

Article

DAFF-Net: A Dual-Branch Attention-Guided Feature Fusion Network for Vehicle Re-Identification

Yi Guo ^{1,2}, Guowu Yuan ^{1,2,*} , Wen Li ¹ and Hao Li ^{1,2}

¹ Yunnan Key Laboratory of Digital Communications, Yunnan Communications Investment & Construction Group Co., Ltd., Kunming 650103, China; guoyi@stu.ynu.edu.cn (Y.G.); liwen@ycicloud.com (W.L.)

² School of Information Science and Engineering, Yunnan University, Kunming 650504, China

* Correspondence: gwyuan@ynu.edu.cn

Abstract

Vehicle re-identification (Re-ID) is a critical task in the fields of intelligent transportation and urban surveillance. This task faces numerous challenges, such as significant changes in shooting angles, strong similarities in appearance between different vehicles of the same model, and difficulties in modeling fine-grained differences. To overcome the shortcomings of existing methods in local feature extraction and multi-scale fusion, this paper proposes an attention-guided dual-branch feature fusion network (DAFF-Net). The network uses ResNet50-ibn as its backbone and designs two complementary feature extraction branches. One branch fuses cross-layer attention between shallow and deep features, introducing a Temperature-Calibration Attention Fusion Module (TCAF) to improve the accuracy of cross-layer feature fusion effectively. The other branch enhances multi-scale attention for mid-layer features, constructing a Multi-Scale Gated Attention Module (MSGa) to extract local details and directional structural information. Finally, the discriminative ability of the enhanced features is improved by concatenating the two branch features and jointly optimizing the network using triplet loss, cross-entropy loss, and center loss. Experimental results on the VeRi-776 and VehicleID public datasets indicate that the proposed DAFF-Net outperforms existing mainstream methods in multiple key metrics. On the VeRi-776 dataset, mAP and CMC@1 increased to 82.2% and 97.5%, respectively. In the three test subsets of the VehicleID dataset, the CMC@1 metric achieved 90.7%, 84.6%, and 82.1%, respectively, demonstrating the effectiveness of the proposed network in vehicle re-identification tasks.



Academic Editor: Yajie Zou

Received: 1 September 2025

Revised: 24 October 2025

Accepted: 27 October 2025

Published: 29 October 2025

Citation: Guo, Y.; Yuan, G.; Li, W.; Li,

H. DAFF-Net: A Dual-Branch Attention-Guided Feature Fusion Network for Vehicle Re-Identification. *Algorithms* **2025**, *18*, 690. <https://doi.org/10.3390/a18110690>

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: vehicle re-identification; intelligent transportation; feature fusion network; multi-scale attention; fine-grained recognition

1. Introduction

Vehicle re-identification technology (Re-ID), or vehicle re-recognition technology, aims to retrieve and match target vehicles from a database of vehicle images captured by surveillance cameras. In recent years, with the rapid development of intelligent transportation management systems, this technology has been widely applied in various practical scenarios, including vehicle search and tracking, automated toll collection, road access control, intelligent parking access, and access control systems [1].

Early vehicle re-identification methods primarily drew on traditional image retrieval techniques, relying on manually designed features to extract vehicle appearance information and perform feature matching based on this information. However, such methods typically extract low-level features with weak discriminative power and limited expressive

capabilities, making it difficult to cope with recognition tasks under complex real-world conditions such as cross-perspective, occlusion, and lighting changes. Thanks to the rapid development of artificial intelligence technology, vehicle re-identification methods based on computer vision technology have made significant progress [2]. Using deep neural networks to learn vehicle image features, the network can automatically extract depth appearance features from images with complex backgrounds, different viewpoints, and different lighting conditions. Accurate vehicle recognition and matching can be achieved by calculating the similarity between these features.

Although vehicle re-identification methods using computer vision technology have achieved good feature extraction and matching results, the task still faces many challenges. On the one hand, image information is often incomplete due to obstruction and background interference caused by different shooting angles. On the other hand, vehicles of the same make and model appear highly similar when captured from the same perspective. However, the same vehicle may exhibit significant differences in appearance depending on the time, location, or condition, resulting in the problem of “small inter-class differences and large intra-class differences,” as shown in Figure 1. These factors severely impact model accuracy, making the enhancement of vehicle re-identification performance a current research hotspot and challenge.

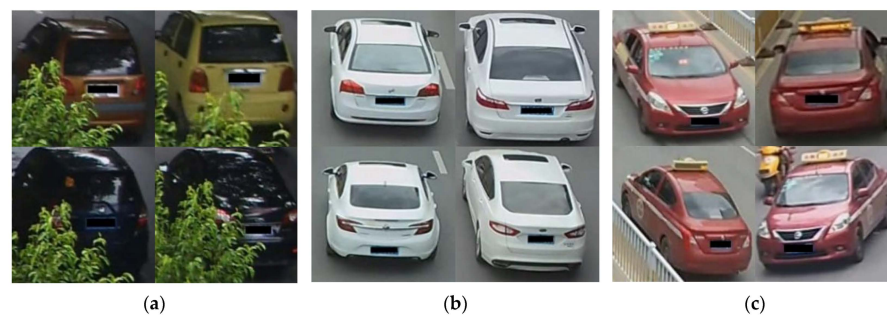


Figure 1. Challenges in vehicle re-identification tasks. (a) Vehicles are obscured in complex environments. (b) Vehicles of different makes appear very similar when photographed from the same angle. (c) Vehicles of the same make appear quite different when photographed from different angles.

To address the issues mentioned above, some studies have developed methods capable of performing global aggregation on feature maps across the entire vehicle image. This generates global appearance features at the image level as the primary descriptive information and further designs various mechanisms to optimize feature distribution. Liu et al. [3] constructed a deep relative distance learning framework based on global features, introduced vehicle ID constraints, and enhanced the feature aggregation of similar samples through triplet loss; Bai et al. [4] also used global features as a basis and proposed an intra-class deviation loss function to guide the model to compress the distribution differences between samples in the same group in the embedding space; Chu et al. [5] introduced a perspective classifier auxiliary branch to enable the model to perceive perspectives and improve feature consistency under different shooting angles. These methods can model the overall appearance structure of vehicles, thereby improving the ability to recognize samples with significant differences in appearance to a certain extent. However, when distinguishing between vehicles with highly similar appearances, relying solely on overall appearance characteristics still has obvious limitations, making it difficult to capture the local details crucial for identification. To address this issue, researchers began focusing on modeling local features of vehicles, supplementing overall appearance features with local features that retain information about key vehicle components such as headlights and grilles. Zhang et al. [6] used a pre-trained object detection model to locate candidate

component regions. They learned the importance of the distribution of each local region, enabling the model to assign higher attention weights to component regions with strong discriminative power. He et al. [7] proposed the PRN network, introducing a vehicle component detection branch into the model to automatically locate multiple key local regions, including headlights, windows, and logos, and map these local regions to the global feature map, thereby effectively enhancing the model's ability to perceive fine-grained differences. These methods, which introduce vehicle component detection regions and use attention mechanisms to extract discriminative local features from images adaptively, highlight the differences in specific details of vehicles while reducing interference from the background and irrelevant regions, thereby improving the model's ability to distinguish between vehicles with highly similar appearances. If the obtained local detail features are further fused with global features using feature fusion techniques to enhance feature representation, more discriminative fused features can be obtained [8]. Based on the above approach, this paper proposes an attention-guided dual-branch feature fusion network (DAFF-Net). This network introduces differentiated attention mechanisms in two branches and combines multi-scale complementary feature extraction strategies to extract features from different levels and scales effectively. Additionally, the network achieves integration of features from different levels through feature fusion operations, enhancing the model's comprehensive perception capabilities for both local details and global semantics, thereby significantly improving vehicle re-identification performance.

The main contributions of this work are summarized as follows:

1. We propose a dual-branch attention-guided feature fusion network (DAFF-Net), which introduces a Temperature-Calibration Attention Fusion (TCAF) module and a Multi-Scale Gated Attention (MSGa) module to fully exploit and fuse the complementary information of low-level, mid-level, and high-level features in the backbone network, thereby achieving effective collaboration between local details and global semantics. Additionally, to enhance the intra-class compactness of features, this paper introduces a center loss function during training, further optimizing the model's representation learning capabilities.
2. We employ an integrated mechanism of "align first, fuse then, and adaptively select", enhancing key scale and directional cues within a unified semantic space. Compared to traditional fusion methods like static concatenation, fixed pyramids, or single-layer attention, this design reduces information dilution and cross-layer interference during multi-level fusion. It achieves a more balanced trade-off between feature alignment, multi-scale modeling, and directional selection at near-linear computational cost, thereby preserving finer-grained component information and enabling more efficient feature collaboration.
3. Extensive experimental results on the VeRi-776 and VehicleID datasets demonstrate that the proposed method outperforms other relevant approaches in terms of feature expression accuracy and cross-layer fusion effectiveness.

2. Related Work

2.1. Methods for Fusing Global and Local Features

A feature learning method that jointly models global appearance features extracted from images with specific regions or structural parts can simultaneously focus on the overall contour and local details of a vehicle, helping to distinguish between vehicles that are similar in appearance but differ in local features, thereby improving the model's discriminative ability. Li et al. [9] utilized a self-supervised learning mechanism to guide the model in automatically identifying local key regions within images, then fused the extracted local features with global features. Furthermore, through downsampling, they achieved

the integration of multi-scale features. Lee et al. [10] designed a multi-branch attention mechanism to extract local discriminative information and introduced a soft segmentation strategy in the feature fusion stage to integrate discriminative features across multiple scales and perspectives, thereby improving re-identification accuracy under cross-camera conditions. Some studies employ object detection algorithms to extract local vehicle features to achieve more precise localization of key vehicle regions. For example, studies [6,7,11] utilize prior information in images, using pre-trained vehicle component detectors to locate key regions, and then fuse these key regions with global features to construct more expressive feature representations. This design of fusing global and local features has significantly improved model accuracy in cross-camera matching tasks across multiple public datasets, demonstrating strong generalizability and effectiveness, and has become one of the key research directions in current vehicle re-identification tasks.

2.2. Methods for Fusing Global and Attribute Features

Vehicle attribute information typically remains relatively stable across different perspectives and environments. Therefore, some researchers use vehicle attribute information to supplement global features to enhance feature expressiveness. Zhang et al. [12] extracted discriminative attribute features related to vehicle identity through an attribute-guided feature disentanglement mechanism and optimized identity recognition and attribute classification tasks using a multi-task learning framework. Li et al. [13] introduced two modules—attribute enhancement and state weakening—to comprehensively model attribute and state information within the global feature embedding module. Yu et al. [14] extracted various attribute features of vehicles and incorporated them into the Transformer as semantic supplements to visual features. Additionally, studies [15–17] introduced an attribute recognition branch to extract semantic attributes and utilize attention mechanisms to enhance responses to key attribute features, thereby suppressing background interference information unrelated to identity in global features. To improve the model's consistent understanding of attributes in complex environments, some methods introduce attribute consistency loss or design attribute-assisted supervision mechanisms to optimize joint features from the loss layer perspective. Overall, these studies demonstrate the significant value of attribute features in enhancing semantic modeling and improving vehicle re-identification accuracy.

2.3. Methods for Fusing Global Features and Spatiotemporal Information

In complex scenarios such as low resolution and limited shooting angles, it is difficult for models to extract visual features with sufficient discriminative power. To address this issue, some researchers have introduced spatio-temporal information into joint modeling based on visual features, using information such as the shooting time, shooting angle, and geographical location of the camera to constrain the matching range of candidate vehicles further. Studies [18,19] model the spatial distance and shooting time intervals between adjacent cameras to determine whether the appearance of a vehicle in different cameras is reasonable. Sun et al. [20] use vehicle perspective information to guide the model in performing two-stage re-ranking of retrieval results, identifying samples with significant directional differences from the target vehicle through a direction-aware mechanism. Li et al. [21] use a knowledge graph transfer strategy to construct a semantic association graph between vehicles, and employ a graph neural network to perform relationship modeling and context information propagation on this structure, enhancing the model's consistent understanding of vehicle features under multi-view changes. Meng et al. [22] trained a U-Net network to parse vehicle images into four perspective masks: front, rear, side, and top. They used mask average pooling to process global features, obtaining aligned

local perspective features. These methods, which fuse global features with vehicle spatio-temporal information, provide stronger prior constraints and semantic supplementation for the vehicle matching process, effectively addressing the issue of insufficient discrimination of pure visual features in complex environments.

3. Proposed Network Model

3.1. Model Overview

Unlike general classification tasks, vehicle re-identification is a fine-grained visual recognition task that requires the network to have high-quality feature extraction capabilities. This paper selects the ResNet50-ibn network, an improved version of ResNet50 [23], as the backbone network and constructs a dual-branch vehicle re-identification feature extraction network, DAFF-Net, as shown in Figure 2.

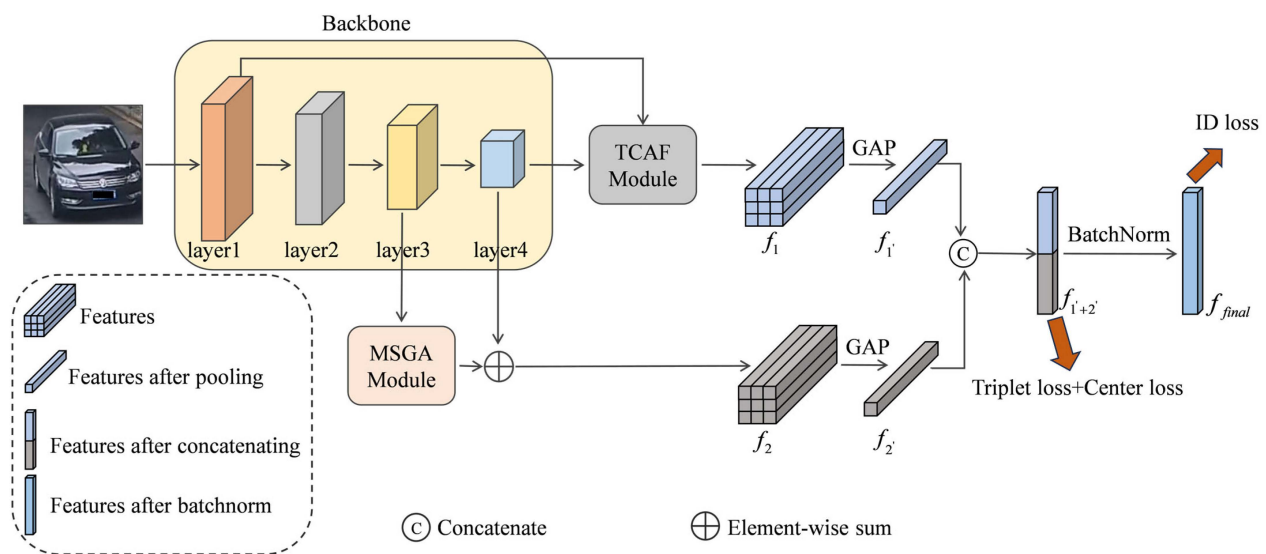


Figure 2. The structure of DAFF-Net.

In Figure 2, the network primarily consists of two feature fusion branches guided by attention mechanisms, which extract discriminative information from different layers of vehicle images and enhance feature expressiveness through cross-layer fusion. The first branch takes the shallow-layer (layer1) and deep-layer (layer4) features from the backbone network as input and uses the Temperature-Calibration Attention Fusion (TCAF) module for fusion, focusing on cross-layer information guidance fusion and emphasizing the complementary roles of shallow-layer and deep-layer structural features. The second branch takes the backbone network's mid-layer features (layer3) as input, processes them through the Multi-Scale Gated Attention (MSGA) module, and then performs a simple additive fusion with layer4 features, emphasizing the modeling of contextual associations within a single layer. Finally, the output features from both branches are concatenated to obtain the final integrated features at low, medium, and high levels across different scales. This approach, which first aligns scales between unrelated layers before merging, captures global semantics without neglecting local details.

To further enhance the clustering effect in the feature space, this paper introduces a center loss in addition to the conventional triplet loss and cross-entropy loss. This loss reduces intra-class variability by bringing samples of the same class closer together in the feature space, resulting in a more compact feature distribution and clearer boundaries between classes.

Section 3.2 below introduces the Temperature-Calibration Attention Fusion Module shown in Figure 3, Section 3.3 introduces the Multi-Scale Gated Attention Module shown in Figure 4, and Section 3.4 introduces the Center Loss Function.

3.2. Temperature-Calibration Attention Fusion Module

In the ResNet50-ibn backbone network, features from different layers exhibit distinct semantic hierarchical characteristics. Shallow-layer features primarily focus on detailed information within images, while deep-layer features contain more abstract, semantically discriminative information. To integrate detailed and semantic information, this paper designs the TCAF module to guide cross-layer fusion in the channel and spatial dimensions. The structure of the TCAF module is shown in Figure 3.

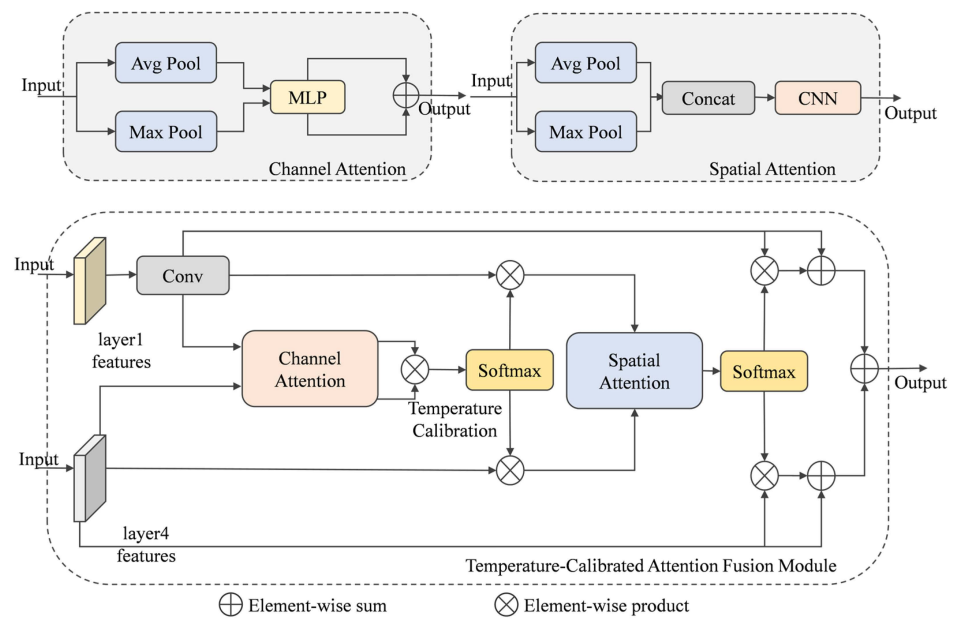


Figure 3. Temperature-Calibration Attention Fusion module structure.

The TCAF module consists of two key submodules: the channel attention cross-module and the spatial attention fusion module [24]. Given the features from layer1 and layer4 of the backbone network, a convolutional layer aligns the dimensions of these two layers. A bidirectional cross-attention mechanism is then applied in the channel dimension to facilitate information exchange and guidance. Specifically, global average and max pooling are performed on the features of each layer, respectively, followed by a shared two-layer MLP to obtain the channel attention vectors a_1, a_2 . Additionally, a learnable temperature parameter τ is introduced into the cross-attention mechanism for temperature calibration, which is used to scale the attention distribution and automatically adjust the intensity of attention focus. The calculation method for attention weights after introducing the temperature parameter τ is as follows:

$$W = \text{softmax}\left(\frac{a_1 a_2^T}{\tau}\right) \quad (1)$$

Subsequently, the two temperature-calibrated feature streams are fed into the spatial attention fusion module. The module first performs global average pooling and max pooling on the channel dimension, respectively, yielding two spatial attention maps. These maps are then concatenated and passed through a lightweight convolution to extract spatial attention, with weights obtained via softmax normalization. These weights are used to apply position-wise weighting to the features, and structural information is preserved by

adding residuals to the input. Finally, the two outputs are aligned in the channel dimension and summed, achieving cross-layer fusion in the channel and spatial dimensions.

3.3. Multi-Scale Gated Attention Module

Given that mid-level features have higher expressive power in capturing local structure and discriminative details of objects, this paper designs an MSGA module for mid-level feature enhancement, as shown in Figure 4. The module takes the features from layer3 of the backbone network as input. It mainly consists of five stages: feature grouping, direction-aware pooling, multi-scale convolution, gated fusion, and cross-spatial attention learning.

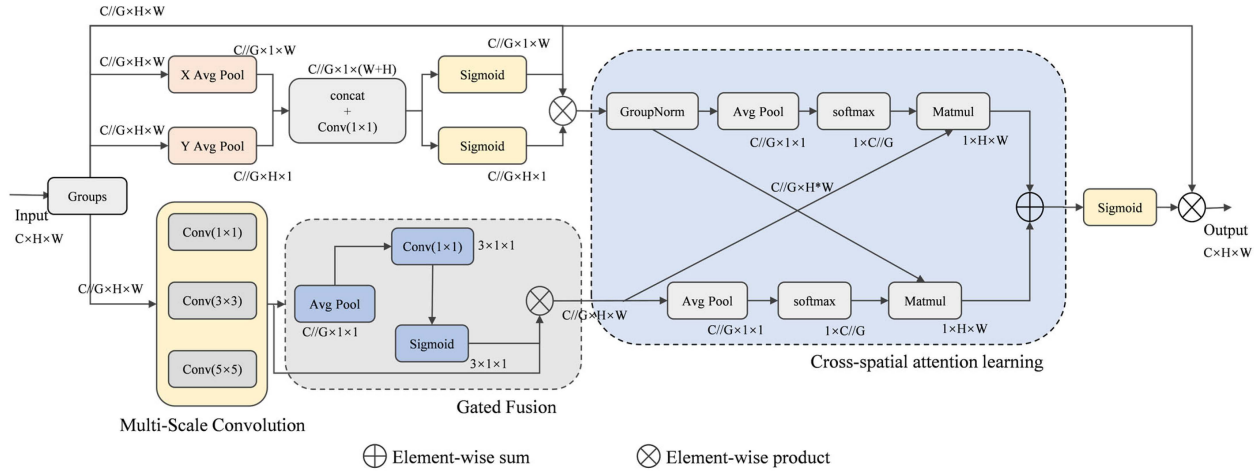


Figure 4. Multi-Scale Gated Attention module structure.

1. Feature grouping: For the input feature map $X \in R^{C \times H \times W}$, MSGA divides it into G sub-feature groups along the channel dimension [25]. Each sub-feature group can be represented as: $X_g \in R^{\frac{C}{G} \times H \times W}$.

2. Direction-aware pooling: Considering that vehicle images exhibit significant structural directionality, such as the horizontal structure of the vehicle body, symmetrical front and rear contours, or symmetrical left and right headlights, the MSGA module introduces a direction-aware pooling mechanism to more effectively model these directional structural features, as shown in Figure 5. This mechanism performs average pooling along the horizontal X-axis and vertical Y-axis directions on the input feature map to extract spatial distribution features corresponding to each direction, thereby generating directional attention maps with structural awareness. Among these, the average pooling operation along the X-axis of the input feature map is expressed as:

$$Z_C^W(W) = \frac{1}{H} \sum_{0 \leq j \leq H} x_c(j, W) \quad (2)$$

Similarly, the average pooling along the Y-axis is expressed as:

$$Z_C^H(H) = \frac{1}{W} \sum_{0 \leq i \leq W} x_c(H, i) \quad (3)$$

where C represents the number of input channels, x_c represents the feature map of the c th channel, and H and W denote the height and width of the feature map, respectively.

3. Multi-scale convolution: The MSGA module introduces parallel multi-scale convolution operations on each sub-feature group, using 1×1 , 3×3 , and 5×5 convolution

kernels to model context information at different scales. This yields feature maps F_1 , F_2 , and F_3 at three distinct scales.

4. Gate-controlled fusion: To perform adaptive multi-scale fusion on the three-scale feature maps generated in the previous step, a lightweight gate-controlled fusion mechanism is designed to emphasize the most discriminative scale features selectively. This mechanism introduces a gate-controlled weight generation module:

$$G = \sigma(\text{Conv}_{1 \times 1}(\text{AvgPool}(X_g))) \in \mathbb{R}^{3 \times 1 \times 1} \quad (4)$$

where $\sigma(\cdot)$ represents the sigmoid activation function, which is used to restrict the weights to the interval $[0, 1]$.

Finally, the fusion of multi-scale features can be expressed as:

$$F_{\text{fused}} = G[0, :, :] \cdot F_1 + G[1, :, :] \cdot F_3 + G[2, :, :] \cdot F_5 \quad (5)$$

This fusion approach avoids the high complexity of standard self-attention, enabling the network to adaptively select between multi-scale receptive fields at near-linear computational cost, thereby more efficiently enhancing component-level details.

5. Cross-spatial attention learning: This stage introduces a cross-attention mechanism to enable effective information exchange between direction-enhanced features and multi-scale fusion features. Specifically, global average pooling is performed on the two feature branches to obtain description vectors with channel-aware capabilities. One branch's description vector serves as the query term to capture the current region of interest, while the other branch acts as the key and value to provide semantic supplementation. By calculating the attention correlation matrix and generating a unified attention map, the original feature maps are weighted and adjusted in the spatial dimension. Due to the differences in direction sensitivity, receptive field size, and context dependency between the two feature branches in their structural design, this cross-spatial bidirectional interaction mechanism enables the model to fully integrate its complementary advantages, ultimately yielding feature representations with greater semantic consistency and more thorough preservation of fine-grained information.

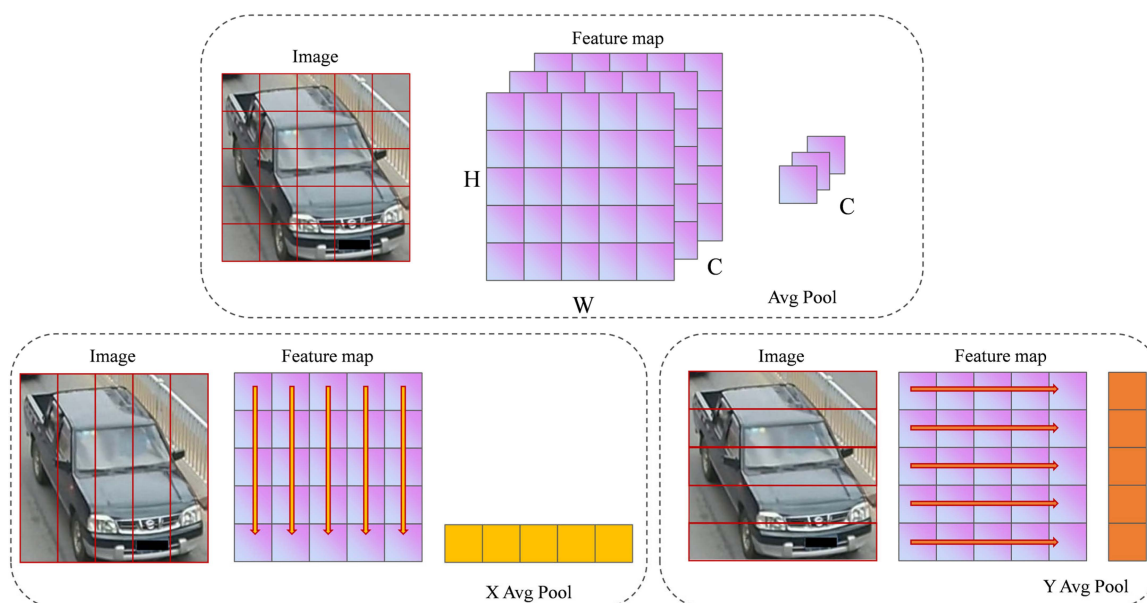


Figure 5. The input feature maps are average-pooled along the horizontal X-axis and vertical Y-axis directions, resulting in a one-dimensional feature vector in each direction.

3.4. Center Loss Function

Unlike most vehicle re-identification tasks that use only traditional triplet loss [26] and cross-entropy loss [27] to optimize the model, this paper introduces center loss [28] to address the issue that these methods cannot directly optimize intra-class compactness. Specifically, let there be K categories in the training set, with each category k corresponding to a learnable feature center vector $c_k \in R^d$, where d is the feature dimension. For a sample feature vector f_i and its corresponding label y_i , the center loss is defined as:

$$L_{center} = \frac{1}{2} \sum_{i=1}^N \|f_i - c_{y_i}\|_2^2 \quad (6)$$

This loss drives similar samples to cluster toward their class center in the feature space, thereby reducing intraclass variance and improving feature separability. During network training, this paper combines the triplet loss $L_{triplet}$, cross-entropy loss L_{ce} , and center loss L_{center} with weighted combinations:

$$L_{total} = L_{triplet} + L_{ce} + \lambda \cdot L_{center} \quad (7)$$

Here, λ is the hyperparameter controlling the loss weight of the control center. By reasonably setting λ , it is possible to further improve the discriminative power of the features without compromising the original optimization objective.

4. Experiment and Results Analysis

4.1. Datasets

To validate the effectiveness of the proposed method, experiments were conducted to evaluate its performance on two publicly available vehicle re-identification datasets: VeRi-776 [29] and VehicleID [3].

VeRi-776: The VeRi-776 dataset contains approximately 50,000 images of 776 vehicles, captured by 20 non-overlapping cameras. Each image is annotated with vehicle ID, camera ID, and auxiliary information such as the vehicle brand, color, and timestamp. The training set includes 37,778 images of 576 vehicles, while the test set includes 11,579 images of 200 vehicles.

VehicleID: The VehicleID dataset is primarily shot from the front or rear of the vehicle, containing 221,763 images of 26,267 vehicles. Each image is labeled with vehicle ID, camera position, vehicle model, and other tag information. The training set includes 110,178 images of 13,134 vehicles, while the test set is divided into three subsets based on the number of vehicles: Small (800 vehicles), Medium (1600 vehicles), and Large (2400 vehicles).

4.2. Evaluation Metrics

This paper uses two mainstream evaluation metrics, mean average precision (mAP) and cumulative matching characteristics (CMC), to assess DAFF-Net's performance [30].

mAP: mAP is a key metric for measuring retrieval performance in vehicle re-identification, which comprehensively considers the accuracy and recall rate of retrieval results. The calculation formula is as follows:

$$mAP = \frac{1}{Q} \sum_{q=1}^Q AP(q) \quad (8)$$

where Q represents the total number of images in the query set, and $AP(q)$ represents the average accuracy of the q th query image. The calculation formula is as follows:

$$AP = \frac{1}{N} \sum_{k=1}^n p(k) \cdot gt(k) \quad (9)$$

where N denotes the number of true matching samples related to the query image, n denotes the total number of images in the query set, $p(k)$ denotes the accuracy of the top k search results, and $gt(k)$ denotes whether there are true matching samples of the query image among the top k search results.

CMC: The CMC curve is used to evaluate the ranking performance of identification results, reflecting the probability that the first correct match appears among the top k search results. This paper mainly uses two indicators, CMC@1 and CMC@5, which represent the accuracy rate of correct matches appearing among the top 1 and top 5 search results, respectively. In the figures, CMC@1 and CMC@5 are labeled as Rank-1 and Rank-5.

$$CMC@K = \frac{1}{Q} \sum_{q=1}^Q gt(q, k) \quad (10)$$

4.3. Experiment Settings

Before training, all input images are resized to 256 pixels $\times$ 256 pixels. Data preprocessing and augmentation sequentially include: random horizontal flipping ($p = 0.5$) [31], zero-padding by 10 pixels followed by random cropping back to 256 pixels $\times$ 256 pixels, tensorized and normalized according to fixed channel means and standard deviations (means [0.485, 0.456, 0.406], standard deviations [0.229, 0.224, 0.225]), and finally applying random erasure ($p = 0.5$) [32]. The backbone network employs ResNet50-ibn, with BNNeck enabled during training and after-BN features used for testing. The Adam [33] optimizer is used during training to facilitate effective parameter updates for the model. The batch size is set to 16, and the total number of training epochs is set to 120. To prevent overfitting, a weight decay strategy is employed, with the base learning rate set to 5×10^{-5} , decreasing to 5×10^{-6} at the 40th epoch and further decreasing to 5×10^{-7} at the 70th epoch. During both the training and testing phases of the model, Euclidean distance is used as the metric to measure the similarity between the query image and the retrieved images in the database.

4.4. Comparison with Different Mainstream Models

To validate the effectiveness of the proposed network in vehicle re-identification tasks, this paper conducted comparative experiments with various representative methods on the mainstream datasets VeRi-776 and VehicleID.

1. Experiments on the VeRi-776 dataset: This paper uses mAP, CMC@1, and CMC@5 as evaluation metrics on the VeRi-776 dataset. The specific experimental results are shown in Table 1.

In Table 1, AAVR, PRN, and PVEN leverage auxiliary clues to optimize feature learning for re-identification. Among these, PVEN utilizes component-level semantic segmentation with viewpoint alignment and attention enhancement to improve representations under conventional viewpoints. However, it heavily relies on mask quality and alignment accuracy, resulting in only moderate re-identification performance. In addition, methods such as HPGN, LSFR, and ASSEN also demonstrate competitive performance, with mAP values above 80% and CMC@1 values above 96%. In particular, the ASSEN network, which learns discriminative vehicle features through attribute enhancement and spatiotemporal weakening, achieves 81.3% mAP and 96.9% CMC@1, demonstrating outstanding performance. However, DAFF-Net achieved further improvements in the above

two key metrics, with mAP increasing to 82.2% and CMC@1 increasing to 97.5%. The TransReID model, based on the Transformer architecture, achieves feature representation through global self-attention across image tokens, yielding a commendable mAP of 80.6%, but DAFF-Net outperforms TransReID on this metric. This proves that the multi-scale feature fusion strategy based on dual-branch attention guidance enhances the expressive richness of features, thereby surpassing current mainstream methods in multiple evaluation metrics.

Table 1. Comparison with other methods in terms of mAP, CMC@1, and CMC@5 on the VeRi-776 dataset.

Method	mAP	CMC@1	CMC@5
AAVER [34]	0.612	0.890	0.947
PRN [7]	0.743	0.943	0.989
MSDeep [35]	0.745	0.951	-
PVEN [6]	0.794	0.956	0.984
TBE-Net [2]	0.795	0.960	0.985
HPGN [36]	0.802	0.967	-
LSFR [11]	0.808	0.964	-
TransReID [37]	0.806	0.969	-
ASSEN [13]	0.813	0.969	-
DAFF-Net (Ours)	0.822	0.975	0.982

2. Experiments on the VehicleID dataset: Since each query image has only one matching target in the VehicleID dataset, mAP cannot be effectively evaluated, so only CMC@1 and CMC@5 are used as evaluation metrics. Table 2 shows the experimental results of each method on the VehicleID dataset.

Table 2. Comparison with other methods in terms of CMC@1 and CMC@5 on the VehicleID dataset.

Method	Small		Medium		Large	
	CMC@1	CMC@5	CMC@1	CMC@5	CMC@1	CMC@5
AAVER [34]	0.747	0.938	0.686	0.900	0.635	0.856
PRN [7]	0.784	0.923	0.750	0.883	0.742	0.864
MSDeep [35]	0.812	0.954	0.780	0.918	0.756	0.893
CFVMNet [38]	0.814	0.941	0.773	0.904	0.747	0.887
HPGN [36]	0.839	-	0.799	-	0.773	-
PVEN [6]	0.847	0.970	0.806	0.945	0.778	0.920
TBE-Net [2]	0.860	0.984	0.823	0.966	0.807	0.949
MsKAT [39]	0.863	0.974	0.818	0.955	0.794	0.939
GLNet [1]	0.872	0.978	0.829	0.956	0.803	0.934
DAFF-Net (Ours)	0.907	0.972	0.846	0.965	0.821	0.956

As shown in Table 2, DAFF-Net achieved CMC@1 of 90.7%, 84.6%, and 82.1% on the Small, Medium, and Large subsets, respectively, demonstrating the best performance among all comparison methods. Among these, HPGN, as a graph model representative, achieves re-identification by constructing hierarchical component relationship graphs and propagating information through graph convolutions. However, it failed to achieve satisfactory results across all three subsets. TBE-Net is a three-branch component-complementary framework with strong local modeling capabilities; however, it is relatively dependent on the quality of component segmentation, as localization errors may compromise fusion stability. Although it achieved the best CMC@5 results on Small and Medium subsets, DAFF-Net outperformed TBE-Net by 4.7% and 2.3% on the CMC@1 metric. Meanwhile, recently proposed methods such as MsKAT and GLNet demonstrate excellent performance across the three subsets. MsKAT constructs a multi-scale knowledge-aware Transformer architecture and introduces a knowledge-guided alignment loss function, enabling it to

achieve superior recognition performance across different scale divisions; GLNet integrates a semantic segmentation module and fuses an adaptive local attention mechanism to enhance the model's adaptability to vehicle re-identification tasks. Compared to the aforementioned methods, DAFF-Net achieves direction-aware local modeling through convergence along height and width dimensions and gated scale selection after cross-layer alignment. This enhances feature-relevant information without relying on explicit feature detection. This design achieves a more reasonable balance between cross-layer alignment, local multi-scale processing, directional selectivity, efficiency, and robustness.

4.5. Ablation Experiments and Analysis

4.5.1. Module Performance Analysis

To further validate the effectiveness of each module in DAFF-Net, ablation experiments were conducted on the VeRi-776 dataset to assess the performance changes after removing specific modules. Table 3 shows the performance comparison results of the baseline model after removing the MSGA module (DAFF-Net w/o MSGA) and the TCAF module (DAFF-Net w/o TCAF).

Table 3. DAFF-Net module ablation experiment results.

Method	mAP	CMC@1	CMC@5
Baseline	0.778	0.963	0.978
DAFF-Net w/o MSGA	0.808	0.969	0.980
DAFF-Net w/o TCAF	0.803	0.968	0.983
DAFF-Net	0.822	0.975	0.982

As shown in Table 3, the baseline model achieves an mAP of 77.8%, CMC@1 of 96.3%, and CMC@5 of 97.8% without any attention fusion modules. When the MSGA module is introduced but the TCAF module is removed (DAFF-Net w/o TCAF), the model performance significantly improves, achieving an mAP of 80.8%, CMC@1 of 96.9%, and CMC@5 of 98.0%, indicating that the MSGA module effectively enhances the multi-scale modeling capabilities of intermediate-level features. Similarly, when only the TCAF module is retained and the MSGA module is removed (DAFF-Net w/o MSGA), the model performance also improves to some extent, with mAP of 80.3%, CMC@1 of 96.8%, and CMC@5 of 98.3%, indicating that the TCAF module facilitates the effective integration of cross-layer information after fusing shallow and deep features. The complete DAFF-Net, which incorporates both the MSGA and TCAF modules, achieves optimal performance on both evaluation metrics, with an mAP of 82.2% and CMC@1 of 97.5%, further validating the synergistic benefits of the two modules in terms of feature representation capability and vehicle discrimination performance.

4.5.2. Experiments on the Combination Methods and Fusion Strategies of the TCAF

To validate the effectiveness of the proposed Temperature-Calibration Attention Fusion Module (TCAF) in cross-layer feature fusion, this paper designed the following ablation experiments to compare the impact of different feature combination methods and fusion strategies at various layers on the final performance. In the experiments, features from different layers of the backbone network (layer1, layer2, and layer3) were fused with those from layer4, and the effects of two strategies—direct addition and fusion via the TCAF module—were compared to evaluate the impact of different configurations on model performance. The experimental results are shown in Table 4.

Table 4. Ablation experiment results of the TCAF module under different fusion strategies.

Method	mAP	CMC@1	CMC@5
Sum (layer1 + layer4)	0.804	0.963	0.983
TCAF (layer1 + layer4)	0.813	0.973	0.985
TCAF (layer2 + layer4)	0.808	0.968	0.986
TCAF (layer3 + layer4)	0.805	0.969	0.981

As can be observed from the table, when the features from layer1 and layer4 are fused using the TCAF module, the model achieves optimal performance, with mAP reaching 81.3% and CMC@1 improving to 97.3%. This represents a significant improvement compared to the direct addition fusion method, validating the effectiveness of TCAF in guiding cross-layer feature fusion. Additionally, when layer2 or layer3 features are fused with layer4 features, performance slightly decreases. This indicates that TCAF is more sensitive to the integration of information between shallow and deep features, enabling it to exploit their potential complementary relationships more fully.

4.5.3. MSGA Module Multi-Scale Convolutional Ensemble Ablation Experiment

This paper conducts ablation experiments on the combination of multi-scale convolution structures in the Multi-Scale Gated Attention Module (MSGA), investigating the impact of different convolution combinations under varying receptive fields on vehicle re-identification performance. A total of seven different configurations were designed, including using a single-scale convolution kernel (1×1 , 3×3 , 5×5), any combination of two convolution kernels ($1 \times 1 + 3 \times 3$, $1 \times 1 + 5 \times 5$, $3 \times 3 + 5 \times 5$), and all combinations of the three convolution kernels (Conv-All), with the latter being the final complete design adopted.

As shown in the results of Table 5, the 3×3 convolution kernel performed best when used alone in a single-scale configuration, achieving an mAP of 81.3%, indicating that the medium receptive field has strong feature extraction capabilities. In the two convolution kernel combination configurations, both $1 \times 1 + 3 \times 3$ and $3 \times 3 + 5 \times 5$ achieved a good mAP of 81.5%, further validating the effectiveness of multi-scale feature fusion. Ultimately, the entire Conv-All configuration achieved the best results across all three metrics, with an mAP of 81.9%, a CMC@1 of 97.4%, and a CMC@5 of 98.5%, indicating that the joint modeling of multi-scale information and the gated fusion mechanism significantly enhances the model's discriminative capability.

Table 5. Ablation experiment results of the MSGA module multi-scale convolution combination method.

Method	mAP	CMC@1	CMC@5
Conv- 1×1	0.791	0.963	0.980
Conv- 3×3	0.813	0.972	0.982
Conv- 5×5	0.810	0.973	0.982
Conv- $1 \times 1 + 3 \times 3$	0.815	0.973	0.984
Conv- $1 \times 1 + 5 \times 5$	0.808	0.969	0.980
Conv- $3 \times 3 + 5 \times 5$	0.815	0.970	0.983
Conv-All	0.819	0.974	0.985

4.5.4. Experiment on the Selection of Hyperparameter λ in Center Loss

To evaluate the impact of the weight hyperparameter λ in the loss function on model performance, this paper sets multiple different λ values while keeping the model structure and other loss terms unchanged to conduct ablation experiments. The experimental results are shown in Table 6.

Table 6. Ablation experiment results on the performance of the center loss weight λ .

$\lambda(\times 10^{-3})$	mAP	CMC@1	CMC@5
0	0.781	0.954	0.979
0.25	0.798	0.957	0.983
0.5	0.813	0.972	0.986
0.75	0.811	0.968	0.983
1	0.807	0.963	0.981

As can be seen from the results in the table, when $\lambda = 0$, i.e., when no center loss is introduced, the model performs relatively poorly in terms of mAP and CMC@1. As λ increases, the model's performance shows an upward trend, reaching its best performance at $\lambda = 0.5$, with mAP at 81.3% and CMC@1 at 97.2%. As λ continues to increase, the model's performance slightly declines, particularly at $\lambda = 1.0$, where CMC@1 drops to 96.3%. The experimental results indicate that the weight of the center loss has a certain impact on model training. Too small a λ value may result in insufficient feature aggregation capability, while too large a λ value may affect inter-class separation, impacting the model's overall performance. In this experiment, $\lambda = 0.5 \times 10^{-3}$ was set as the optimal weight parameter.

4.6. Visualization of Experimental Results

Figure 6 shows the retrieval visualization results of the baseline and the DAFF-Net proposed in this paper on the VeRi-776 dataset. Each set of results includes a query image and the corresponding top 10 retrieval results, with red borders indicating incorrectly retrieved images. In scenarios with similar visual distractions (such as identical colors or similar contours), multiple mismatches appear among the top results of the baseline model. Under identical queries, DAFF-Net significantly reduces such early-stage false positives, enabling correct samples of the same vehicle to rank higher earlier. This phenomenon is observable in both light-colored and dark-colored vehicles in the figure, indicating that the proposed method is more robust to fine-grained distinguishing cues and less sensitive to interference from vehicle silhouette and color similarity. Figure 7 presents an attention drift heatmap, illustrating the evolution of attention distribution for the same query image across the backbone network's layer3 and layer4, as well as two enhancement modules: TCAF and MSGA. From layer3 to layer4, attention gradually converges from dispersed patterns onto the vehicle body and front region. After processing through the TCAF module, the focus on background areas diminishes. Finally, the MSGA module further compacts the hotspots. This overall progression demonstrates a shift in attention toward vehicle identity-related components, highlighting these modules' role in enhancing focus on discriminative regions.

**Figure 6.** Retrieval examples on VeRi-776 for the baseline and DAFF-Net. For each query, the top 10 gallery images are ranked from left to right by Euclidean distance; red boxes indicate false matches.

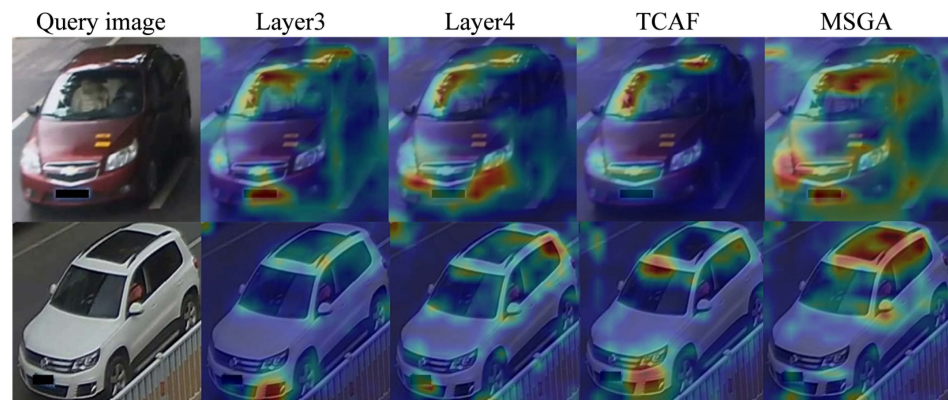


Figure 7. Attention heatmaps across backbone depth and modules. Warmer colors denote stronger attention.

To validate the effectiveness of DAFF-Net in vehicle re-identification tasks, this paper compares the cumulative matching feature (CMC) curve performance between the baseline and DAFF-Net. As shown in Figure 8, DAFF-Net outperforms the baseline model in both CMC@1 and CMC@5 metrics, fully demonstrating that the design based on the multi-scale attention fusion strategy effectively enhances the model’s ability to capture vehicle detail information, thereby improving the overall retrieval performance in vehicle re-identification tasks.

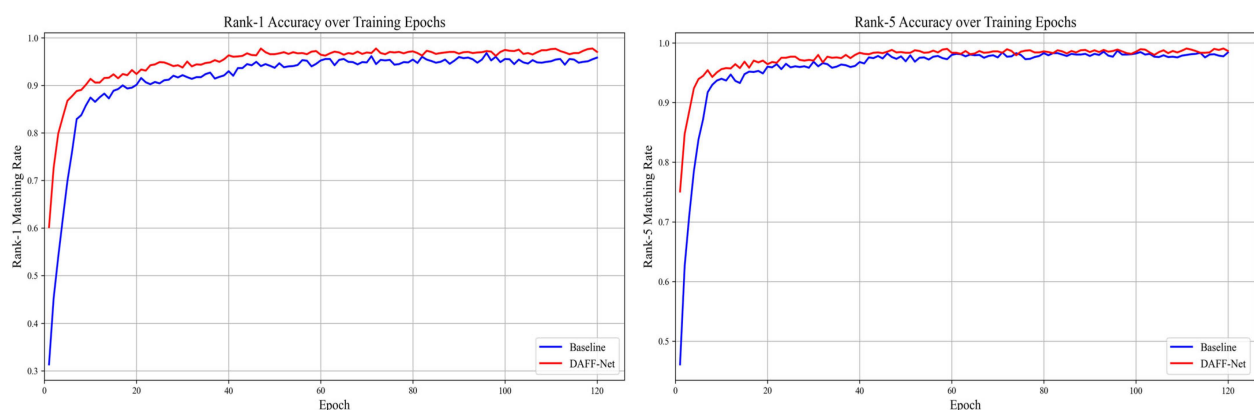


Figure 8. Training curves comparing baseline and DAFF-Net. The left panel shows CMC@1 across epochs. The right panel shows CMC@5 across epochs.

5. Conclusions

This paper addresses the challenges of extracting and fusing local features in vehicle re-identification tasks by proposing a feature fusion network based on dual-branch attention guidance—DAFF-Net. The network uses ResNet50-ibn as its backbone, with separate branches for cross-layer feature fusion and multi-scale enhancement at the intermediate layer. By introducing the Temperature-Calibration Attention Fusion module and the Multi-Scale Gated Attention module into these two branches, the network achieves synergistic optimization of cross-layer information integration and multi-scale contextual modeling, thereby enhancing the expressive power of features. Finally, the obtained features are optimized using a triplet loss, center loss, and cross-entropy loss, further improving the clustering and discriminative properties of the feature space. Experimental results on the public datasets VeRi-776 and VehicleID demonstrate that DAFF-Net outperforms existing mainstream methods in metrics such as mAP, CMC@1, and CMC@5, fully validating the effectiveness of the proposed method.

Although the method proposed in this paper achieves superior performance in vehicle re-identification tasks, some shortcomings still warrant further research and improvement. On one hand, DAFF-Net introduces an attention mechanism in its dual-branch structure to enhance the expressive capability of multi-level features, thereby improving the model's discriminative power. However, in practical applications, the distribution of attention weights is susceptible to factors such as background interference. This leads to situations where the focus area deviates from the target's critical regions, limiting the effective extraction of local detail features. On the other hand, DAFF-Net primarily relies on static image appearance features for re-identification and cannot model relatively stable auxiliary information, such as vehicle attribute information and spatiotemporal information. This results in performance degradation in vehicle re-identification under severe occlusion or cross-domain application conditions. Future research efforts could further focus on enhancing the semantic alignment capabilities of attention mechanisms to improve their stability and interpretability in complex scenarios. Meanwhile, multimodal information and cross-domain adaptation mechanisms can be incorporated to enhance the model's generalization capability in real-world scenarios. Alternatively, one could draw upon relevant research on transport trucks to conduct typological assessments in both truck and public transit scenarios [40], providing several retrieval examples to accommodate better practical application needs such as urban traffic management.

Author Contributions: Conceptualization, Y.G. and G.Y.; methodology, Y.G. and G.Y.; software, Y.G.; validation, Y.G. and W.L.; formal analysis, Y.G. and H.L.; investigation, W.L.; resources, W.L.; data curation, Y.G.; writing—original draft preparation, Y.G. and G.Y.; writing—review and editing, Y.G. and H.L.; visualization, Y.G.; supervision, Y.G.; project administration, Y.G. and H.L.; funding acquisition, Y.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by 2023 Opening Research Fund of Yunnan Key Laboratory of Digital Communications (YNKLDC-KFKT-202303).

Data Availability Statement: The raw data supporting the conclusions of this article will be made available by the authors on request.

Conflicts of Interest: We declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

References

1. Yu, H. Research on Vehicle Re-Identification Methods Based on Multi-Perception of Global and Local Features. Master's Thesis, Shandong Technology and Business University, Yantai, China, 2024. (In Chinese).
2. Sun, W.; Dai, G.; Zhang, X.; He, X.; Chen, X. TBE-Net: A three-branch embedding network with part-aware ability and feature complementary learning for vehicle re-identification. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 14557–14569. [\[CrossRef\]](#)
3. Liu, H.; Tian, Y.; Yang, 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, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 2167–2175. [\[CrossRef\]](#)
4. Bai, Y.; Lou, Y.; Gao, F.; Wang, S.; Wu, Y.; Duan, L.-Y. Group-sensitive triplet embedding for vehicle reidentification. *IEEE Trans. Multimed.* **2018**, *20*, 2385–2399. [\[CrossRef\]](#)
5. Chu, R.; Sun, Y.; Li, Y.; Zhang, C.; Wei, Y. Vehicle re-identification with viewpoint-aware metric learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 8282–8291. [\[CrossRef\]](#)
6. Zhang, X.; Zhang, R.; Cao, J.; Gong, D.; You, M.; Shen, C. Part-guided attention learning for vehicle instance retrieval. *IEEE Trans. Intell. Transp. Syst.* **2020**, *23*, 3048–3060. [\[CrossRef\]](#)
7. He, B.; Li, J.; Zhao, Y.; Tian, Y. Part-regularized near-duplicate vehicle re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3997–4005. [\[CrossRef\]](#)
8. Ping, C.; Li, L.; Liu, D.; Lin, H.; Shi, J. Research Progress on Vehicle Re-Identification Based on Deep Learning. *Comput. Eng. Appl.* **2025**, *61*, 1–26. (In Chinese)

9. Li, M.; Huang, X.; Zhang, Z. Self-supervised Geometric Features Discovery via Interpretable Attention for Vehicle Re-Identification and Beyond (Complete Version). *arXiv* **2023**, arXiv:2303.11169.
10. Lee, S.; Woo, T.; Lee, S.H. Multi-attention-based soft partition network for vehicle re-identification. *J. Comput. Des. Eng.* **2023**, *10*, 488–502. [\[CrossRef\]](#)
11. Liang, Y.; Gao, Y.; Shen, Z.Y. Transformer vehicle re-identification of intelligent transportation system under carbon neutral target. *Comput. Ind. Eng.* **2023**, *185*, 109619. [\[CrossRef\]](#)
12. Zhang, H.; Kuang, Z.; Cheng, L.; Liu, Y.; Ding, X.; Huang, Y. AIVR-Net: Attribute-based invariant visual representation learning for vehicle re-identification. *Knowl.-Based Syst.* **2024**, *289*, 111455. [\[CrossRef\]](#)
13. Li, H.; Li, C.; Zheng, A.; Tang, J.; Luo, B. Attribute and state guided structural embedding network for vehicle re-identification. *IEEE Trans. Image Process.* **2022**, *31*, 5949–5962. [\[CrossRef\]](#)
14. Yu, Z.; Pei, J.; Zhu, M.; Zhang, J.; Li, J. Multi-attribute adaptive aggregation transformer for vehicle re-identification. *Inf. Process. Manag.* **2022**, *59*, 102868. [\[CrossRef\]](#)
15. Quispe, R.; Lan, C.; Zeng, W.; Pedrini, H. Attributenet: Attribute enhanced vehicle re-identification. *Neurocomputing* **2021**, *465*, 84–92. [\[CrossRef\]](#)
16. Wang, H.; Peng, J.; Chen, D.; Jiang, G.; Zhao, T.; Fu, X. Attribute-guided feature learning network for vehicle reidentification. *IEEE Multimed.* **2020**, *27*, 112–121. [\[CrossRef\]](#)
17. Qian, J.; Jiang, W.; Luo, H.; Yu, H. Stripe-based and attribute-aware network: A two-branch deep model for vehicle re-identification. *Meas. Sci. Technol.* **2020**, *31*, 095401. [\[CrossRef\]](#)
18. Zheng, Z.; Ruan, T.; Wei, Y.; Yang, Y.; Mei, T. VehicleNet: Learning robust visual representation for vehicle re-identification. *IEEE Trans. Multimed.* **2020**, *23*, 2683–2693. [\[CrossRef\]](#)
19. Tu, J.; Chen, C.; Huang, X.; He, J.; Guan, X. Discriminative feature representation with spatio-temporal cues for vehicle re-identification. *arXiv* **2020**, arXiv:2011.06852. [\[CrossRef\]](#)
20. 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*, 109304. [\[CrossRef\]](#)
21. Li, Z.; Zhang, X.; Tian, C.; Gao, X.; Gong, Y.; Wu, J.; Zhang, G.; Li, J.; Liu, H. TVG-ReID: Transformer-based vehicle-graph re-identification. *IEEE Trans. Intell. Veh.* **2023**, *8*, 4644–4652. [\[CrossRef\]](#)
22. Meng, D.; Li, L.; Liu, X.; Li, Y.; Yang, S.; Zha, Z.-J.; Gao, X.; Wang, S.; Huang, Q. Parsing-based view-aware embedding network for vehicle re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 7103–7112. [\[CrossRef\]](#)
23. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 770–778. [\[CrossRef\]](#)
24. Zhou, H.; Luo, F.; Zhuang, H.; Weng, Z.; Gong, X.; Lin, Z. Attention multihop graph and multiscale convolutional fusion network for hyperspectral image classification. *IEEE Trans. Geosci. Remote Sens.* **2023**, *61*, 1–14. [\[CrossRef\]](#)
25. Ouyang, D.; He, S.; Zhang, G.; Luo, M.; Guo, H.; Zhan, J.; Huang, Z. Efficient multi-scale attention module with cross-spatial learning. In Proceedings of the ICASSP 2023—2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Rhodes, Greece, 4–10 June 2023; IEEE: Piscataway, NJ, USA, 2023; pp. 1–5. [\[CrossRef\]](#)
26. Hermans, A.; Beyer, L.; Leibe, B. In defense of the triplet loss for person re-identification. *arXiv* **2017**, arXiv:1703.07737. [\[CrossRef\]](#)
27. Zhang, Z.; Sabuncu, M. Generalized cross entropy loss for training deep neural networks with noisy labels. *Adv. Neural Inf. Process. Syst.* **2018**, *31*, 8792–8802.
28. Wen, Y.; Zhang, K.; Li, Z.; Qiao, Y. A discriminative feature learning approach for deep face recognition. In Proceedings of the Computer Vision—ECCV 2016, 14th European Conference, Amsterdam, The Netherlands, 11–14 October 2016; Proceedings, part VII 14. Springer International Publishing: Cham, Switzerland, 2016; pp. 499–515. [\[CrossRef\]](#)
29. Liu, X.; Liu, W.; Mei, T.; Ma, H. Provid: Progressive and multimodal vehicle reidentification for large-scale urban surveillance. *IEEE Trans. Multimed.* **2017**, *20*, 645–658. [\[CrossRef\]](#)
30. Zheng, L.; Shen, L.; Tian, L.; Wang, S.; Wang, J.; Tian, Q. Scalable person re-identification: A benchmark. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 1116–1124. [\[CrossRef\]](#)
31. Krizhevsky, A.; Sutskever, I.; Hinton, G.E. Imagenet classification with deep convolutional neural networks. *Adv. Neural Inf. Process. Syst.* **2012**, *25*, 1106–1114. [\[CrossRef\]](#)
32. Shin, H.C.; Roth, H.R.; Gao, M.; Lu, L.; Xu, Z.; Nogues, I.; Yao, J.; Mollura, D.; Summers, R.M. Deep convolutional neural networks for computer-aided detection: CNN architectures, dataset characteristics and transfer learning. *IEEE Trans. Med. Imaging* **2016**, *35*, 1285–1298. [\[CrossRef\]](#) [\[PubMed\]](#)
33. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. *arXiv* **2014**, arXiv:1412.6980.
34. Khorramshahi, P.; Kumar, A.; Peri, N.; Rambhatla, S.S.; Chen, J.-C.; Chellappa, R. A dual-path model with adaptive attention for vehicle re-identification. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Republic of Korea, 27 October–2 November 2019; pp. 6132–6141. [\[CrossRef\]](#)

35. Cheng, Y.; Zhang, C.; Gu, K.; Qi, L.; Gan, Z.; Zhang, W. Multi-scale deep feature fusion for vehicle re-identification. In Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), Barcelona, Spain, 4–8 May 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1928–1932. [\[CrossRef\]](#)
36. Shen, F.; Zhu, J.; Zhu, X.; Xie, Y.; Huang, J. Exploring spatial significance via hybrid pyramidal graph network for vehicle re-identification. *IEEE Trans. Intell. Transp. Syst.* **2021**, *23*, 8793–8804. [\[CrossRef\]](#)
37. He, S.; Luo, H.; Wang, P.; Wang, F.; Li, H.; Jiang, W. Transreid: Transformer-based object re-identification. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 11–17 October 2021; pp. 15013–15022. [\[CrossRef\]](#)
38. Sun, Z.; Nie, X.; Xi, X.; Yin, Y. CFVMNet: A multi-branch network for vehicle re-identification based on common field of view. In Proceedings of the 28th ACM International Conference on Multimedia, Seattle, WA, USA, 12–16 October 2020; pp. 3523–3531. [\[CrossRef\]](#)
39. Li, H.; Li, C.; Zheng, A.; Tang, J.; Luo, B. MsKAT: Multi-scale knowledge-aware transformer for vehicle re-identification. *IEEE Trans. Intell. Transp. Syst.* **2022**, *23*, 19557–19568. [\[CrossRef\]](#)
40. Zou, Y.; Hu, Q.; Han, W.; Zhang, S.; Chen, Y. Analyzing Travel and Emission Characteristics of Hazardous Material Transportation Trucks Using BeiDou Satellite Navigation System Data. *Remote Sens.* **2025**, *17*, 423. [\[CrossRef\]](#)

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.