

Article

Vehicle Re-Identification Method Based on Multi-Task Learning in Foggy Scenarios

Wenchao Gao , Yifan Chen , Chuanrui Cui and Chi Tian

School of Artificial Intelligence, China University of Mining & Technology-Beijing, Beijing 100083, China; cyf08132024@163.com (Y.C.); zqt2310405024@student.cumtb.edu.cn (C.C.); 17854333900@163.com (C.T.)

* Correspondence: gaowc@cumtb.edu.cn; Tel.: +86-13811786356

Abstract: Vehicle re-identification employs computer vision to determine the presence of specific vehicles in images or video sequences, often using vehicle appearance for identification due to the challenge of capturing complete license plate information. Addressing the performance issues caused by fog, such as image blur and loss of key positional information, this paper introduces a multi-task learning framework incorporating a multi-scale fusion defogging method (MsF). This method effectively mitigates image blur to produce clearer images, which are then processed by the re-identification branch. Additionally, a phase attention mechanism is introduced to adaptively preserve crucial details. Utilizing advanced artificial intelligence techniques and deep learning algorithms, the framework is evaluated on both synthetic and real datasets, showing significant improvements in mean average precision (mAP)—an increase of 2.5% to 87.8% on the synthetic dataset and 1.4% to 84.1% on the real dataset. These enhancements demonstrate the method's superior performance over the semi-supervised joint defogging learning (SJDL) model, particularly under challenging foggy conditions, thus enhancing vehicle re-identification accuracy and deepening the understanding of applying multi-task learning frameworks in adverse visual environments.

Keywords: vehicle re-identification; foggy scenarios; multi-task learning framework; phase attention mechanism

MSC: 68T45; 68U10



Citation: Gao, W.; Chen, Y.; Cui, C.; Tian, C. Vehicle Re-Identification Method Based on Multi-Task Learning in Foggy Scenarios. *Mathematics* **2024**, *12*, 2247. <https://doi.org/10.3390/math12142247>

Academic Editor: Jonathan Blackledge

Received: 17 June 2024

Revised: 10 July 2024

Accepted: 11 July 2024

Published: 19 July 2024



Copyright: © 2024 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/>).

1. Introduction

The rapid escalation in global motor vehicle numbers, particularly in urban areas, has challenged traditional traffic management methods, revealing their limitations. This increase underscores the growing importance of vehicle re-identification—a field that leverages advancements in computer vision and machine learning to enhance intelligent transportation systems [1]. Vehicle re-identification [2], pivotal in these systems, was initially developed to track and analyze vehicle movements. It is now essential not only for managing vehicle flow but also for ensuring the accuracy of vehicle identification in various environmental conditions [3,4].

Vehicle re-identification becomes particularly critical in scenarios where vehicle attributes such as license plates are obscured or unreadable due to factors like distance, angle, or adverse weather conditions such as fog [5]. These situations necessitate robust systems that can identify vehicles based on appearance alone, making it crucial to develop methods that can operate effectively in less than ideal visual conditions [6]. Traditional methods, such as license plate recognition, though foundational in early vehicle identification systems, prove effective only under clear conditions and falter in suboptimal environments like fog. To overcome these limitations, contemporary approaches employ artificial fog in network training and combine metric learning with object detection. However, these methods typically require a cumbersome two-step process involving first removing the fog, then

identifying the vehicle, which leads to slower inference times and significant challenges. These include the inability to simultaneously remove fog and identify vehicles, a learning bias toward synthetic data over real data, and the loss of crucial vehicle details during the fog removal process, all of which complicate re-identification [7]. Therefore, addressing these issues requires the development of more sophisticated, integrated technologies [8]. These technologies reflect the ongoing evolution of vehicle re-identification, transitioning from simple plate recognition to complex, environment-adaptive systems capable of efficiently and accurately processing vehicle identities under varying conditions [9].

To address the limitations of traditional vehicle re-identification methods and meet the advanced requirements of modern intelligent transportation systems, this article introduces a novel multi-task learning framework. Among the algorithms employed, convolutional neural networks form the backbone of both the defogging and re-identification networks, leveraging their ability to learn high-level features from complex visual data. Specifically, the defogging network integrates adaptive thresholding techniques to differentiate between foreground moving objects and foggy backgrounds effectively, enhancing the subsequent feature extraction process. Furthermore, the re-identification network utilizes a combination of feature pyramid networks (FPNs) and region-based convolutional neural networks (R-CNNs) to improve the accuracy of vehicle detection and classification across different scales. This is particularly beneficial in urban settings, where vehicles may appear at various distances and angles due to congested traffic conditions. The FPN architecture ensures that the network captures precise details from multi-scale images, which are crucial for identifying vehicles accurately in heavily obscured scenarios.

Additionally, the framework incorporates attention mechanisms to refine the feature maps generated by the CNNs, focusing on critical areas within an image that likely contain the vehicle. This selective attention helps to mitigate the effects of variable lighting and weather conditions, which can distort the visual appearance of vehicles. By dynamically adjusting the focus of the neural network based on the contextual information within the scene, the framework enhances its ability to discern and track vehicles reliably.

These advanced techniques collectively contribute to a robust vehicle re-identification system that not only improves the clarity and precision of vehicle images in foggy conditions, but also ensures high reliability and real-time processing capabilities. The integration of these sophisticated methods within a multi-task learning framework marks a significant advancement over traditional vehicle re-identification approaches, offering a comprehensive solution that addresses the dynamic challenges of modern intelligent transportation systems.

We outline the contributions of this article as follows:

- (1) This paper proposes a novel multi-task learning framework that synergistically integrates a defogging network and a re-identification network. As shown in Figure 1, the defogging and re-identification branches concurrently perform their respective tasks, thereby improving re-identification accuracy under foggy conditions.
- (2) To mitigate the loss of vehicle information due to fog removal, which impacts re-identification accuracy, a phase attention mechanism is employed to extract weighted feature map information. This is then combined with the original image to discern key vehicle information across various dimensions, such as channel and spatial levels, ultimately improving re-identification quality [10].
- (3) This paper introduces a multi-scale fusion module (Msf) designed specifically for superior fog removal. The Msf module fuses features from initial fog removal, utilizing multi-scale feature fusion and a specially designed loss function to improve both the fog removal network and the quality of image defogging.

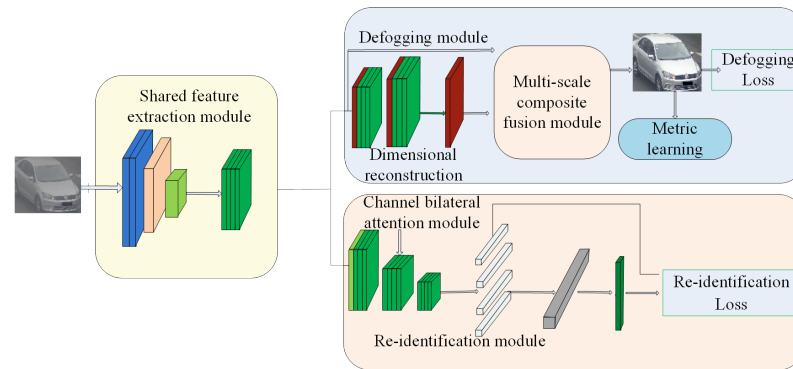


Figure 1. Structure diagram of the fogging re-identification network.

2. Related Works

With the advancement of deep convolutional neural networks (DCNN) and the release of extensive benchmarks (e.g., Vehicle ID, VeRi-776, VERI-Wild, Vehicle-1M) [11–13], vehicle re-identification has attracted increased attention, resulting in numerous methods that demonstrate excellent performance. Vehicle re-identification technology utilizing DCNN can be categorized into three types. First, the use of meta-information to enhance the fusion of embedded features. Ye et al. [14] discuss using transformers to capture global dependencies and local features more effectively than traditional CNNs. He et al. [15] utilize the pure transformer network as a backbone to extract image features. The jigsaw patch module (JPM) rearranges embedded sequences and generates robust features through displacement and patch shuffling. Roman-Jimenez et al. [16] explore vehicle re-identification by evaluating distance metrics in CNN latent spaces, showing that MCD-based track-to-track processing with DenseNet201 significantly improves accuracy compared with conventional approaches. Building on these concepts, Zheng et al. [17] introduce the viewpoint-conditioned network that leverages vehicle attributes and spatio-temporal information to learn comprehensive global vehicle representations. This method significantly enhances re-identification accuracy by effectively handling variations from different viewpoints, demonstrating the power of integrating diverse meta-information in improving feature fusion.

The second category focuses on end-to-end training that combines local and global constraints. He B et al. [18] combine component-based regularization with a global re-identification module. He S et al. [19] preserve partially regularized discriminative features. Wu H et al. [20] introduce a sample-proxy dual triplet (SPDT) loss function with a multi-proxy softmax (MPS) loss function to effectively represent classes with multiple proxies, enhance inter-class distinction, and reduce intra-class variability, thus demonstrating superior performance in vehicle re-identification on large datasets such as VeRi-776 and DukeMTMC-ReID. Yang et al. [21] propose GLFNet, which combines global and local information to enhance vehicle re-identification accuracy. By leveraging both globally invariant features and discriminative local details, their method significantly improves the model's generalization ability. Zheng A et al. [22] propose a unified attribute-guided network that learns global features, camera views, vehicle types, and colors.

The third category involves generating adversarial networks for feature learning. Wang Q et al. [23] abstract various layers of vehicle visual modes, integrate multi-level features, and assign a virtual label to the generated data based on the color domain CVLSR. To reduce the translation noise between images, Xu Z et al. [24] employ a graph neural network (GNN) method to construct a hierarchical spatial structure graph with global and local regions as nodes and a two-layer relationship as edges. Under the constraints of metric learning, the GCN module is utilized to learn discriminative structural features, integrating classification loss with metric learning loss. Zhou and Shao et al. [25,26] learn global multi-view feature representation by proposing a single-view input attention model based on GAN.

Despite the advancements in vehicle re-identification through deep convolutional neural networks, significant gaps persist. Current methods often fail to effectively integrate meta-information with feature fusion, leading to discrepancies between synthetic training data and real scenarios. Furthermore, end-to-end systems, although improved, still struggle with scalability and adapting to diverse environments, requiring extensive tuning to manage dataset variability. Additionally, the use of adversarial networks for feature learning has shown promise in enhancing feature discrimination but suffers from issues of stability and consistency across different domains and visibility conditions. Our research proposes a novel multi-task learning framework that combines an advanced defogging network with a robust re-identification network, aiming to enhance feature integration, scalability, and real-time processing across diverse environments. This approach sets out to redefine benchmarks for vehicle re-identification, especially under challenging conditions such as fog and low light.

3. Methodology

3.1. Multi-Task Learning Vehicle Re-Identification Framework

In this paper, we address vehicle re-identification in foggy scenarios through a novel multi-task learning framework designed to simultaneously perform defogging and re-identification tasks. As illustrated in Figure 1, the workflow begins with a shared feature extraction module where raw input images are processed to extract essential features, which are concurrently directed into two integrated pathways: the defogging and the re-identification modules. Advanced image processing techniques in the defogging module enhance clarity by mitigating fog-related distortions while preserving critical details obscured by fog through a multi-scale composite fusion module that merges the enhanced features with the original data. Simultaneously, the refined features are utilized in the re-identification module where a metric learning component optimizes their use to ensure high precision in vehicle identification. This dual-process approach allows for efficient and accurate identification by addressing the challenges posed by fog directly within the feature extraction and processing stages, with further details on the mechanisms of the phase attention ResNet re-identification branch and the semi-supervised multi-scale defogging branch being discussed in subsequent sections.

3.2. Phase Attention ResNet Re-Identification Branch

ResNet, short for Residual Network, is a type of convolutional neural network that utilizes skip connections or shortcuts to bypass some layers. Typical CNNs become increasingly inefficient and difficult to train as they grow deeper; however, ResNet's architecture allows it to maintain high performance even when scaling up the network depth, making it ideal for tasks requiring deep feature extraction like vehicle re-identification. In foggy scenarios, ResNet's ability to handle deep convolutional features becomes particularly crucial.

A key challenge in image processing is the presence of noise, predominantly found in the shallow features of images. This noise often results in the loss of critical vehicle details during the fog removal stage. To address this issue, our approach emphasizes the importance of analyzing and preserving the fine-grained features across various channels, which is crucial in minimizing mist interference and effectively retain essential vehicle information.

To achieve this, we employ the phase attention (pixel location attention, PLA) mechanism for feature map extraction. This mechanism allows the input image to utilize a shared feature layer within the network, thereby facilitating the extraction of convolutional features. Specifically, the attention mechanism is strategically applied to one layer within the last three layers of the convolutional stack. This application is critical for extracting the feature map's weight information, enabling the segmentation of the image's fine-grained details and enhancing clarity.

In foggy conditions, a unique challenge arises as the fog concentration increases linearly with distance, complicating the extraction process. To overcome this, we propose a novel phase attention method that innovatively designs the vehicle extraction network to

accommodate diverse image channels, integrating both channel weights and spatial local information. This design allows for a more nuanced approach to managing the variability of fog concentration across different image regions.

As illustrated in Figure 2, the phase attention design divides the features extracted by the module into two distinct parts. One part is dedicated to calculating the relevant weights of the positions, addressing the spatial variability of fog. The other part focuses on determining the channel weights, ensuring that the network captures the necessary depth and detail from each channel. By mapping the training channels and image depth onto these two branches, our method achieves a more detailed and clearer image representation. This dual-branch approach is fundamental in extracting and preserving the fine-grained details necessary for accurate vehicle re-identification in challenging foggy conditions.

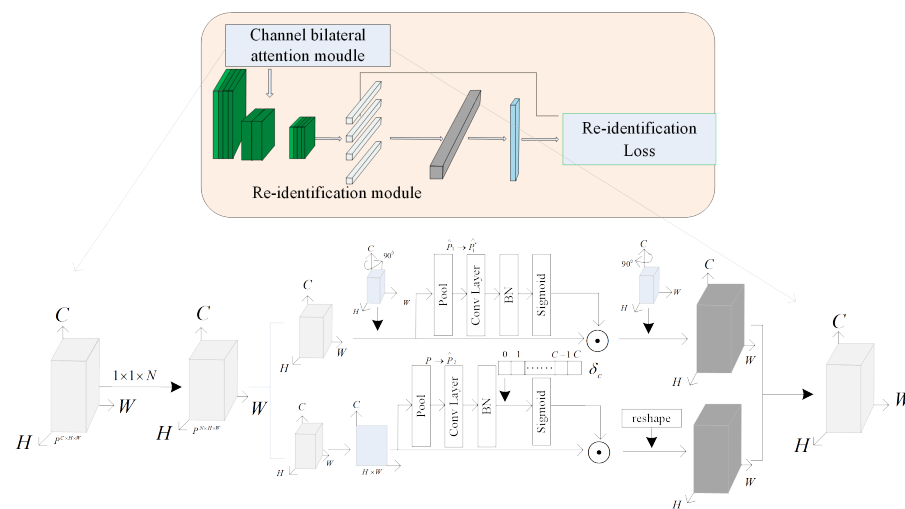


Figure 2. Phase attention structure diagram.

The attention method is used on the re-identification branch feature extraction of the middle layer. After extracting the convolutional feature map $P^{C \times H \times W}$, where C is the number of channels, H is the height, and W is the width, it uses the $1 \times 1 \times N$ dimensional transformation and turns into $P^{N \times H \times W}$ by feature weight transformation through the two-sided attention mechanism.

In order to capture spatial information such as position, the first branch of the bilateral attention mechanism establishes the interaction between position dimensions and channel dimensions. The specific processing process is to rotate the input tensor 90° counterclockwise along the channel dimension to increase the position dimensions between H , W , and C . Next, the tensor is processed through contractualization of pooling, turned into $\hat{P}_1^*$, then through 1×1 standard convolution layer filtering and feature extraction, the batch normalized layer is increased to standardize output and the output tensor process through into the activation layer of $(0, 1)$ interval to generate the attention weight tensor. Finally, this will generate the attention weight tensor applied on the original input tensor along the position dimension H clockwise rotation 90° , making the input tensor keep the original shape of P ; the new tensor introduces higher attention, improving the model performance.

The channel of the feature map P of the second branch transforms H and W by image dimension and is expanded into $(C \times S \times 1)$ features. Starting with $S = H \times W$, under the action of *Avgpool* with the new neurons $\hat{P}_2^*$ with an area of 1, the tensor shape is $(C \times S \times 1)$, through the standard convolution layer defined by the kernel size k , and then through the batch normalization layer. We set a trainable variable $(\delta_c \times 1 \times 1)$ representing the weight mapping relationship between depth and channel $W_d = \delta_c \times C$, where W_d represents the weight of channel C mapping to depth, and the output of the original feature after matrix multiplication, the activation layer *Sigmoid* generates the attention weight of the shape $(C \times 1 \times 1)$, and applies it to the input X .

The refined shape tensor ($C \times H \times W$) generated by each of the branches is then aggregated by simple averaging. For the input tensor $P^{C \times H \times W}$, the process of obtaining the fine attention from the triple attention can be represented as follows:

$$y = \frac{1}{3}(\hat{P}_1\sigma(\psi_1(\hat{P}_1)) + p\sigma(\psi_2(\hat{P}_2))), \quad (1)$$

where σ represents *Sigmoid*, which represents the activation function; ψ_1, ψ_2 represent the standard 2D convolution layer of convolution kernel size k in the two branches of phase attention. The formula can be simplified to

$$y = \frac{1}{3}(\hat{P}_1\alpha_1 + \hat{P}_2\alpha_2) = \frac{1}{3}(y_1 + y_2), \quad (2)$$

where α_1 and α_2 are the three cross-dimensional attention weights computed in phase attention. Combining the weight of phase attention with the original image can obtain the weight information of the vehicle at the channel level and space and above. The channel level information can reduce the influence of noise and the spatial level information can depict the fine-grained characteristics of the picture in some areas, providing sufficient information for subsequent comparison. The heat map generated by ResNet is compared with the heat map after adding PLA. As shown in Figure 3, ResNet can extract the whole vehicle information, and the key position information is more obvious after adding PLA features, helping the model to retain more useful image details. The heat maps presented in the figure compare the attention distribution of the ResNet model alone and enhanced with the PLA mechanism. The ResNet subfigure shows the heat map generated by the basic ResNet model, indicating where the model focuses within the vehicle image, typically on distinguishing features such as the car's outline and windows. The ResNet + PLA subfigure illustrates how the addition of PLA shifts and intensifies the focus toward more specific features like the wheels and the rear parts of the vehicle, demonstrating the PLA's ability to direct the model's attention to areas that are crucial for more accurate vehicle identification. This comparative visualization highlights the effectiveness of integrating attention mechanisms to enhance feature extraction capabilities of convolutional networks in vehicle re-identification tasks.



Figure 3. PLA and ResNet.

3.3. Multi-Scale Fusion Defogging Branch

When utilizing semi-supervised methods, there is a tendency for the learned information to be biased toward synthetic data, given the larger scale of synthetic data compared with real data. Therefore, it is crucial to refine the fog removal process by analyzing the

discrepancies between corresponding synthetic and original images. This paper introduces a multi-scale fusion module (MsF) to address this issue. As illustrated in Figure 4, the initial defogging is executed by the dimension reconstruction module. As depicted in Figure 5, this process is further enhanced by the multi-scale fusion module. The following three methods are employed to discern the differences between initial and dense fog, thereby restoring the detailed information of the original image: (1) extracting the multidimensional sign map as a feature pyramid; (2) integrating the relatively clean feature map with the original image to preserve its characteristics; and (3) utilizing the feature sensing loss function and metrics loss function for training.

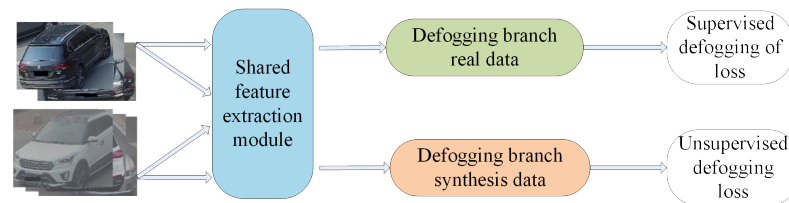


Figure 4. A semi-supervised learning framework diagram.

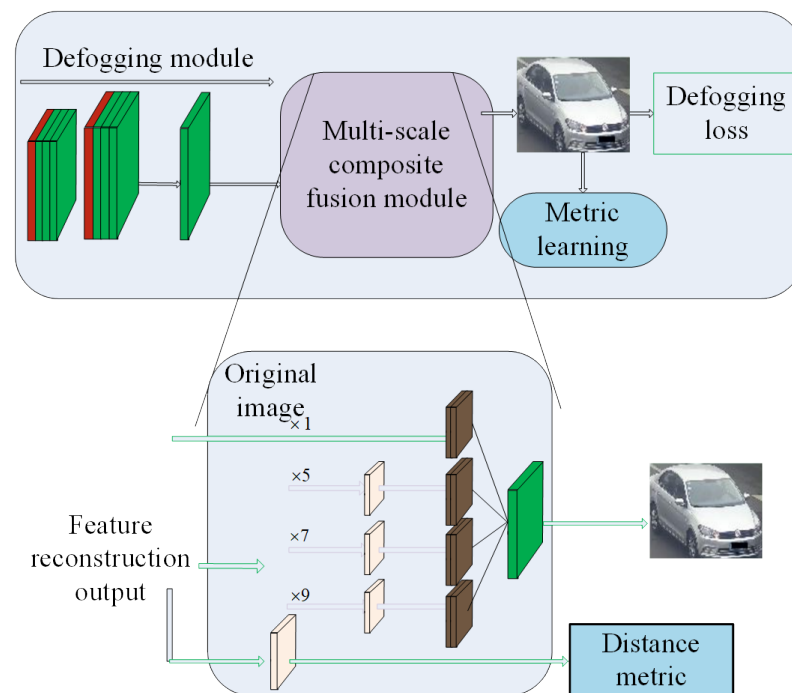


Figure 5. Multi-scale fusion network diagram.

Re-identification requires cleaning and extracting fine-grained features. In this paper, the ResNet50 network is used to extract feature maps, and instance regularized IBN structure is used to map features to different spaces for feature extraction. The choice of ResNet50 is particularly advantageous for its deep residual learning framework, which helps in alleviating the vanishing gradient problem that often occurs with deep networks. This capability allows ResNet50 to learn rich and complex feature representations effectively. Additionally, the architecture of ResNet50, with its skip connections, ensures that it can learn optimized features without degradation of network performance as depth increases, which is crucial for maintaining high accuracy in the detailed feature analysis required for re-identification tasks. First, the image rough features are extracted from the global feature sharing module to share the feature maps of the re-identification branch and the defogging branch. Then, the fog information reconstruction module is entered, from which the dimension reconstruction module can extract shared features and cause accuracy loss

due to the dimension reduction of the module, samples on its feature map, expands the feeling field, obtains the rough feature map information f_p , and then enters the multi-scale fusion module after two 3×3 convolution layers. Next, the original image feature map and the rough defogging results are entered. Among them, the four branches of a multi-scale pyramid, the output of the front convolution layer as the first three branch input through the average pooling layer, respectively, with 9×9 , 7×7 , and 5×5 convolution kernel downsampling—the last branch for the original share feature module extraction, to build a four scale pyramid, and then each scale layer 1×1 convolution dimension reduction. Features are then upsampled and linked together. Finally, the cascade features are convolved 3×3 , and the fog-free feature graph f_e is obtained by fusion. Meanwhile, the lowest input can be identified with f_e through the final metrics to accelerate the training of the dimension reconstruction module, so as to improve the accuracy.

4. Experiments

4.1. Advanced Verification of PLA Method

The current evaluation metrics for vehicle re-identification algorithms primarily consist of top-k accuracy and mean average precision (mAP). Given the challenges associated with collecting real-time datasets under variable and unpredictable foggy conditions, coupled with the scarcity of readily available fog-specific vehicle re-identification data, this paper introduces a new dataset, VERI-Heavy, specifically designed for vehicle re-identification in foggy scenarios. VERI-Heavy, derived from the official VERI-Wild [27] and Vehicle-1M datasets [28], includes a selection of real images that encompass both foggy and normal weather conditions. This dataset comprises 42,558 vehicle samples, of which 3000 are specifically from foggy conditions. This extensive collection provides a robust annotation foundation and ensures the dataset's relevance and applicability, facilitating controlled testing and validation of re-identification models across diverse environmental scenarios.

To assess the effectiveness of the proposed pixel location attention (PLA) method, we conducted a comparative analysis on the VERI-Heavy dataset using several advanced methodologies, as detailed in Table 1. The comparison involved using the query set as the retrieval image and the test set as the gallery. The algorithms compared include VOC (visual object classes), VEHICLEX, DMT (deep matching transduction), PVEN (pose viewpoint energy network), TransReID (transforming and retrieving identity), and SJDL (semi-supervised joint defogging learning). The experimental results demonstrate the superior performance of our method. Among these metrics, the Top-1, Top-5, and Top-10 accuracy rates indicate the precision of identifying the Top-1, 5, and 10 vehicles with the highest similarity, respectively.

Table 1. Comparison results of PLA and advanced re-identification algorithm.

Algorithm	Top-1%		Top-5%		Top-10%		mAP%	
	S	R	S	R	S	R	S	R
VOC [29]	86.1	82.8	94.3	94.0	95.6	96.6	59.7	57.4
VEHICLEX [30]	86.5	83.2	95.0	95.2	97.4	97.9	63.6	61.5
DMT [11]	93.4	93.2	97.2	97.4	97.9	98.5	73.9	71.7
PVEN [31]	63.7	66.4	84.3	86.5	89.6	91.2	72.8	75.3
TransReID [19]	82.4	77.7	92.3	88.8	98.4	94.0	62.9	64.0
SJDL [7]	94.6	94.6	97.9	98.1	98.9	99.2	85.3	82.7
Ours	96.3	96.3	98.3	98.9	98.7	99.5	87.8	84.1

Note: Values in bold represent the highest scores in each column.

The results in Table 1 indicate that PLA outperforms other methods for both synthetic (S) and real (R) data sets. In the synthetic data set, PLA showed improvements in metrics like mAP, Top-1, and Top-5 by at least 2.45%, 1.72%, and 0.41%, respectively; in the real data set, improvements in mAP, Top-1, Top-5, and Top-10 were at least 1.42%, 1.70%, 0.80%, and 0.30%, respectively. The effectiveness of this method stems from the phase

attention approach's ability to extract relevant information at the channel and position levels, and the metrics loss function's enhanced accuracy in distinguishing between anchor, positive, and negative samples in both foggy and non-foggy images. In the synthetic data, the Top-10 performance of PLA is slightly lower than that of the SJDL model, yet it remains at a relatively high level with only a 0.2% gap, which falls within an acceptable deviation range. The corresponding CMC curves are depicted in Figure 6.

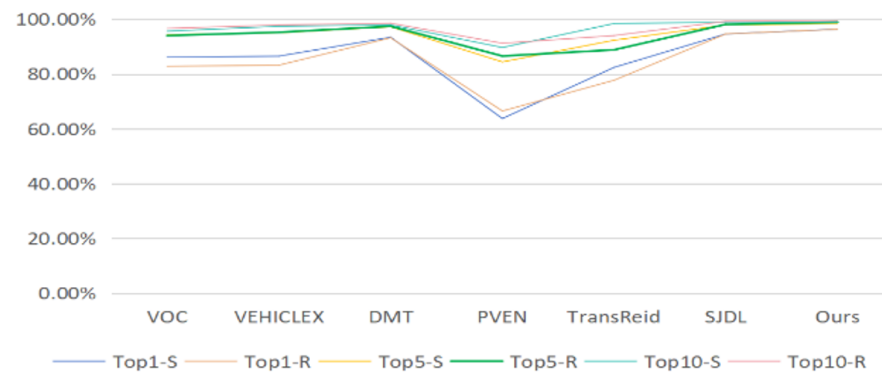


Figure 6. Re-identification algorithm CMC, line diagram.

This paper presents the experimental results of the PLA method through visual representation. The query image is used to locate the same sample vehicle in the gallery, with green indicating the correct image and red denoting the incorrect image. Serial numbers 1–5 in the figure represent the top five most similar images in the gallery.

In Figure 7, three vehicles with IDs 238, 503, and 411 are shown, and the re-identification results of the PLA model under the Top-5 evaluation, show 100% accuracy for IDs 503 and 411, with only one identification error for ID 238. The possible reasons for the identification error include the following: (1) the vehicle's appearance being very similar to the queried vehicle; (2) the vehicle in question being partially obscured, leading to confusion with another vehicle of similar appearance in the gallery.



Figure 7. Top-5 visualization result.

4.2. Advanced Verification of PLA Method

The current evaluation metrics for defogging algorithms include $PSNR_{fake}$, $SSIM_{fake}$, $PSNR_{ehn}$, and $SSIM_{ehn}$. To validate the effectiveness of the MsF module introduced in this section, the results for each evaluation metric are as follows: After integrating the MsF module into the SJDL method, there were increases of 0.529%, 0.003%, 0.025%, and 0.007% in $PSNR_{fake}$, $SSIM_{fake}$, $PSNR_{ehn}$, and $SSIM_{ehn}$, respectively. This improvement is attributed to

the acquisition of multi-scale image detail information, particularly optimizing the original image details for enhanced fog effect mitigation. These specific results are presented in Table 2.

Table 2. Msf is compared with the other defogging methods.

	PSNR _{fake}	SSIM _{fake}	PSNR _{ehn}	SSIM _{ehn}
YOLO [32]	21.840	0.819	22.384	0.901
TMSI [33]	28.265	0.861	27.972	0.951
PMHLD [34]	29.744	0.933	28.125	0.933
RefineDNet [35]	27.132	0.931	28.014	0.891
D4 [35]	27.732	0.932	27.982	0.956
SJDL [7]	28.125	0.933	29.744	0.933
Ours	28.769	0.940	29.854	0.936

Note: Values in bold represent the highest scores in each column.

This paper presents a two-stage fog removal method, where the first stage focuses on feature reconstruction with minimal constraints, thereby preserving more image details. Furthermore, convolution at various scales is employed to capture details at different levels, and metric learning is utilized during the supervised learning phase to optimize information gain. The feature sharing module benefits from the information feedback from the re-identification branch to the defogging task. Consequently, this fusion of image features at various scales significantly enhances the quality of fog removal.

4.3. Fusion Experiment of Multi-Scale Defogging and Phase Attention Re-Identification

To assess the compatibility of the semi-supervised multi-scale composite-aware defogging method, MsF, with traditional re-identification methods, this paper replaces the defogging branch in SJDL with MsF and evaluates its impact on re-identification tasks. This replacement aims to determine how the integration of an advanced defogging technique affects the performance of established re-identification models, which is crucial for applications in diverse environmental conditions. To further assess the compatibility of the semi-supervised multi-scale fusion sensing defog method, MsF, with the phase attention-based re-identification method, this study employs MsF as the defogging branch and adds PLA to the re-identification branch to evaluate the task. The addition of PLA is intended to explore the synergistic effects of advanced feature extraction methods in enhancing the overall accuracy of vehicle re-identification. The specific experimental results are detailed in Table 3.

Table 3. MsF and PLA fusion results.

Method	Top-1%	Top-5%	Top-10%	mAP%
SJDL	94.61	97.90	98.90	85.36
SJDL+MsF	94.65	97.91	98.90	85.38
SJDL+PLA	96.30	98.30	98.71	87.80
SJDL+MsF+PLA	96.31	98.27	98.72	87.80

Note: Values in bold represent the highest scores in each column.

The experimental results demonstrate that the proposed MsF defogging method, in combination with the high compatibility of PLA, leads to enhanced performance across all experimental groups. However, the combination of SJDL + MsF + PLA, when compared with SJDL + PLA, shows a slight decrease in performance by 0.03% in re-identification results. Consequently, this experiment validates that a model can simultaneously handle both defogging and re-identification tasks, reducing computing power costs and enhancing network accuracy, thereby confirming the method's superiority.

5. Conclusions

This paper introduces a multi-task re-identification (re-ID) framework that effectively performs dehazing and re-ID tasks in two different branches. Leveraging advanced deep learning architectures, especially phase attention networks, the framework significantly improves the re-ID accuracy. Experimental results show that the mean average precision (mAP) can be improved up to 87.8% in foggy conditions, which illustrates the effectiveness of our multi-scale composite fusion network in merging key information in the original image with the feature pyramid network to minimize information loss during processing.

Despite these technical advances, the framework still faces several implementation challenges, especially when scaling to larger datasets or adapting to dynamic environmental conditions such as varying fog density and lighting levels. Adapting to a variety of meteorological conditions, including dust, rain, and snow also poses significant challenges. Future research will focus on improving these systems to more skillfully manage the complexity of real-world environments, thereby improving their real-time operation capabilities and accuracy in adverse weather conditions. At the same time, the technology is expected to reduce traffic accidents and enhance traffic flow management, especially in urban areas where low visibility often leads to road accidents. Furthermore, the ability to accurately identify vehicles under a variety of conditions could enhance safety measures and support smarter urban planning with significant impacts on societal well-being.

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

Funding: This research was funded by the Beijing Municipal Society of Higher Education grant number MS2022314, the National Research Project on Higher Education in the Coal Industry grant number 2021MXJG44, the Ministry of Education’s Industry-University Cooperation Collaborative Educational Programme grant number 202102210008, the Open Fund of Anhui Engineering Research Center for Intelligent Applications and Security of Industrial Internet grant number IASII24-09, and the Fundamental Research Funds for the Central Universities grant number 2024ZKPYZN01.

Data Availability Statement: The data presented in this study are not publicly available due to privacy or ethical restrictions.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Abbreviations

The following abbreviations are used in this manuscript:

DCNN	Deep convolutional neural network
JPM	Jigsaw patch module
SPDT	Sample-proxy dual triplet
MPS	Multi-proxy softmax
GNN	Graph neural network
MsF	Multi-scale fusion
PLA	Pixel location attention
CMC	Cumulative match characteristic
PSNR	Peak signal-to-noise ratio
SSIM	Structural similarity index
IBN	Instance batch normalization
VERI	Vehicle re-identification

References

1. Alsrehin, N.O.; Klaib, A.F.; Magableh, A.I. Intelligent transportation and control systems using data mining and machine learning techniques: A comprehensive study. *IEEE Access* **2019**, *7*, 49830–49857. [\[CrossRef\]](#)
2. Tang, Z.; Naphade, M.; Liu, M.-Y.; Yang, X.; Birchfield, S.; Wang, S.; Kumar, R.; Anastasiu, D.; Hwang, J.-N. CityFlow: A City-Scale Benchmark for Multi-Target Multi-Camera Vehicle Tracking and Re-Identification. *arXiv* **2019**, arXiv:1903.09254.
3. Chen, X.; Liu, L.; Deng, Y.; Kong, X. Vehicle detection based on visual attention mechanism and adaboost cascade classifier in intelligent transportation systems. *Opt. Quantum Electron.* **2019**, *51*, 263. [\[CrossRef\]](#)
4. Khorramshahi, P.; Shenoy, V.; Chellappa, R. Robust and Scalable Vehicle Re-Identification via Self-Supervision. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Vancouver, BC, Canada, 17–24 June 2023; pp. 5295–5304.
5. Zhang, Z.; Lin, H.; Liu, Y.; Li, Z.; He, Y. Vehicle Re-Identification in Foggy Weather Using Domain Adaptation and Generative Adversarial Networks. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, Paris, France, 2–6 October 2023; pp. 2348–2357.
6. Li, H.; Chen, J.; Zheng, A.; Wu, Y.; Luo, Y. Day-Night Cross-domain Vehicle Re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 17–21 June 2024; pp. 12626–12635.
7. Chen, W.T.; Chen, I.H.; Yeh, C.Y.; Yang, H.H.; Ding, J.J.; Kuo, S.Y. SjdI-vehicle: Semi-supervised joint defogging learning for foggy vehicle re-identification. In Proceedings of the AAAI Conference on Artificial Intelligence, Vancouver, BC, Canada, 28 February–1 March 2022; Volume 36, pp. 347–355.
8. Wu, F.; Chen, H.; Xie, M. Adaptive Multi-Scale Feature Fusion for Vehicle Re-Identification. In Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV) Workshops, Paris, France, 2–6 October 2023; pp. 1289–1298.
9. Kwasniewska, A.; MacAllister, A.; Nicolas, R.; Garza, J. Multi-Sensor Ensemble-Guided Attention Network for Aerial Vehicle Perception Beyond Visible Spectrum. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Vancouver, BC, Canada, 17–24 June 2023; pp. 345–353.
10. Wu, W.; Ke, W.; Sheng, H. A Transformer-based Unsupervised Clustering Method for Vehicle Re-identification. In Proceedings of the 2022 6th International Conference on Universal Village (UV), Boston, MA, USA, 22–25 October 2022; pp. 1–5. [\[CrossRef\]](#)
11. He, S.; Luo, H.; Chen, W.; Zhang, M.; Zhang, Y.; Wang, F.; Li, H.; Jiang, W. Multi-domain learning and identity mining for vehicle re-identification. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Identification Workshops, Seattle, WA, USA, 13–19 June 2020; pp. 582–583.
12. Lee, Y.; Tang, Q.; Choi, J.; Jo, K. Low Computational Vehicle Re-Identification for Unlabeled Drone Flight Images. In Proceedings of the IECON 2022—48th Annual Conference of the IEEE Industrial Electronics Society, Brussels, Belgium, 17–20 October 2022; pp. 1–6. [\[CrossRef\]](#)
13. ElRashidy, A.; Ghoneima, M.; Munim, H.E.A.E.; Hammad, S. Recent Advances in Vision-based Vehicle Re-identification Datasets and Methods. In Proceedings of the 2021 16th International Conference on Computer Engineering and Systems (ICCES), Cairo, Egypt, 15–16 December 2021; pp. 1–6. [\[CrossRef\]](#)
14. Ye, M.; Chen, S.; Li, C.; Zheng, W.S.; Crandall, D.; Du, B. Transformer for Object Re-Identification: A Survey. *arXiv* **2024**, arXiv:2401.06960.
15. Wang, Y. Deep learning technology for re-identification of people and vehicles. In Proceedings of the 2022 IEEE International Conference on Electrical Engineering, Big Data and Algorithms (EEBDA), Changchun, China, 24–26 February 2023; pp. 972–975. [\[CrossRef\]](#)
16. Roman-Jimenez, G.; Guyot, P.; Malon, T.; Chambon, S.; Charvillat, V.; Crouzil, A.; Péninou, A.; Pinquier, J.; Sèdes, F.; Senac, C. Improving vehicle re-identification using CNN latent spaces: Metrics comparison and track-to-track extension. *IET Comput. Vis.* **2021**, *15*, 85–98. [\[CrossRef\]](#)
17. Zheng, A.; Zhang, C.; Li, C.; Tang, J.; Tan, C. Multi-Query Vehicle Re-Identification: Viewpoint-Conditioned Network, Unified Dataset and New Metric. *IEEE Trans. Image Process.* **2023**, *32*, 5948–5960. [\[CrossRef\]](#) [\[PubMed\]](#)
18. 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, 16–17 June 2019; pp. 3997–4005.
19. 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, Virtual, 11–17 October 2021; pp. 15013–15022.
20. Wu, H.; Shen, F.; Zhu, J.; Zeng, H.; Zhu, X.; Lei, Z. A sample-proxy dual triplet loss function for object re-identification. *IET Image Process.* **2022**, *16*, 3781–3789. [\[CrossRef\]](#)
21. Yang, Y.; Liu, P.; Huang, J.; Song, H. GLFNet: Combining global and local information in vehicle re-recognition. *Sensors* **2024**, *24*, 616. [\[CrossRef\]](#) [\[PubMed\]](#)
22. Li, H.; Lin, X.; Zheng, A.; Li, C.; Luo, B.; He, R.; Hussain, A. Attributes guided feature learning for vehicle re-identification. *IEEE Trans. Emerg. Top. Comput. Intell.* **2021**, *6*, 1211–1221. [\[CrossRef\]](#)
23. Wang, Q.; Min, W.; Han, Q.; Yang, Z.; Xiong, X.; Zhu, M.; Zhao, H. Viewpoint adaptation learning with cross-view distance metric for robust vehicle re-identification. *Inf. Sci.* **2021**, *564*, 71–84. [\[CrossRef\]](#)

24. Xu, Z.; Wei, L.; Lang, C.; Feng, S.; Wang, T.; Bors, A.G. HSS-GCN: A Hierarchical Spatial Structural Graph Convolutional Network for Vehicle Re-identification. In Proceedings of the ICPR 2021: Pattern Identification. ICPR International Workshops and Challenges, International Workshop on Human and Vehicle Analysis for Intelligent Urban Computing (IUC), Virtual Event, 10–15 January 2021.
25. Yan, H.; Huang, K.; Wang, X.; Li, Y. A Vehicle Re-Identification Method Based on Fine-Grained Features and Metric Learning. In Proceedings of the 2022 6th Asian Conference on Artificial Intelligence Technology (ACAIT), Changzhou, China, 9–11 December 2022; pp. 1–6. [\[CrossRef\]](#)
26. Du, L.; Huang, K.; Yan, H. ViT-ReID: A Vehicle Re-identification Method Using Visual Transformer. In Proceedings of the 2023 3rd International Conference on Neural Networks, Information and Communication Engineering (NNICE), Guangzhou, China, 24–26 February 2023; pp. 287–290. [\[CrossRef\]](#)
27. Lou, Y.; Bai, Y.; Liu, J.; Wang, S.; Duan, L. VERI-Wild: A Large Dataset and a New Method for Vehicle Re-Identification in the Wild. In Proceedings of the in 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, Long Beach, CA, USA, 16–17 June 2019; pp. 5295–5304.
28. Guo, H.; Zhao, C.; Liu, Z.; Wang, J.; Lu, H. Learning coarse-to-fine structured feature embedding for vehicle re-identification. In Proceedings of the AAAI Conference on Artificial Intelligence, New Orleans, LA, USA, 2–7 February 2018.
29. Zhu, X.; Luo, Z.; Fu, P.; Ji, X. VOC-ReID: Vehicle re-identification based on vehicle-orientation-camera. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 13–19 June 2020; pp. 602–603.
30. Yao, Y.; Zheng, L.; Yang, X.; Naphade, M.; Gedeon, T. Simulating content consistent vehicle datasets with attribute descent. In Proceedings of the Computer Vision—ECCV 2020: 16th European Conference, Glasgow, UK, 23–28 August 2020; Proceedings, Part VI; Springer International Publishing: Berlin/Heidelberg, Germany, 2020; pp. 775–791.
31. 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 (CVPR), Seattle, WA, USA, 13–19 June 2020; IEEE: Piscataway, NJ, USA, 2020.
32. Redmon, J.; Divvala, S.; Girshick, R.; Farhadi, A. You Only Look Once: Unified, Real-Time Object Detection. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 779–788.
33. Chen, P.; Liu, S.; Kolmanič, S. Research on Vehicle Re-Identification Algorithm Based on Fusion Attention Method. *Appl. Sci.* **2023**, *13*, 4107. [\[CrossRef\]](#)
34. Liang, J.; Gong, S.; Cheng, X.; Wang, L.; Zheng, F. Progressively Multi-Head Lightweight Network for Vehicle Re-Identification. *IEEE Trans. Image Process.* **2020**, *29*, 1489–1502.
35. Zhang, H.; Patel, V.M. Densely Connected Pyramid Dehazing Network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3194–3203.

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.