

Comparison and analysis of deep learning models for vehicle re-identification

Liyuan Tang

College of Computer Science and Technology, Donghua University, Shanghai,
201620, China

lytang@mail.dhu.edu.cn

Abstract. The rapid urbanization and increasing number of vehicles on the roads demand efficient and accurate vehicle re-identification (Re-ID) techniques for intelligent transportation systems, traffic monitoring, and surveillance applications. This paper offers a detailed analysis of deep learning approaches to vehicle Re-ID, covering feature learning, attention mechanism, unsupervised learning, self-supervised learning, and specialized loss function. The efficiency of these methods is assessed using VeRi-776 and VehicleID datasets. Key metrics, such as mean Average Precision (mAP) and Rank-n accuracy, are employed to gauge their success. Results show that feature learning, attention mechanism, and specialized loss function play pivotal roles in achieving high performance in vehicle Re-ID tasks, with unsupervised and self-supervised learning methods displaying potential for practical applications due to their scalability. The paper also highlights several challenges, including enhancing the interpretability of attention mechanisms, exploring the relationship between popular loss functions, and addressing the infeasibility of existing methods for real-time applications. The paper concludes with several recommendations for the prospects of vehicle Re-ID, including developing advanced real-time algorithms, enhancing deep learning techniques, investigating innovative approaches, and addressing existing challenges. These proposed advances could significantly improve the efficacy of feature learning and spark fresh innovations in this field.

Keywords: vehicle re-identification, comparative study, deep learning.

1. Introduction

Vehicle re-identification (Re-ID) serves as a pivotal element in intelligent transportation systems, traffic monitoring, and surveillance applications [1, 2]. This task focuses on recognizing a specific vehicle across different cameras and locations, thereby enabling the tracking of vehicles for numerous purposes, including traffic management, crime prevention, and parking enforcement [1-3]. In light of the rapid urbanization and the escalating number of vehicles on the roads, the importance of efficient and accurate vehicle Re-ID cannot be overstated as it forms an essential component of modern transportation systems.

Deep learning, emerging from the broader domain of machine learning, has exhibited exceptional performance in numerous computer vision applications, such as identifying, detecting, and tracking objects [4]. It surfaces as a promising approach for vehicle Re-ID, given its capacity to autonomously learn hierarchical feature representations from large-scale data [4]. Utilizing deep learning techniques, both global and local features of vehicles like color, shape, and distinctive parts crucial for precise Re-

ID can be captured effectively. Furthermore, by leveraging large-scale annotated datasets and complex network architectures, deep learning models can adeptly adapt to various challenges inherent in vehicle Re-ID, including viewpoint variations, occlusions, and illumination changes [5].

The comprehensive comparative study of various deep learning techniques for vehicle Re-ID is presented. These methods are classified into five major themes: feature learning, attention mechanism, unsupervised learning, self-supervised learning, and specialized loss function. The focal point and key contributions of each method are discussed, with an emphasis on their respective strengths and limitations in addressing the vehicle Re-ID issue.

In this paper, for fair comparison, the models are compared on VeRi-776 and VehicleID datasets. The performance of these approaches is measured using performance metrics, including mean Average Precision (mAP) and Rank-n accuracy. The analysis unearths that feature learning and attention mechanism have a vital role in achieving high performance in vehicle Re-ID tasks. Clustering algorithms manifest potential for enhancement, while unsupervised and self-supervised learning methods exhibit moderate performance, hinting at the possibility of further advancement in these techniques. Specialized loss functions emerge as effective means to augment the discriminative power of learned features, ultimately leading to high performance in vehicle Re-ID.

In subsequent chapters, section 2 presents an overview of deep learning-based vehicle Re-ID techniques and corresponding datasets. Section 3 evaluates these techniques through a comparative analysis based on the results from both datasets. Section 4 delves into the strengths and limitations of each approach, touching upon the challenges of vehicle Re-ID. Lastly, Section 5 encapsulates the methodologies examined throughout this paper, addressing the complexities inherent in vehicle Re-ID and suggesting potential avenues for future investigations.

2. Method

2.1. Feature learning

Feature learning methods aim at extracting robust embeddings from vehicle images to improve Re-ID performance. The process of extracting discriminative and robust features is crucial for effective vehicle Re-ID [2]. Numerous approaches have been suggested for extracting both global and local features. Graph Interactive Transformer (GiT) is proposed, which combines graphs and transformers to leverage multiple-level features [2]. Sun et al. designed the Multi-Feature Learning Model with Enhanced Local Attention (MFELA) model to enhance global representation by using multiple levels of features, while incorporating Region Batch Dropblock to learn discriminative local features [1].

Implementing variational representation learning, Alfasly et al. expertly derived discriminative features and employed long short-term memory (LSTM) to grasp the association among multiple vehicle angles of view [6]. In the innovative Multi-Label-Based Similarity Learning (MLSL) approach, Alfasly et al. skillfully employed a Siamese network to analyze three distinct vehicle attributes (ID, color, and type), while utilizing a standard CNN-based feature learner for the vehicle ID attribute representation [7].

To overcome the challenges of restricted training data and efficiently employ multiple public datasets. Zheng et al. presented VehicleNet, an extensive dataset amalgamating various public vehicle datasets [8]. They introduced a two-stage progressive learning model to learn descriptive representations for vehicles. The Viewpoint-Aware Metric Learning method, proposed by Chu et al., studies two metrics at various viewpoints, creating a viewpoint-aware network (VANet) that boosts vehicle Re-ID accuracy [9].

2.2. Attention mechanism

Attention mechanisms enable models to concentrate on crucial aspects of images. A framework incorporating multi-scale mechanisms and attention (MSA) techniques was proposed by Zheng et al. [4]. Jiang et al. offered a Global Reference Attention Network (GRA-Net), which leverages connections between nodes and a global reference node for attention learning [10]. Zheng et al. pioneered the

innovative Dual-relational attention network (DRA-Net), which employs a dual-relational attention module for assessing the feature significance across spatial and channel dimensions [3]. Yang et al. explored the synergy of multi-level features for effective vehicle Re-ID by implementing a two-branch network that leverages pyramid-based local and spatial attention global feature learning (PSA) [11]. Attention mechanisms enhance vehicle Re-ID by directing the model's focus to relevant and informative regions within images.

2.3. *Unsupervised learning*

Unsupervised learning methods tackle the challenges of limited annotations and scalability in real-world scenarios by extracting discriminative features and clustering vehicles without the need for labeled data [5, 12]. Bashir et al. developed VR-PROUD, a method that utilizes unsupervised learning techniques for self-paced progressive learning within a cascaded framework and incorporates contextual information for vehicle ReID [13]. Wei et al. introduced the Transformer-Based Domain-Specific Representation (TDSR) learning framework, which emphasizes domain-specific details for each domain while optimizing feature distribution using Contrastive Clustering Loss [12]. Zheng et al. proposed a novel viewpoint-aware progressive clustering framework (VAPC) model that divides the feature space into subspaces based on predicted viewpoints and performs progressive clustering [5]. The innovative Manifold-based Aggregation Clustering (MAC) algorithm, developed by Zhu and Peng, skillfully addresses unsupervised vehicle Re-ID challenges by accommodating an unknown quantity of clusters [14]. By concentrating on extracting distinct features and clustering vehicles, unsupervised learning methods for vehicle Re-ID improve scalability and practicality in real-world situations.

2.4. *Supervised learning*

Self-supervised learning techniques involve training models using automatically generated labels or supervision signals from the data itself, enabling them to learn meaningful representations without relying on human-annotated labels. Li et al. introduced a groundbreaking approach for vehicle Re-ID, adeptly encoding both local and global representations through self-supervised learning [15]. Utilizing a feature dictionary, Yu and Oh developed a self-supervised metric learning (SSML) approach [16]. Self-supervised learning techniques enhance vehicle Re-ID by reducing the reliance on labeled data and enabling the model to extract inherent structures of data.

2.5. *Loss function*

By employing specialized loss functions in a range of methods, the discriminative power of the features learned is enhanced, addressing specific challenges in the process [17]. Taufique and Savakis introduced Class Balanced Loss in the Local Graph Aggregation Network with Class Balanced Loss (LABNet) to compensate for sample distribution imbalances [17]. Wang et al. used Triplet Center Loss in the Triplet Center Loss based Part-aware Model (TCPM) to emphasize part details and learn discriminating features [18]. Gu et al. devised the Angular Triplet loss method for vehicle ReID, streamlining global representation by integrating a zero-bias batch normalization layer and leveraging cosine metric space for triplet loss computation [19]. Specialized loss functions significantly contribute to vehicle Re-ID by enhancing feature discrimination and tackling problems like imbalanced sample distribution.

2.6. *Dataset*

VeRi-776, extensively employed for vehicle Re-ID, comprises 49,325 samples depicting 776 distinct vehicles. They are recorded by 20 cameras in uncontrolled traffic situations [2, 17]. The dataset, a comprehensive resource, includes a training (37,746 images) and a testing set (11,579 images) [2]. During the validation process, only cross-camera vehicle pairs are considered [2]. The VeRi-776 dataset provides a large-scale evaluation platform for vehicle Re-ID algorithms, as it covers various viewpoints and camera settings [17].

VehicleID is another standard benchmark for vehicle Re-ID, consisting of 221,763 images featuring 26,267 distinct vehicles [2, 17]. It is separated into two distinct sets [9]. The training one comprises

110,178 samples. It covers 13,164 unique vehicles. In contrast, samples for evaluation includes 111,585 images of 13,113 separate vehicles [9]. Unlike VeRi-776, VehicleID only includes front or rear viewpoints of the vehicles [2]. The testing set is subsequently partitioned into three unique parts of diverse magnitudes, which include Small, Medium, and Large [9]. Each subset contains varying numbers of gallery and probe images, allowing for performance evaluation at different data scales [2]. This method is executed till the expected quantity of images is obtained [17].

To gauge the success of vehicle Re-ID algorithms, the two datasets play a pivotal role as benchmarks. By providing diverse data samples and varying evaluation settings, these datasets enable researchers to develop and improve upon existing vehicle Re-ID techniques.

3. Result

In this section, the outcomes of Re-ID techniques using VeRi-776 and VehicleID datasets are showcased. The evaluation of each method's performance is carried out using mAP and Rank-1 on VeRi-776. Additionally, Rank-1 and -5 accuracies are utilized for assessment on the VehicleID dataset. This multi-faceted evaluation process guarantees a robust examination of the performance of each method. The results can be observed within the subsequent pair of tables.

Table 1. mAP and Rank-1 of each model on VeRi-776.

Models		mAP	Rank-1
Feature Learning	GiT [2]	80.34	96.86
	MFELA [1]	81.9	96.3
	Mob.VFL-LSTM [6]	58.08	87.18
	MLSL [7]	61.13	90.04
	VANet [9]	66.34	89.78
Attention Mechanism	MSA [4]	62.89	92.07
	GRA-Net [10]	80.5	95.2
	DRA-Net [3]	80.5	94.7
	PSA [11]	75.9	93.9
Unsupervised Learning	VAPC-DA [5]	40.3	77.4
	MAC [14]	44.29	72.40
	VR-PROUD [13]	22.7	55.7
	TDSR [12]	40.0	86.8
Self-Supervised Learning	SSML [16]	26.7	74.5
Loss Function	LABNet [17]	84.6	97.9
	TCPM [18]	74.59	93.98
	Angular Triplet [19]	78.1	95.9

Among the numerous techniques under consideration, feature learning methods consistently show effectiveness across both datasets. Specifically, GiT and MFELA distinguish themselves in both mAP and Rank-1 accuracy measures. These two methods exhibit an impressive mAP of 80.34% and 81.9%, respectively, with Rank-1 accuracy both surpassing the 96%. They further demonstrate a robust performance on the VehicleID dataset, especially MFELA. It has Rank-1 and -5 accuracy at 85% and 97% respectively in the small dataset category.

Conversely, unsupervised Learning methods display a considerable degree of variability in their results. While TDSR exhibits an exceptional Rank-1 accuracy of 86.8% on the VeRi-776 dataset, this impressive performance does not extend to the VehicleID dataset, where it reaches a mere Rank-1 accuracy of 60% for the small dataset. On both small datasets, the MAC demonstrates underwhelming performance, achieving Rank-1 accuracies of merely 72.4% and 54.27%, respectively.

Table 2. Rank-1 and -5 accuracies of each model on VehicleID.

Methods		Small		Medium		Large	
		Rank-1	Rank-5	Rank-1	Rank-5	Rank-1	Rank-5
Feature Learning	GiT [2]	84.65	-	80.52	-	77.94	-
	MFELA [1]	85.5	97.0	80.2	93.9	78.7	91.8
	Mob.VFL-LSTM [6]	73.37	85.52	69.52	81.00	67.41	78.48
	MLSL [7]	74.21	88.38	69.23	81.48	66.55	78.67
	VANet [9]	83.26	95.97	81.11	94.71	77.21	92.92
Attention Mechanism	MSA [4]	77.55	90.50	74.41	86.26	72.91	84.35
	GRA-Net [10]	83.3	96.2	78.2	93.5	75.1	90.8
	DRA-Net [3]	82.8	96.2	78.1	93.2	75.3	90.9
	PSA [11]	80.4	94.7	79.5	92.1	76.3	88.9
Un-Sup Learning	VAPC-DA [5]	75.3	89.0	69.0	85.5	61.0	79.7
	MAC [14]	54.27	71.10	47.49	66.83	44.37	65.85
	TDSR [12]	60.0	73.0	56.1	71.6	52.0	68.5
Self-Sup Learning	SSML [16]	49.6	71.0	43.9	64.9	34.7	55.4
Loss Function	LABNet [17]	84.0	-	80.2	-	77.2	-
	TCPM [18]	81.96	96.38	78.82	94.29	74.58	90.71
	Angular	-	-	-	-	77.3	91.9
	Triplet [19]	-	-	-	-	-	-

Regarding the attention mechanism category, GRA-Net and DRA-Net emerge as formidable contenders, exhibiting virtually identical performance metrics. They both attain mAP of 80.5% on the VeRi-776 dataset, while their Rank-1 accuracies stand at 95.2% and 94.7%, respectively. Their robust performance reflects on VehicleID, with Rank-1 accuracies above 82% surpassing 96% on the small dataset.

Self-supervised learning methods, epitomized by SSML, demonstrate a somewhat limited degree of success. It garners mAP of 26.7% and Rank-1 accuracy at 74.5%, while achieving Rank-1 accuracy of 49.6% in the small dataset category of the VehicleID dataset.

In contrast, the loss function category displays some of the highest performance metrics. LABNet excels with mAP of 84.6% and Rank-1 accuracy of 97.9%. On the VehicleID dataset, LABNet shows Rank-1 accuracies above 77% across all dataset sizes.

To conclude, it is crucial to note that while feature learning and loss function categories generally display superior performance on the studied datasets, there are considerable intra-category variations. Therefore, a comprehensive grasp of task-specific needs and careful evaluation of the dataset is crucial for choosing the ideal vehicle Re-ID method.

4. Discussion

Through a comparative analysis of multiple techniques applied to two datasets, this section evaluates the strengths and limitations of each method. In addition, the challenges of vehicle Re-ID are discussed.

Feature learning methods, such as GiT and MFELA, exhibit commendable performance in vehicle Re-ID tasks by deriving robust features from vehicle images. Nevertheless, difficulties may arise when handling local features and contending with the high cost of manual part annotations [1, 2]. The efficiency of attention mechanisms, specifically GRA-Net and DRA-Net, is demonstrated through their ability to concentrate the model's attention on relevant and information-rich parts of images, consequently boosting vehicle Re-ID outcomes. However, the interpretability of these mechanisms and their capacity to unveil vital information for recognition persist as challenges [1, 15].

The benefits of scalability and the capacity to learn from inherent data structures without human-annotated labels are offered by unsupervised and self-supervised learning techniques such as VAPC-DA and SSML. However, these methods generally display subpar performance when compared to feature learning and attention mechanism due to obstacles posed by limited annotations and dependency on automatically generated labels or supervision signals from the data itself [13].

By leveraging specialized loss functions like Class Balanced Loss in LABNet and Triplet Center Loss in TCPM, the discriminative prowess of the learned features is amplified, effectively handling issues such as imbalanced sample distribution. Exploring the link between common loss functions, including softmax loss as well as triplet loss for vehicle Re-ID remains an open field of research [18].

In terms of practical applications, a significant challenge with existing methods is their infeasibility for real-time applications due to substantial time and space requirements, and the discrimination capabilities of most existing representations are far from ideal. Additionally, overcoming the obstacle of learning stable visual cues in the face of intra-class pose, illumination, and partial occlusion changes remains a daunting task.

5. Conclusion

In this paper, a thorough comparative evaluation of various deep learning methodologies employed for vehicle Re-ID is articulated, encompassing areas such as feature learning, attention mechanism, unsupervised learning, self-supervised learning, and specialized loss function. The comprehensive study on VeRi-776 and VehicleID datasets highlights the importance of feature learning, attention mechanism, and specialized loss function in attaining exceptional performance in vehicle Re-ID tasks. Amongst the methods studied, LABNet, leveraging Class Balanced Loss, displayed remarkable competence in managing uneven sample distributions. Moreover, methods like GiT and MFELA proved adept at extracting resilient and discriminative attributes from vehicle images.

Nonetheless, unsupervised and self-supervised learning methods such as VAPC-DA and SSML, despite their marginally less competitive performance, show potential in pragmatic applications due to their scalability and capacity to discern the intrinsic structure of data, independent of human-annotated labels. It is recognized, however, that these methods require augmentation, particularly with respect to their performance and their applicability to real-world situations.

The paper further illuminates the challenges inherent in this field. These encompass the task of enhancing the explicability of attention mechanisms, investigating the relationship between prevalent loss functions for vehicle Re-ID, and addressing the impracticability of current methodologies for real-time applications given their significant time and space requisites. In addition, the problem of learning visual cues that remain invariant despite variations in intra-class pose, illumination, and partial occlusion continues to pose a significant challenge.

Therefore, future inquiries ought to emphasize the creation of efficient algorithms suitable for real-time applications, the increased application of deep learning for superior discriminative potential, the exploration of innovative loss functions for enhanced robustness, and the broadening of unsupervised methodologies for improved capitalization of unlabeled data. Moreover, integrating attention mechanisms with other methods could offer insights into the decision-making processes. Lastly, surmounting the challenges of local feature management and manual part annotations through automated and trustworthy annotation procedures could lead to significant advancements in the effectiveness of feature learning in vehicle Re-ID systems.

References

- [1] Sun, W., Chen, X., Zhang, X. R., Dai, G. Z., Chang, P. S., & He, X. (2021). A multi-feature learning model with enhanced local attention for vehicle re-identification. *Computers, Materials & Continua*, 69(3), 3549-3560.
- [2] Shen, F., Xie, Y., Zhu, J., Zhu, X., & Zeng, H. (2023). Git: Graph interactive transformer for vehicle re-identification. *IEEE Transactions on Image Processing*, 32, 1039-1051.

- [3] Zheng, Y., Pang, X., Jiang, G., Tian, X., & Meng, Q. (2023). Dual-relational attention network for vehicle re-identification. *Applied Intelligence*, 53(7), 7776-7787.
- [4] Zheng, A., Lin, X., Dong, J., Wang, W., Tang, J., & Luo, B. (2020). Multi-scale attention vehicle re-identification. *Neural Computing and Applications*, 32, 17489-17503.
- [5] Zheng, A., Sun, X., Li, C., & Tang, J. (2021). Viewpoint-aware progressive clustering for unsupervised vehicle re-identification. *IEEE Transactions on Intelligent Transportation Systems*, 23(8), 11422-11435.
- [6] Alfasly, S. A. S., Hu, Y., Liang, T., Jin, X., Zhao, Q., & Liu, B. (2019). Variational representation learning for vehicle re-identification. In 2019 IEEE International Conference on Image Processing, 3118-3122.
- [7] Alfasly, S., Hu, Y., Li, H., Liang, T., Jin, X., Liu, B., & Zhao, Q. (2019). Multi-label-based similarity learning for vehicle re-identification. *IEEE Access*, 7, 162605-162616.
- [8] Zheng, Z., Ruan, T., Wei, Y., Yang, Y., & Mei, T. (2020). VehicleNet: Learning robust visual representation for vehicle re-identification. *IEEE Transactions on Multimedia*, 23, 2683-2693.
- [9] Chu, R., Sun, Y., Li, Y., Liu, Z., Zhang, C., & Wei, Y. (2019). Vehicle re-identification with viewpoint-aware metric learning. In Proceedings of the IEEE/CVF international conference on computer vision, 8282-8291.
- [10] Jiang, G., Pang, X., Tian, X., Zheng, Y., & Meng, Q. (2022). Global reference attention network for vehicle re-identification. *Applied Intelligence*, 1-16.
- [11] Yang, J., Xing, D., Hu, Z., & Yao, T. (2021). A two - branch network with pyramid - based local and spatial attention global feature learning for vehicle re - identification. *CAAI Transactions on Intelligence Technology*, 6(1), 46-54.
- [12] Wei, R., Gu, J., He, S., & Jiang, W. (2022). Transformer-Based Domain-Specific Representation for Unsupervised Domain Adaptive Vehicle Re-Identification. *IEEE Transactions on Intelligent Transportation Systems*, 24(3), 2935-2946.
- [13] Bashir, R. M. S., Shahzad, M., & Fraz, M. M. (2019). Vr-proud: Vehicle re-identification using progressive unsupervised deep architecture. *Pattern Recognition*, 90, 52-65.
- [14] Zhu, W., & Peng, B. (2022). Manifold-based aggregation clustering for unsupervised vehicle re-identification. *Knowledge-Based Systems*, 235, 107624.
- [15] Li, M., Huang, X., & Zhang, Z. (2021). Self-supervised geometric features discovery via interpretable attention for vehicle re-identification and beyond. In Proceedings of the IEEE/CVF International Conference on Computer Vision, 194-204.
- [16] Yu, J., & Oh, H. (2021). Unsupervised vehicle re-identification via self-supervised metric learning using feature dictionary. In 2021 IEEE/RSJ International Conference on Intelligent Robots and Systems, 3806-3813.
- [17] Taufique, A. M. N., & Savakis, A. (2021). LABNet: Local graph aggregation network with class balanced loss for vehicle re-identification. *Neurocomputing*, 463, 122-132.
- [18] Wang, H., Peng, J., Jiang, G., Xu, F., & Fu, X. (2021). Discriminative feature and dictionary learning with part-aware model for vehicle re-identification. *Neurocomputing*, 438, 55-62.
- [19] Gu, J., Jiang, W., Luo, H., & Yu, H. (2021). An efficient global representation constrained by Angular Triplet loss for vehicle re-identification. *Pattern Analysis and Applications*, 24, 367-379.