

Doi:10.32604/cmc.2025.062950

ARTICLE



Tech Science Press

AG-GCN: Vehicle Re-Identification Based on Attention-Guided Graph Convolutional Network

Ya-Jie Sun¹, Li-Wei Qiao¹ and Sai Ji^{1,2,*}

¹School of Computer Science, Nanjing University of Information Science & Technology, Nanjing, 210044, China
²College of Information Engineering, Taizhou University, Taizhou, 225300, China

*Corresponding Author: Sai Ji. Email: jisai@nuist.edu.cn

Received: 31 December 2024; Accepted: 21 March 2025; Published: 09 June 2025

ABSTRACT: Vehicle re-identification involves matching images of vehicles across varying camera views. The diversity of camera locations along different roadways leads to significant intra-class variation and only minimal inter-class similarity in the collected vehicle images, which increases the complexity of re-identification tasks. To tackle these challenges, this study proposes AG-GCN (Attention-Guided Graph Convolutional Network), a novel framework integrating several pivotal components. Initially, AG-GCN embeds a lightweight attention module within the ResNet-50 structure to learn feature weights automatically, thereby improving the representation of vehicle features globally by highlighting salient features and suppressing extraneous ones. Moreover, AG-GCN adopts a graph-based structure to encapsulate deep local features. A graph convolutional network then amalgamates these features to understand the relationships among vehicle-related characteristics. Subsequently, we amalgamate feature maps from both the attention and graph-based branches for a more comprehensive representation of vehicle re-identification. Comprehensive qualitative and quantitative analyses on two publicly available datasets verify the efficacy of AG-GCN in addressing intra-class and inter-class variability issues.

KEYWORDS: Vehicle re-identification; a lightweight attention module; global features; local features; graph convolution network

1 Introduction

Vehicle re-identification technology is employed to identify and monitor vehicles across surveillance camera networks. This technology holds promise to assist in resolving illicit activities, such as combating license plate forgery and investigating hit-and-run cases, while also providing valuable data for urban traffic management and planning. It constitutes an essential instrument for improving community services and upholding public safety. With the growing interest in vehicle re-identification, several datasets have been curated for advanced research. The VeRi-776 [1] and VehicleID [2] datasets, introduced in 2016, are pivotal for research purposes, offering a wide range of vehicle images that enable comprehensive testing and enhancement of re-identification techniques. Although progress has been made with deep learning approaches for re-identification, challenges such as vehicle occlusions, diverse viewpoints, and variations in lighting conditions still present substantial obstacles, as illustrated in Fig. 1. To overcome these challenges and refine identification precision, researchers are actively seeking and devising innovative solutions.





Figure 1: Difficulties in vehicle re-identification. (a) intra-class difference; (b) inter-class similarity

To confront the challenges identified in the realm of vehicle re-identification, the field has been systematically segmented into two evolutionary phases. Foundational investigations [3] enhanced established attributes such as color and texture, advancing re-identification methodologies. Despite inherent constraints, these pioneering efforts established the essential theoretical framework and catalyzed subsequent advancements. The variable nature of vehicle appearances across divergent viewpoints hinders the exclusive dependence on global attributes. In response, Wang et al. [4] delineated local features from strategic points to construct resilient, orientation-neutral representations, integrating a log-normal distribution to mitigate temporal and spatial discrepancies and heighten the precision of retrieval outcomes. Progress in the analysis of local characteristics has been pivotal for contemporary vehicle Re-ID schemas. He et al. [5] accentuated the efficacy of models that conserve distinctive local attributes, thereby securing their pivotal role in the identification procedure. Sun et al. [6] unveiled the MFELA, an avant-garde vehicle Re-ID model that skillfully amalgamates local and global feature assimilation, enhancing accuracy and focal regional discernment, as validated by benchmark analyses. These methodologies, by encoding intricate structural data of vehicles, furnish improved exactness. Nevertheless, prevalent techniques often rely on rudimentary feature concatenation, neglecting the intricate, inherent correlations among vehicular attributes. To thoroughly scrutinize these complex affinities and amplify re-identification precision, Zhu et al. [7] contrived the Structured Graph Attention Network (SGAN). This network, by meticulously leveraging graph structures, offers accurate delineation and distinction of subtle disparities among various vehicle models or makes. Exploration of attention mechanisms has permeated vehicle re-identification inquiries, albeit frequently to the detriment of computational efficiency [8,9]. The quest for proficient attention frameworks with minimal parameters and reduced operational demands persists as a critical endeavor. It is noteworthy that the present study incorporates the Effective Channel Attention (ECA) module [10], which curtails computational overheads whilst securing pertinent information extraction.

Our study introduces a novel dual-branch architecture, ingeniously merging graph neural networks with a streamlined attention mechanism, to enhance the vehicle re-identification paradigm. This unified framework simplifies the complexity of the process by leveraging the detailed structure analysis of graphbased methods while employing a lightweight attention scheme to effectively pinpoint critical features. The result is a more compact, yet powerful model that offers both efficiency and accuracy in addressing vehicle re-identification challenges. Given the considerations mentioned, we introduce an Attention-Guided Graph Convolutional Network (AG-GCN) to learn the relationship between local features and then combine global features to realize vehicle re-identification. Initially, we introduce effective channel attention (ECA) into the ResNet50 backbone to enhance its ability to capture global vehicle features during training. To address the issue of neglecting subtle features in the global branch, we propose the use of a local branch, which employs GCN to transmit local information between different parts and better capture the vehicle's relationships. By combining the information from both branches, the AG-GCN generates an optimal vehicle feature representation. To validate the effectiveness of our method, we conduct experiments on standard large-scale datasets and compare our results with those of other representative methods.

To sum up, the contribution includes the following three aspects:

We propose a global and local dual-branch network, which simultaneously learns global features and local fine-grained features.

We optimized the ResNet50 network by adding the ECA module which only has a few parameters, bringing significant performance gain.

We adopt a GCN in the local branch to learn structured information about vehicle features.

Section 2 reviews the work related to this paper. Section 3 introduces the proposed reidentification model in detail. Section 4 makes experimental comparisons and analyses. Lastly, we summarize the paper.

2 Related Works

In this section, his paper introduces three related works: vehicle re-identification, attention mechanisms and graph convolutional network.

2.1 Vehicle Re-Identification

The core task of vehicle re-identification (Re-ID) is to match the query vehicle with the corresponding vehicle image in the database, so as to accurately determine the identity (ID) of the target vehicle. This technology not only simplifies the search process and improves the search efficiency, but also has become an important part of the automated management of intelligent transportation systems. However, vehicle re-identification still faces many challenges in terms of missing features, local information utilization, and data consistency modeling due to a variety of factors such as camera view angle changes, uneven illumination, and occlusion.

In recent years, deep neural networks have made significant breakthroughs in visual recognition tasks. Compared with traditional manual feature extraction methods, deep learning has obvious advantages in suppressing visual noise and improving model robustness. Early vehicle Re-ID research mainly relied on global feature extraction, such as the DRDL method proposed by Liu et al. [2], which enhances the differentiation ability by combining the vehicle ID and vehicle model supervisory information through a two-branch network. However, relying on global features alone often makes it difficult to distinguish vehicles with similar appearances, which leads to the degradation of the recognition performance, and therefore some researchers have started to explore local feature fusion strategies, as exemplified by the mesh space transformation network designed by Ma et al. [3], which extracts local features through the top, middle, and bottom branches, and strengthens the expression of details by combining with the attention mechanism to enhance the accuracy of the Re-ID. The accuracy of Re-ID is improved.

In the vehicle re-recognition task, multi-view information complementation and data consistency are the keys to improve the recognition performance. For example, the BWIC-TIMC method proposed by Yao et al. [11] utilizes inter-view and intra-view information to complete the missing viewpoint data, and introduces a dual tensor constraint module in the tensor space to achieve higher-order graph completion, which provides a new way to solve the problem of missing features caused by changes in viewpoints. Meanwhile, the MIMB method proposed by Wang et al. [12] emphasizes on maintaining the consistency with the original data during the data recovery process, and improves the clustering performance through the stream embedding and inverse regularization strategies, which effectively reduces the noise and outlier interference in the multi-view matching and feature learning Due to the large differences in the viewing angles of vehicle images captured by different cameras, the lack of effective data consistency modeling can easily lead to matching errors. Based on the strategy of MIMB in multi-view clustering, vehicle Re-ID can also optimize feature learning by introducing data consistency constraints. For example, Chen et al. [13] use adversarial training to generate vehicle features in different viewpoints to improve the robustness of the model viewpoints, while Sun et al. [14] design a detail enhancement module based on stripe segmentation to strengthen the local feature expression. All of these methods are based on the concept of completing the missing information in high-dimensional space and fusing local and global features to optimize the recognition effect.

In addition to relying on image feature extraction and deep learning techniques, sensor-based vehicle re-recognition methods have also shown their advantages in recent years. Yang et al. [15] proposed the TIMS system, which fuses CAN bus data with visual information, and makes full use of multi-dimensional sensor data (e.g., acceleration, braking, and steering angle, etc.), to achieve stronger robustness in scenes with drastic changes in occlusion and illumination. However, at the same time, it is difficult for such methods to completely overcome the intra-class differences caused by lighting changes, obstacle occlusion or background complexity, which affect the stability of vehicle appearance features.

To solve the above problems, this study introduces an approach based on the attention mechanism to effectively improve the system's adaptability to environmental inconsistencies. The method is able to dynamically focus on key regions during feature extraction and suppress redundant information and noise interference, thus improving the overall recognition performance and robustness. The experimental results show that compared with the traditional method, this strategy shows more significant advantages in complex scenes, which ensures the reliability and stability of the Re-ID system in practical applications.

2.2 Attention Mechanisms

In human vision research, scholars have been investigating visual attention mechanisms [16] to understand how they filter essential information from the non-essential, recognizing the boundaries of our capacity for information processing. These mechanisms are indispensable for information processing systems, which must parse immense volumes of data to isolate pivotal information for subsequent analysis. Advancements in this domain promise to enhance machine vision systems, potentially equating their effectiveness with that of humans. Echoing this line of inquiry, recent studies in vehicle re-identification have integrated attention modules to mitigate the effects of variable viewing angles and detection inaccuracies. Liu et al. [17] proposed an attention model designed to refine the extraction of discriminative features while attenuating background interference, marking a significant step forward in the application of attention mechanisms to vehicle re-identification. In consonance with these developments, Zheng et al. [18] delved into multi-scale attention, leveraging spatial and channel attention branches to sharpen the precision of feature extraction. Guo et al. [19] proposed a novel dual-pooled attention module focused on improving vehicle Re-ID performance in Unmanned Aerial Vehicle (UAV) filming scenarios, which significantly improves the accuracy and robustness of vehicle recognition by introducing an attention mechanism. Xiang et al. [20] proposed LSKA-ReID to further optimize the application of the self-attention mechanism. LSKA-ReID breaks through the limitation of local receptive fields, and effectively combines the global information with

local details by combining the Large Separable Kernel Attention (LSKA) with the Hybrid Channel Attention (HCA) to enable the model to extract the features of the vehicle in a more comprehensive way.

The emphasis on attention mechanisms is thus paramount for the accurate discernment and recognition of critical vehicle features under diverse circumstances. Mindful of the imperative to harmonize model intricacy with operational efficacy, our research presents the Efficient Channel Attention (ECA) approach. Characterized by its minimal parameter footprint, the ECA outperforms alternative attention schemes in vehicle re-identification by maximizing accuracy and simplifying model complexity.

2.3 Graph Convolutional Network

Graph Convolutional Networks (GCNs) have emerged as robust analyzers of graph-structured data, adept at integrating the attributes of individual nodes with the contexts provided by their neighbors, thus effectively encapsulating the intrinsic diversity of graphs. This capacity has propelled GCNs to prominence in several computer vision applications, notably in object recognition, image classification, and semantic segmentation. In these fields, GCNs have proven especially competent in managing datasets of varied sizes and complex semantic content. The foundational work by Kipf et al. [21] has been pivotal in expediting the adoption of GCNs within semi-supervised learning frameworks, thereby ushering in innovative methodologies for training graph-based neural networks. In the context of re-identification (Re-ID) processes, which span the identification of both individuals and vehicles, GCNs have proven indispensable in extracting subtle relational patterns among images, surpassing the capabilities of previous techniques. The Spatial-Temporal Graph Convolutional Network, introduced by Yang et al. [22] in 2020, represents a significant advancement in video-based person Re-ID, dissecting the complex spatial-temporal interplay present in video data. Parallelly, the PrGCN model put forward by Liu et al. in 2021 [23] has augmented the robustness of similarity determinations crucial for person Re-ID, by integrating probability prediction within the GCN structure. In 2022, Pan et al.'s AAGCN [24] introduced an adjacency-aware mechanism, markedly reducing intra-class variance and enhancing the accuracy of Re-ID systems. Furthermore, the graph convolution operators designed by Zhang et al. [25] specifically for Re-ID re-ranking purposed to refine and exemplify the adaptability and strength of GCNs. In the same vein, the part-guided variation of GCNs as delineated by Zhang et al. [26] in 2021, which emphasizes the synthesis of localized features, heralds a new viable direction for person Re-ID research. In alignment with the diversification trend, the cutting-edge GCN development by Saber et al. in 2022 [27], incorporating Triplet Attention Learning, marks a pivotal advancement in amplifying person Re-ID through attention-based learning pathways. The deployment of GCNs in vehicle Re-ID initiatives, extracting features from automotive datasets, testifies to the versatile utility of GCNs and presages a transformative chapter in computational vision, reinforcing the essential role of GCNs in contemporary academic investigation.

3 The Proposed Method

This section commences by delineating the overarching workflow of the network in a comprehensive manner. Following this, it offers an exhaustive exposition of the feature extraction process for each constituent segment. The discourse culminates with the introduction of the formulation pertaining to the network's loss function.

3.1 Method Overview

Fig. 2 presents the architecture of the Attention-Guided Graph Convolution Network for vehicle reidentification (AG-GCN). The system is composed of three primary components: the backbone network, the local branch, and the global branch. The ResNet50 architecture serves as the foundational backbone of the method, which is augmented with the Efficient Channel Attention (ECA) mechanism to attenuate extraneous features and amplify salient ones, thereby preserving both primary and ancillary information within the global features. In the global branch, a modified ResNet-50 extracts the overall characteristics of the image, while the local branch processes nine segmented local features as vertex attributes. These attributes are then synthesized and propagated via a graph convolutional network (GCN), facilitating the learning of structured information. Subsequently, the assimilation of global and local structural features is achieved. The paper concludes with the employment of joint optimization across multiple loss functions to enable precise predictions.



Figure 2: AG-GCN for vehicle re-identification

3.1.1 Backbone Network

In this paper, we use ResNet50 as the underlying network for feature extraction with some modifications. Specifically, ResNet50 eliminates the downsampling operation at layer 4, which makes the feature map more prominent. In order to introduce channel attention in the residual module of ResNet50, the attention module ECA is proposed and embedded. As shown in Fig. 3, channel weights are applied to the original feature maps, thus allowing the network to selectively focus on specific information of interest instead of spreading the attention evenly across the scene.

The ECA module is designed to address the potential issue where the dimensionality reduction operation in the SE module may negatively affect channel attention prediction while increasing model complexity. As an enhanced, lightweight attention module derived from SE, ECA introduces only a minimal number of additional parameters and negligible computational overhead. Compared with other attention modules, ECA is designed to achieve improved performance by employing a local cross-channel interaction strategy without dimensionality reduction. This is implemented using a one-dimensional convolution, whose kernel size is adaptively determined by a specific function. The specific operations are depicted and explained as follows in Fig. 4. Using the input feature map as a starting point, first create an $1 \times 1 \times C$ vector using global average pooling independently for each channel, and then utilize a local cross-channel interaction approach without dimensionality reduction, which is accomplished by implementing one-dimensional convolution. The size of the convolution kernel for one-dimensional convolution is adaptively chosen by a specific function. Subsequently, the weight of each channel is determined by a sigmoid function, as demonstrated

in Eq. (1). Only a few parameters are involved, allowing layers with a larger number of channels to have more cross-channel interactions. The formula for calculating the adaptive convolution kernel's size is:

$$k = \psi(C) = \left| \frac{\log_2(C)}{\gamma} + \frac{b}{\gamma} \right|$$

$$\omega = \sigma(CID_k(\gamma))$$
(1)
(2)

where *k* is a one-dimensional convolution kernel's size, which determines the scope of local interaction. y is the channel feature. *C* indicates the number of channels. y and b are hyperparameters, which y = 1, b = 1. *CID* is one-dimensional convolution. ω represents the weight of the channel.



Figure 3: Residual module. (a) ECA-Resnet; (b) Resnet



Figure 4: ECA module

3.1.2 Local Branch

Subsequently, this study derives the final local features by constructing a structured graph representation of the vehicle, which is then processed using a Graph Convolutional Network (GCN) to capture fine-grained spatial relationships among different vehicle parts. The local branch extracts intermediate feature maps from

the fourth layer of the backbone network, where rich local details are preserved. These feature maps are then partitioned into multiple regions, each serving as a graph node, while edges are defined based on spatial adjacency and feature similarity. The constructed graph is fed into the GCN, which propagates and aggregates information among neighboring nodes, allowing the network to learn structural dependencies and relationships between vehicle components. This structured feature representation enhances discriminative power, reduces intra-class variations, and mitigates the impact of visually similar vehicles. Finally, the refined local features from the GCN are fused with global features, forming a comprehensive feature representation for robust vehicle re-identification. The mathematical expression corresponding to each layer is as follows:

$$H^{l+1} = \sigma \left(\tilde{D}^{-\frac{1}{2}} \tilde{A} \tilde{D}^{-\frac{1}{2}} H^l W^l \right)$$
(3)

where $H^l \in \mathbb{R}^{N \times D}$ is the matrix of activations in the l_{th} . $\sigma(\cdot)$ is nonlinear activation function. $\tilde{D} \in \mathbb{R}^{C \times C}$ is the degree matric of \tilde{A} . $\tilde{A} \in \mathbb{R}^{C \times C}$ is the adjacent. The adjacency matrix of the undirected graph G with additional self-connections is shown here as $\tilde{A} = A + I_N$. The identity matrix is I_N . W^l is a layer-specific trainable weight matrix. $H^0 = [r_1, \ldots, r_N]^T$ is an initial matrix, which obtained from regional features. It is possible to generate a more discriminative fine-grained local feature representation by feeding the graph structure into GCN to learn the internal outcome relationship between vehicles.

3.1.3 Global Branch

The local branch's proficiency lies in its ability to capture deep vehicle attributes; however, the holistic perspective of the global branch is indispensable, encompassing characteristics such as vehicle color and type. The distinctiveness of vehicle color aids in differentiating vehicles using graph recognition techniques. Existing re-identification datasets that include annotations for both model and color facilitate the refinement of Re-ID models.

The global branch constitutes the first substantial divergence following the backbone network. Its detailed network structure is illustrated in the upper portion of Fig. 2. Feature maps from the backbone network undergo an initial compression via a global average pooling layer, extracting both global features and attributes of expansive flat regions. Subsequent to this pooling, the features are propagated through a convolutional layer followed by batch normalization to refine the feature set. The process culminates with classification, effected through a fully connected layer utilizing a cross-entropy loss function. Additionally, to sharpen vehicle contour discernment, triplet loss is employed to augment model optimization.

3.2 Loss Functions

In terms of loss function design, AG-GCN implements a multi-task learning strategy, utilizing two loss functions—cross entropy loss and triplet loss with batch hard mining (TriHard loss)—which simultaneously exert influence on the model. By striking a balance in the weight ratio of distinct losses relative to the total loss, the model possesses the ability to acquire a more robust feature mapping method, thereby enhancing the model's performance.

The cross-entropy loss function can steer the model towards learning a feature space that accommodates interfaces between each class, thereby possessing favourable classification properties. The definition of this function is illustrated in Eq (4):

$$L_{ce} = -\sum_{i=1}^{k} y_i \log(y')$$
(4)

where *k* means the number of images in a training batch. *y*' represents the probability of belonging to *i* categories. y_i is the true label and $y_i = \frac{e^{z_i}}{\sum_{i=1}^k e^{z_k}}$. *z* represents the output score of the classification layer.

The conventional triplet loss metric learning method, which relies on random sampling of training samples, has the limitation of only being able to gather some easily discriminative samples, thus falling short of effectively training the expressive re-identification (re-id) model. For this reason, the experiment employs the hard triplet loss method for difficult samples. The triple loss of difficult samples is an improved triple loss function. It can focus on more difficult triplet samples during training, greatly improving the convergence effect of the network. The calculation formula is as Eq. (5):

$$L_t = \frac{1}{p \times k} \sum_{a \in batch} \left(\max_{p \in A} d_{a,p} - \min_{p \in B} d_{a,n} + \alpha \right)_+$$
(5)

where in a training batch, we select *P* vehicles, with *K* images randomly selected for each vehicle. This is a mini batch that contains $P \times K$ images. *a* represents the randomly selected vehicle identity, $p \in A$ means that the image and the selected vehicle have the same vehicle identity ID, and $p \in B$ means that the image and the selected vehicle have different vehicle identity IDs. Besides, *d* represents the similarity between two vehicle images' features. α is a threshold and the interior distance. $(z)_+$ is equivalent to max (z, 0). So global branch loss L_g and the local branch loss L_l are as follows:

$$L_g = L_t\left(g\right) + L_{ce}\left(g\right) \tag{6}$$

$$L_l = L_t \left(l \right) + L_{ce} \left(l \right) \tag{7}$$

As mentioned in the review, the two loss functions are jointly trained in this paper, and the final loss L_{total} is shown in Eq. (8):

$$L_{total} = L_g + aL_l \tag{8}$$

where *a* is a weighing hyperparameter and a = 1.

4 Experiments

This section commences with an introduction to the datasets and evaluation metrics employed, discussed in detail in Sections 4.1 and 4.2, respectively. It then proceeds to Section 4.3, which delineates the experimental setup associated with the AG-GCN model. Following this setup, the AG-GCN's performance is scrutinized and benchmarked in Section 4.4. To elucidate the impact of the modules introduced in this study on vehicle re-identification, ablation studies are conducted and their findings are assessed in Section 4.5. Concluding this section, Section 4.6 showcases the practicability of the proposed method via a visual exposition of results.

4.1 Dataset

Vehicle re-identification datasets predominantly originate from traffic surveillance cameras and comprise multi-angle vehicular images. The datasets utilized in this research are publicly accessible and encompass:

The VeRi-776 dataset, procured from real-world settings devoid of control, comprises 49,360 images of 776 individual vehicles. Each vehicle is meticulously annotated with attributes such as color, make, and model. This dataset covers 10 vehicular hues and 9 model types, supplemented by extensive spatiotemporal

metadata, including time-stamps and camera locations. The images are harvested from an array of 20 cameras, with each typically capturing a subject vehicle on six separate occasions, thus fostering a comprehensive dataset conducive to intricate model assessments.

Furthermore, the VehicleID dataset, assembled by Peking University's National Engineering Laboratory for Video Technology (NELVT), boasts a repository of 221,763 images spanning 26,267 unique vehicles. Amassed through the lens of 40 high-definition cameras, the dataset not only provides vehicle identities, but also classifies their types and colors, embracing an impressive diversity of 250 car models. Accommodating diverse testing requirements, the dataset is structured to offer query sets containing 800, 1600, and 2400 images, respectively.

4.2 Evaluation Metrics

To address the problem of target re-identification, common evaluation metrics utilized include Mean Average Precision (mAP) and Rank-k Accuracy. Both of these metrics serve to measure the search capability of an algorithm and can effectively evaluate and quantify the results of re-identification.

We use average accuracy to evaluate overall performance. Calculate average precision (AP) for each query:

$$AP(q) = \frac{\sum_{k=1}^{n} p(k) \times rel(k)}{N_{gt}}$$
(9)

where p(k) means the accuracy of sorting the *k*th. $N_{gt}(q)$ means the total number of true retrievals for q. *rel*(k) is the indicator function. 1 if the *k*th result is the correct target vehicle, 0 otherwise. *mAP* reflects the real ranking results and calculates the average of the average accuracy in all queries. *mAP* can reflect the stability of the model evaluation accuracy. *mAP* is expressed as follows:

$$mAP = \frac{\sum\limits_{q=1}^{Q} AP(q)}{Q}$$
(10)

Rank(k) means that for all the images of the vehicles to be retrieved, sort the images in the matching image library by similarity in turn to see if the first k images of the sorting result are retrieved correctly. The calculation equation is as follows:

$$Rank(k) = CMC(k) = \frac{\sum_{k=1}^{Q} gt(q,k)}{Q}$$
(11)

where *k* represents the *k*th image in the list of retrieved images. The model performs better when the rank-*k* value is higher. Generally, Rank-1, Rank-5 and other values are used to reflect the performance of the model. *q* represents the *q*th image to be queried, *Q* represents all images to be queried. For the *k*th image, if the retrieval results are ranked within the top *k* names in order of similarity, then gt(q, k) equals 1, otherwise gt(q, k) equals 0.

4.3 Implementation Details

The computational experiments are facilitated by a Tesla V100 GPU equipped with 32 GB of onboard memory. For initializing the network, we adopt the ImageNet pre-trained parameters on the ResNet-50

architecture. The Adaptive Moment Estimation (Adam) algorithm is utilized for the optimization of model parameters. In an effort to conserve the integrity of the image data, the stride parameter within the final bottleneck layer of ResNet-50 is adjusted to one.

Moreover, we augment every bottleneck layer with an attention block. Throughout the experimental procedures, input images are maintained at a resolution of 256 × 256 pixels. We employ an assortment of data augmentation techniques, such as random horizontal flipping, random cropping, and random erasing, to enhance the model's resilience in handling data variance; each of these is probabilistically applied with a rate of 0.5. For the model training phase, Stochastic Gradient Descent (SGD) optimizer is set with a batch size of 64 within the training dataset. The model undergoes a training span covering 120 epochs; this includes an application of label smoothing as a safeguard against overfitting. The learning rate is initially established at 0.00003 and is scheduled to decrease by a factor of 10 at both the 40th and 80th epochs. A warmup strategy is employed in the initial 10 epochs, incrementally elevating the learning rate from 0.00003 to 0.0003. For the testing phase, the Euclidean distance measure is employed to evaluate the similarity metrics between the query image and the database images.

4.4 Comparison with State-of-the-Art Methods

To assess AG-GCN performance, experiments are performed on two open datasets: VeRi-776 or VehicleID. The compared methods include:(1) PROVID [1]; (2) RAM [28]; (3) VAMI [29]; (4) TAMR [30]; (5) LOMO [31]; (6) AAVER [32]; (7) QD DLF [33]; (8) PRN [5]; (9) SAVER [34]; (10) AGW [35]; (11) GiT [36]; (12) SOFCT [37].

In the comparative analysis presented in Table 1, AG-GCN's performance on the VeRi-776 benchmark dataset is evaluated against other referenced algorithms.

Methods	mAP (%)	Rank-1 (%)	Rank-5 (%)
LOMO [31]	9.60	25.30	46.50
VAMI [32]	50.13	77.03	90.82
PROVID [1]	53.42	81.56	95.11
RAM [28]	61.50	88.60	94.00
QD-DLF [33]	61.83	88.50	94.46
AAVER [32]	61.18	88.97	94.70
PRN [5]	74.30	94.30	98.70
GiT [36]	80.03	95.90	98.40
SOFCT [37]	80.07	95.80	98.40
AG-GCN (ours)	80.10	96.10	98.30

Table 1: Results comparisons over VeRi-776

The comparative analysis, delineated in Table 1, reveals that AG-GCN exhibits marked improvements over the traditional LOMO [31] algorithm, with substantial increments of 65.9% for mean Average Precision (mAP), 69.8% for Rank-1, and 51.6% for Rank-5, evidence of our methodology's accuracy advantage. AG-GCN delivers a 95.1% Rank-1 match rate, a 98.10% Rank-5 match rate, and secures a 75.50% mAP. Despite these results, AG-GCN is still slightly behind PRN [5], GiT [36] and SOFCT [37] in Rank-5 metrics. Moreover, in the context of methods leveraging multifaceted feature extraction such as VAMI [29], AAVER [30], and QD-DLF [33], AG-GCN presents notable competitiveness. The results endorse the efficacy of incorporating attention mechanisms with graph convolutional networks in enhancing the model's discernment of both holistic and granular vehicle features.

AG-GCN underwent rigorous testing on the VehicleID benchmark dataset, which categorizes the dataset into three scaled subsets: small, medium, and large, containing 800, 1600, and 2400 images, respectively. The results of these tests are consolidated in Table 2, which compares the performance across the different scales. The data indicates that AG-GCN surpasses competing methodologies in terms of performance metrics. While other approaches, such as AAVER [32] and AGW [35], leverage attention mechanisms to zone in on prominent local features, they frequently rely on intricate attention frameworks. The ECA attention mechanism employed in this research, despite its parameter efficiency, yields significant enhancements in performance. AG-GCN's superior results across all three test sets attest to its robustness and adaptability.

Methods	Test	-800	Test	-1600	Test-2	2400
	Rank-1	Rank-5	Rank-1	Rank-5	Rank-1	Rank-5
LOMO [31]	19.76	32.14	18.95	29.46	15.26	25.63
DRDL [2]	49.00	73.50	42.80	66.80	38.20	61.60
VAMI [29]	63.12	83.25	52.87	75.12	47.24	70.29
TAMR [30]	66.02	79.71	62.90	76.80	59.69	73.87
QD-DLF [31]	72.32	92.48	70.66	88.90	64.14	83.37
RAM [28]	75.20	91.50	72.30	87.00	67.70	84.50
AAVER [32]	74.69	93.82	68.62	89.95	63.54	85.64
PRN [5]	78.40	92.30	75.00	88.30	74.20	86.40
SAVER [34]	79.90	95.20	77.60	97.10	75.30	88.30
AGW [35]	80.20	95.10	77.10	89.80	75.20	88.40
AG-GCN (ours)	83.40	97.30	79.00	95.10	77.20	93.00

Table 2: Results comparisons over VehicleID

In culmination, this article substantiates that the integration of the graph convolution network with a fusion of global discriminative features and local granular features can amplify the efficacy of vehicle re-identification networks. The proposed model substantially escalates the precision of vehicle re-identification, reinforcing its practicality and applicability in the field.

4.5 Ablation Study

During the experimental phase, to ascertain the contributory roles of critical components within AG-GCN and to pinpoint the most effective structural configuration, a series of experiments were conducted utilizing the VeRi-776 dataset. The outcomes of these experiments are systematically presented in Table 3 and encompass four distinct experimental setups.

Model	mAP	Rank-1	Rank-5
Baseline	74.9	94.4	97.5
Baseline + ECA	77.9	95.3	97.7
Baseline + GCN	78.3	95.7	98.0
Proposed AG-GCN	80.1	96.1	98.3

Table 3: The ablation study of AG-GCN on the VeRi-776 dataset

In the context of these experiments, the term "Baseline" denotes the fundamental model framework engaged for vehicle re-identification purposes. The abbreviation "ECA" signifies the deployment of the Efficient Channel Attention mechanism, whereas "GCN" is indicative of the implementation of the Graph Convolutional Network. The impact of the various modules, particularly with regard to the mean Average Precision (mAP), is graphically articulated in Fig. 5. This visual representation underscores the modules' instrumental roles in enhancing the model's re-identification capabilities.



Figure 5: Methods' performance on VeRi-776 dataset

The mAP, Rank-1, and Rank-5 of the benchmark network reached 74.9%, 94.4%, 97.5%. Compared with the benchmark network, the improved resnet50 model after adding ECA were used, which improves the mAP, Rank-1, and Rank-5 by 3.0%, 0.9%, and 0.2%, respectively. This is because the modified resnet50 incorporates features from the middle layer and the last layer. Therefore, more effective global feature extraction is obtained. Using GCN, compared with the reference network, it is improved by 3.4%, 1.2%, and 0.5%, because using GCN can better obtain the internal relationship of the vehicle through the graph structure.

When the two branches exist separately, their performance is not as good as the combined AG-GCN. This is because when only a single branch is retained, the network can only learn partial vehicle features, which negatively affects the re-identification performance. Compared with ResNet-50, the Rank-1 on the dataset VeRi-776 is improved by 1.7%. The mAP is improved by 5.2%. It can be seen that there is a complementary effect between both two branches. The vehicle re-identification method is more robust by creating a more comprehensive vehicle feature by combining global features and local fine-grained features.

4.6 Visual Result

In order to more vividly display the performance of the proposed AG-GCN, we visualized some experimental results on the VeRi-776 dataset. Fig. 6 shows the qualitative query results of the methods proposed in this chapter and global model on the VeRi-776 dataset.

In this chapter, the proposed AG-GCN method is adopted to randomly select three different vehicles for training. The re-identification results are displayed in a visually sorted order, from the most similar to the least similar, with the top five results shown. The first line of each group represents the search result of the AG-GCN method proposed in this chapter, and the second line represents the query result of the global model. The target image in each group is the same query image. AG-GCN, which uses the improved ResNet50 as the

backbone network and combines two branches, is able to learn more comprehensive vehicle characteristics. As shown in Fig. 6, the first column is the vehicle to be queried. From the second column to the sixth column, the vehicle with red frame means the retrieval error, while the vehicle with green frame means the retrieval is correct. The global model retrieves many negative samples because of their great similarity in overall appearance. The model in this chapter filters out a lot of irrelevant vehicle information by focusing on details. AG-GCN can be well adapted to similar vehicles, which verifies the effectiveness of the proposed method. The results show that this method can learn meaningful feature representation and similarity measure under limited data conditions. As can be seen from Fig. 6, the proposed method has higher accuracy and robustness for different perspectives.



Figure 6: Query visualization images of Rank-5 results

5 Conclusions

The study at hand expounds upon the Attention-Guided Graph Convolutional Network (AG-GCN), an avant-garde approach devised primarily for the task of vehicle re-identification. This methodology exhibits significant potential in mitigating the challenges posed by pronounced intra-class disparities and notable inter-class resemblances. The process commences with the utilization of an augmented ResNet50 architecture, which is now equipped with an Efficient Channel Attention (ECA) module that aids not only in the acquisition of global vehicular features but also substantially elevates the extraction of intricate local details. At the local branch, the Graph Convolutional Network (GCN) plays a pivotal role by enabling the diffusion of localized information across various segments of the vehicle. The pivotal local features that are harvested through this process are imperative for the subsequent identification of vehicles. Culminating the process, the proposed method concurrently learns and amalgamates global and local features, eliciting refined attributes capable of discerning nuanced variances between images, thereby increasing the accuracy of re-identification significantly. The model introduced in this paper proves to be remarkably efficacious in

extracting distinct and salient vehicle attributes. Empirical validation conducted on two publicly accessible datasets corroborates the superior performance of the proposed method. It is pertinent to note, thus far, the methodologies discussed are geared towards refining the model's recognition capabilities within a singular domain. Aiming to further the practical utility of vehicle re-identification, future endeavors will be concentrated on streamlining cross-domain coordination, an initiative we anticipate addressing imminently.

Acknowledgement: The authors wish to express their sincere gratitude to the anonymous reviewers and the editor, whose valuable suggestions have significantly improved the quality of this manuscript.

Funding Statement: This research was funded by the National Natural Science Foundation of China (grant number: 62172292).

Author Contributions: The authors confirm contribution to the paper as follows: study conception and design: Ya-Jie Sun and Sai Ji; data collection: Ya-Jie Sun and Li-Wei Qiao; analysis and interpretation of results: Ya-Jie Sun, Li-Wei Qiao, and Sai Ji; draft manuscript preparation: Ya-Jie Sun, Li-Wei Qiao, and Sai Ji; manuscript final layout and preparation for submission: Ya-Jie Sun, Li-Wei Qiao, and Sai Ji. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: The data used to support the findings of this study are available from the corresponding author upon request.

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

- Liu X, Liu W, Mei T, Ma H. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In: Leibe B, Matas J, Sebe N, Welling M, editors. Computer Vision—ECCV 2016: 14th European Conference; 2016 Oct 11–14; Amsterdam, The Netherlands. Cham, Switzerland: Springer. doi:10.1007/978-3-319-46475-6_53.
- Liu H, Tian Y, Wang Y, Pang L, Huang T. Deep relative distance learning: tell the difference between similar vehicles. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2016 Jun 27–30; Las Vegas, NV, USA. p. 2167–75. doi:10.1109/CVPR.2016.238.
- Ma X, Zhu K, Guo H, Wang J, Huang M, Miao Q. Vehicle re-identification with refined part model. In: 2019 IEEE International Conference on Multimedia & Expo Workshops (ICMEW); 2019 Jul 8–12; Shanghai, China. p. 603–6. doi:10.1109/icmew.2019.00110.
- Wang Z, Tang L, Liu X, Yao Z, Yi S, Shao J, et al. Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification. In: 2017 IEEE International Conference on Computer Vision (ICCV); 2017 Oct 22–29; Venice, Italy. p. 379–87. doi:10.1109/ICCV.2017.49.
- He B, Li J, Zhao Y, Tian Y. Part-regularized near-duplicate vehicle re-identification. In: Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2019 Jun 15–20; Long Beach, CA, USA. doi:10.1109/cvpr.2019.00412.
- 6. Sun W, Chen X, Zhang X, Dai G, Chang P, He X. A multi-feature learning model with enhanced local attention for vehicle re-identification. Comput Mater Contin. 2021;69(3):3549–61. doi:10.32604/cmc.2021.021627.
- Zhu Y, Zha ZJ, Zhang T, Liu J, Luo J. A structured graph attention network for vehicle re-identification. In: Proceedings of the 28th ACM International Conference on Multimedia; 2020 Oct 12–16; Seattle, WA, USA. doi:10. 1145/3394171.3413607.
- Chen TS, Lee MY, Liu CT, Chien SY. Viewpoint-aware channel-wise attentive network for vehicle re-identification. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); 2020 Jun 14–19; Seattle, WA, USA. doi:10.1109/cvprw50498.2020.00295.

- 9. Lee S, Woo T, Lee SH. Multi-attention-based soft partition network for vehicle re-identification. J Comput Des Eng. 2023;10(2):488–502. Cambridge, UK: Cambridge University Press; 2023. doi:10.1093/jcde/qwad014.
- Wang Q, Wu B, Zhu P, Li P, Zuo W, Hu Q. ECA-net: efficient channel attention for deep convolutional neural networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA. p. 11531–9. doi:10.1109/cvpr42600.2020.01155.
- 11. Yao M, Wang H, Chen Y, Fu X. Between/within view information completing for tensorial incomplete multi-view clustering. IEEE Trans Multimed. 2024;27:1538–50. doi:10.1109/TMM.2024.3521771.
- 12. Wang H, Yao M, Chen Y, Xu Y, Liu H, Jia W, et al. Manifold-based incomplete multi-view clustering via bi-consistency guidance. IEEE Trans Multimed. 2024;26:10001–14. doi:10.1109/TMM.2024.3405650.
- 13. Chen Y, Ke W, Lin H, Lam CT, Lv K, Sheng H, et al. Local perspective based synthesis for vehicle re-identification: a transformation state adversarial method. J Vis Commun Image Represent. 2022;83:103432.
- 14. Sun Z, Nie X, Bi X, Wang S, Yin Y. Detail enhancement-based vehicle re-identification with orientation-guided re-ranking. Pattern Recognit. 2023;137(1):109304. doi:10.1016/j.patcog.2023.109304.
- 15. Yang H, Cai J, Zhu M, Liu C, Wang Y. Traffic-informed multi-camera sensing (TIMS) system based on vehicle re-identification. IEEE Trans Intell Transp Syst. 2022;23(10):17189–200. doi:10.1109/TITS.2022.3154368.
- 16. Hassanin M, Anwar S, Radwan I, Khan FS, Mian A. Visual attention methods in deep learning: an in-depth survey. Inf Fusion. 2024;108(3):102417. doi:10.1016/j.inffus.2024.102417.
- Liu K, Xu Z, Hou Z, Zhao Z, Su F. Further non-local and channel attention networks for vehicle re-identification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW); 2020 Jun 14–19; Seattle, WA, USA. doi:10.1109/cvprw50498.2020.00300.
- 18. Zheng A, Lin X, Dong J, Wang W, Tang J, Luo B. Multi-scale attention vehicle re-identification. Neural Comput Appl. 2020;32(23):17489–503. doi:10.1007/s00521-020-05108-x.
- 19. Guo X, Yang J, Jia X, Zang C, Xu Y, Chen Z. A novel dual-pooling attention module for UAV vehicle reidentification. Sci Rep. 2024;14(1):2027. doi:10.1038/s41598-024-52225-x.
- 20. Xiang X, Ma Z, Li X, Zhang L, Zhen X. Vehicle re-identification with large separable kernel attention and hybrid channel attention. Image Vis Comput. 2025;155(5):105442. doi:10.1016/j.imavis.2025.105442.
- 21. Kipf TN, Welling M. Semi-supervised classification with graph convolutional networks. arXiv:1609.02907. 2016. doi:10.48550/arXiv.1609.02907.
- 22. Yang J, Zheng WS, Yang Q, Chen YC, Tian Q. Spatial-temporal graph convolutional network for video-based person re-identification. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR); 2020 Jun 13–19; Seattle, WA, USA.
- 23. Liu H, Xiao Z, Fan B, Zeng H, Zhang Y, Jiang G. PrGCN: probability prediction with graph convolutional network for person re-identification. Neurocomputing. 2021;423(12):57–70. doi:10.1016/j.neucom.2020.10.019.
- 24. Pan H, Bai Y, He Z, Zhang C. AAGCN: adjacency-aware Graph Convolutional Network for person reidentification. Knowl Based Syst. 2022;236:107300. Amsterdam, The Netherlands: Elsevier; 2022. doi:10.1016/j. knosys.2021.107300.
- Zhang Y, Qian Q, Liu C, Chen W, Wang F, Li H, et al. Graph convolution for re-ranking in person re-identification. In: 2022 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP); 2022 May 23–27; Singapore. doi:10.1109/ICASSP43922.2022.9747298.
- 26. Zhang Z, Zhang H, Liu S, Xie Y, Durrani TS. Part-guided graph convolution networks for person re-identification. Pattern Recognit. 2021;120:108155. doi:10.1016/j.patcog.2021.108155.
- 27. Saber S, Amin K, Pławiak P, Tadeusiewicz R, Hammad M. Graph convolutional network with triplet attention learning for person re-identification. Inf Sci. 2022;617(2):331–45. doi:10.1016/j.ins.2022.10.105.
- 28. Liu X, Zhang S, Huang Q, Gao W. RAM: a region-aware deep model for vehicle re-identification. In: 2018 IEEE International Conference on Multimedia and Expo (ICME); 2018 Jul 23–27; San Diego, CA, USA. p. 1–6. doi:10. 1109/ICME.2018.8486589.
- 29. Zhouy Y, Shao L. Viewpoint-aware attentive multi-view inference for vehicle re-identification. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition; 2018 Jun 18–23; Salt Lake City, UT, USA. p. 6489–98. doi:10.1109/CVPR.2018.00679.

- 30. Guo H, Zhu K, Tang M, Wang J. Two-level attention network with multi-grain ranking loss for vehicle reidentification. IEEE Trans Image Process. 2019;28(9):4328–38. doi:10.1109/TIP.2019.2910408.
- Liao S, Hu Y, Zhu X, Li SZ. Person re-identification by Local Maximal Occurrence representation and metric learning. In: 2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR); 2015 Jun 7–12; Boston, MA, USA. doi:10.1109/CVPR.2015.7298832.
- Khorramshahi P, Kumar A, Peri N, Rambhatla SS, Chen JC, Chellappa R. A dual-path model with adaptive attention for vehicle re-identification. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV); 2019 Oct 27–Nov 2; Seoul, Republic of Korea. p. 6131–40. doi:10.1109/iccv.2019.00623.
- 33. Zhu J, Zeng H, Huang J, Liao S, Lei Z, Cai C, et al. Vehicle re-identification using quadruple directional deep learning features. IEEE Trans Intell Transport Syst. 2020;21(1):410–20. doi:10.1109/TITS.2019.2901312.
- 34. Khorramshahi P, Peri N, Chen JC, Chellappa R. The devil is in the details: self-supervised attention for vehicle re-identification. In: Computer Vision—ECCV 2020: 16th European Conference; 2020 Aug 23–28; Glasgow, UK.
- 35. Ye M, Shen J, Lin G, Xiang T, Shao L, Hoi SCH. Deep learning for person re-identification: a survey and outlook. IEEE Trans Pattern Anal Mach Intell. 2022;44(6):2872–93. doi:10.1109/TPAMI.2021.3054775.
- Shen F, Xie Y, Zhu J, Zhu X, Zeng H. GiT: graph interactive transformer for vehicle re-identification. IEEE Trans Image Process. 2023;32(2):1039–51. doi:10.1109/TIP.2023.3238642.
- Yu Z, Huang Z, Pei J, Tahsin L, Sun D. Semantic-oriented feature coupling transformer for vehicle re-identification in intelligent transportation system. IEEE Trans Intell Transp Syst. 2024;25(3):2803–13. doi:10.1109/TITS.2023. 3257873.