer-based models are good at understanding the bigger picture. Using large-scale pretraining and vision-language foundation models could further improve vehicle identification.

Multi-Modal and Context-Aware Approaches: Adding other types of data, such as infrared images, motion patterns, or traffic information, can even help systems perform better in tough conditions, like at night or during bad weather.

Real-World Deployment and Privacy Preservation: Fu-ture systems should include ways to protect privacy and use efficient strategies for the large-scale transportation networks.

## IX. Conclusion

This paper provides a brief overview of vehicle re-identification methods for ground-based, aerial-view, and cross-view surveillance. We described the main steps involved in vehicle re-identification systems and grouped existing meth-ods by feature representation learning, feature aggregation, cross-domain generalisation, and matching strategies.

We also reviewed common benchmark datasets and dis-cussed their key performance results from the literature. The comparison shows that the deep learning methods, especially those that use attention and transformer architectures, have made strong progress. Still, identifying vehicles in aerial and cross-view images is difficult due to viewpoint changes, domain differences, and the need to search large datasets.

Future research should focus on geometry-aware represen-tation learning, cross-domain generalisation, and multimodal modelling to improve robustness in real-world intelligent transportation systems.

## References


[1] Xinchen Liu, Wu Liu, Huadong Ma, and Huiyuan Fu. Large-scale vehicle re-identification in urban surveillance videos. In 2016 IEEE International Conference on Multimedia and Expo (ICME), pages 1–6, 2016.

[2] Hongye Liu, Yonghong Tian, Yaowei Wang, Lu Pang, and Tiejun Huang. Deep relative distance learning: Tell the difference between similar vehicles. In 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 2167–2175, 2016.

[3] Yihang Lou, Yan Bai, Jun Liu, Shiqi Wang, and Lingyu Duan. Veri-wild: A large dataset and a new method for vehicle re-identification in the wild. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 3230–3238, 2019.

[4] Zheng Tang, Milind Naphade, Ming-Yu Liu, Xiaodong Yang, Stan Birchfield, Shuo Wang, Ratnesh Kumar, David Anastasiu, and Jenq-Neng Hwang. Cityflow: A city-scale benchmark for multi-target multi-camera vehicle tracking and re-identification. In 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), pages 8789–8798, 2019.

[5] Bingliang Jiao, Lu Yang, Liying Gao, Peng Wang, Shizhou Zhang, and Yanning Zhang. Vehicle re-identification in aerial images and videos: Dataset and approach. IEEE Transactions on Circuits and Systems for Video Technology, 34(3):1586–1603, 2024.

[6] Shangzhi Teng, Shiliang Zhang, Qingming Huang, and Nicu Sebe. Viewpoint and scale consistency reinforcement for uav vehicle re-identification. International Journal of Computer Vision, 129(3):719–735, 2021.

[7] Yiming Sun, Bing Cao, Pengfei Zhu, and Qinghua Hu. Drone-based rgb-infrared cross-modality vehicle detection via uncertainty-aware learning. IEEE Transactions on Circuits and Systems for Video Technology, 32(10):6700–6713, 2022.

[8] Dawei Du, Yuankai Qi, Hongyang Yu, Yifan Yang, Kaiwen Duan, Guorong Li, Weigang Zhang, Qingming Huang, and Qi Tian. The unmanned aerial vehicle benchmark: Object detection and tracking. In Proceedings of the European conference on computer vision (ECCV), pages 370–386, 2018.

[9] Linzhi Shang, Chen Min, Juan Wang, Liang Xiao, Dawei Zhao, and Yiming Nie. Aerial-ground cross-view vehicle re-identification: A benchmark dataset and baseline. Remote Sensing, 17(15), 2025.

[10] Lingxi He, Jian Liang, Hai Li, and Zhenan Sun. A strong baseline for vehicle re-identification. In Proceedings of the IEEE International Conference on Multimedia and Expo (ICME), 2019.

[11] Bingyi He, Yadong Li, and Jian Zhang. Part-regularized near-duplicate vehicle re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2019.

[12] Yichu Liu, Yifan Wang, and Shiliang Zhang. Attentive part-based align-ment network for vehicle re-identification. Electronics, 11(10):1617, 2022.

[13] Yifan Zhao, Hao Liu, and Shiliang Wang. Part-aware attention network for vehicle re-identification. IEEE Transactions on Intelligent Trans-portation Systems, 2020.

[14] Zheng Tang, Milind Naphade, and Ming-Yu Liu. Viewpoint-aware attention for vehicle re-identification. In Proceedings of the European Conference on Computer Vision Workshops (ECCVW), 2018.

[15] Yi Zhou and Ling Shao. Viewpoint-aware attentive multi-view inference for vehicle re-identification. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2018.

[16] Xuejing Liu, Jian Zhang, and Qi Tian. Glamor: Global-local attention module for vehicle re-identification. IEEE Transactions on Image Processing, 2020.

[17] Dechao Meng, Liang Li, Xuejing Liu, Jian Zhang, and Qi Tian. Parsing-based view-aware embedding network for vehicle re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 2020.

[18] Zhenfeng Wang, Zheng Tang, and Milind Naphade. Attributes-guided feature learning for vehicle re-identification. IEEE Signal Processing Letters, 24(10):1570–1574, 2017.

[19] Xin Zhao, Yanhua Li, and Shengjin Chen. Attribute-aware attention model for fine-grained vehicle re-identification. In Proceedings of the IEEE International Conference on Image Processing (ICIP), pages 3925–3929, 2019.

[20] Shuting He, Hao Luo, Pichao Wang, Fan Wang, Hao Li, and Wei Jiang. Transreid: Transformer-based object re-identification, 2021.

[21] Yu Wang, Rui Li, and Yihan Shao. Vehicle re-identification method based on efficient self-attention cnn-transformer and multi-task learning optimization. Sensors, 25(10), 2025.

[22] Sun Wei, Hu Yahua, Dai Guangzhao, Zhang Xiaorui, Xu Fan, and Zhao Yuhuang. Part coupled transformer network for vehicle re-identification. Journal of Computer-Aided Design Computer Graphics, 35(8):1289–1298, 2023.

[23] Yantao Shen, Tong Xiao, Hongsheng Li, Shuai Yi, and Xiaogang Wang. Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals. In Proceedings of the IEEE International Conference on Computer Vision (ICCV), Oct 2017.

[24] Hung-Min Hsu, Jiarui Cai, Yizhou Wang, Jenq-Neng Hwang, and Kwang-Ju Kim. Multi-target multi-camera tracking of vehicles using metadata-aided re-id and trajectory-based camera link model. IEEE Transactions on Image Processing, 30:5198–5210, 2021.
[25] Yun-Lun Li, Hao-Ting Li, and Chen-Kuo Chiang. Multi-camera ve-hicle tracking based on deep tracklet similarity network. Electronics, 11(7):1008, 2022.
[26] Jongmin Yu, Junsik Kim, Minkyung Kim, and Hyeontaek Oh. Camera-tracklet-aware contrastive learning for unsupervised vehicle re-identification. In 2022 International Conference on Robotics and Automation (ICRA), pages 905–911. IEEE, 2022.
[27] Peng Wang, Bingliang Jiao, Lu Yang, Yifei Yang, Shizhou Zhang, Wei Wei, and Yanning Zhang. Vehicle re-identification in aerial imagery: Dataset and approach. In 2019 IEEE/CVF International Conference on Computer Vision (ICCV), pages 460–469, 2019.
[28] Xiao Xiao, Xinyue Yang, Yucheng Wang, Yilong Hui, Jianchuan Zhou, and Guoqiang Mao. Unsupervised domain adaptive vehicle re-identification: A federated learning scheme. IEEE Transactions on Intelligent Transportation Systems, 26(10):14636–14647, 2025.
[29] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pages 3652–3661, 2017.
[30] J. Tu, C. Chen, X. Huang, J. He, and X. Guan. Discriminative feature representation with spatio-temporal cues for vehicle re-identification, 2020.
[31] Fei Shen, Yi Xie, Jianqing Zhu, Xiaobin Zhu, and Huanqiang Zeng. Git: Graph interactive transformer for vehicle re-identification, 2023.