

## Joint Multiple Fine-grained feature for Vehicle Re-Identification

Yan Xu <sup>\*</sup>, Leilei Rong <sup>\*\*</sup>, Xiaolei Zhou, Xuguang Pan, Xianglan Liu

*College of Electronic and Information Engineering, Shandong University of Science & Technology, Qingdao 266590, China*

### ARTICLE INFO

**Keywords:**

Image retrieval  
Deep learning  
Vehicle re-identification  
Fine-grained feature  
Feature map segmentation  
mINP

### ABSTRACT

The process of recognizing the same vehicle in different scenes is called vehicle re-identification. However, due to the different locations of the surveillance cameras, there may be obstacles in the captured vehicle pictures and multiple viewpoints may make the same vehicle look different. In order to effectively reduce the interference of obstacle occlusion, multiple viewpoints, and other factors on vehicle re-identification, in this paper, we propose a multi-fine-grained feature extraction network. While retaining the global information of vehicles, we extract the finegrained features of vehicles precisely by segmenting the vehicle feature map. In addition, we introduce a new evaluation metric mean Inverse Negative Penalty (mINP) to evaluate the vehicle re-identification model more comprehensively. Our method achieves superior accuracy over the state-of-the-art methods on the challenging vehicle datasets: VeRi-776, VehicleID, and VRIC.

### 1. Introduction

Vehicle re-identification (re-id) aims to identify the same vehicle from different scenes, and belongs to the subtask of image retrieval. With the promotion of deep learning technology, vehicle re-identification has become a hot topic in the field of computer vision. However, due to the production of vehicle database from different surveillance cameras in the real world, the captured vehicle images have interference factors such as motion blur, dark background, low resolution, obstacle occlusion, multi-viewpoint, and so on. Hence, how to weaken the adverse factors and improve the accuracy of vehicle re-identification has become a key research direction in this field.

In recent years, researchers have mainly designed new network structures [1–7] based on convolutional neural networks (CNN) to learn more discriminative vehicle features or introduced additional information [8–10] to improve the performance of vehicle re-identification models. Early research [11] focused on using the global information of the vehicle to complete the re-identification task, while ignoring the local information of the vehicle. The local information contains key features that distinguish different vehicles. In order to obtain local feature information of vehicles, Liu et al. [12] proposed a region-aware deep Model (RAM), which can not only extract global features of vehicles but also learn discriminant features of different local regions. He et al. [13] proposed a simple and effective local regularization method, which improves the network's ability to perceive subtle feature

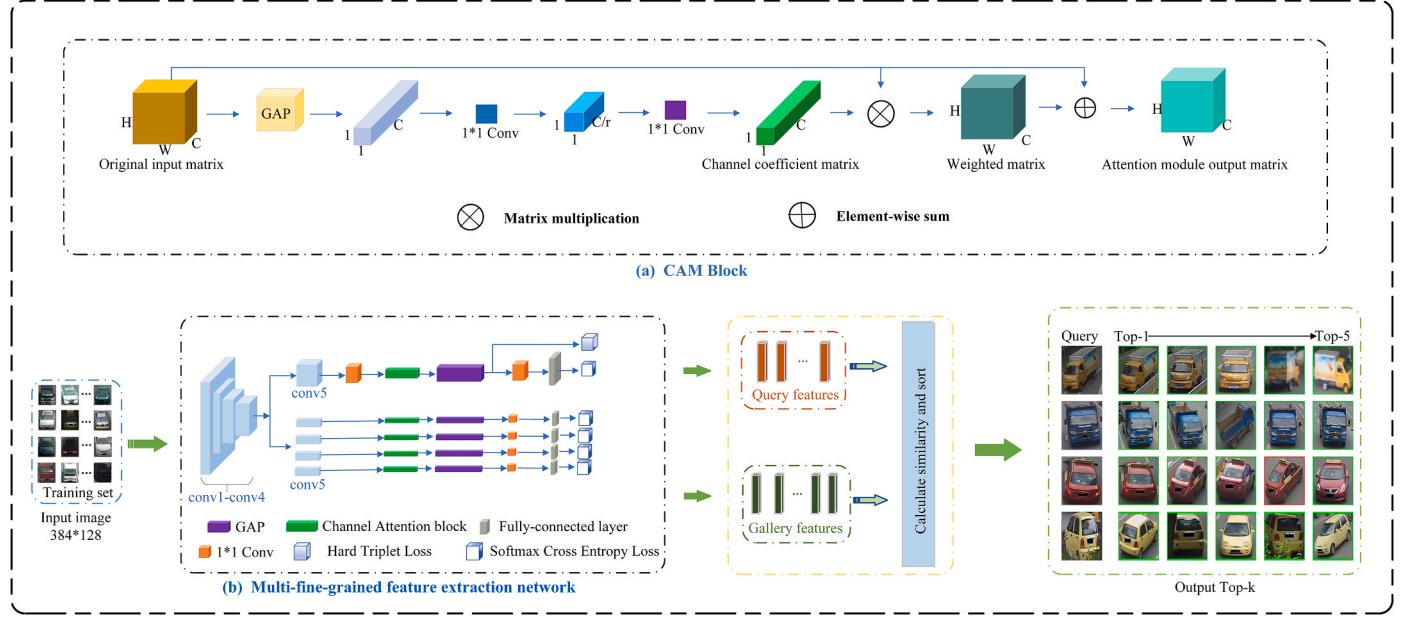
differences, strengthens the network's learning of local features, and further expands the differences between similar instances. Khorramshahi et al. [14] adopted the idea of feature extraction from rough to fine. In the first stage, the proposed adaptive key point selection module was used to select the key points with the largest amount of information from the initial layer of the global feature extraction network to roughly extract local features around the selected key points. In the second stage, the vehicle features of the first stage are refined through the two-layer hourglass network with jump connections, and the relatively fine vehicle features obtained are used for vehicle feature matching. Therefore, using local feature information for vehicle re-identification can make the network model extract more refined vehicle features, and combine with the global features of vehicles, can significantly improve the model's ability to distinguish vehicles with similar appearance but different identities, and improve the accuracy of vehicle re-identification. Later, scholars [15] combined global and local information together. Although it effectively improved the accuracy of vehicle re-identification, they failed to make full use of the vehicle's local information.

To solve this problem, we divide the vehicle feature map horizontally and keep the global information of the vehicle at the same time. The advantage of convolutional neural networks lies in the perception of local information in the image. By feature map segmentation, on the one hand, it can make the feature extraction network pay more attention to the fine-grained features of the vehicle (such as vehicle logo, lights,

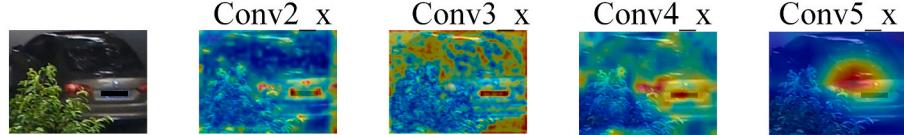
\* Corresponding author.

\*\* Corresponding author.

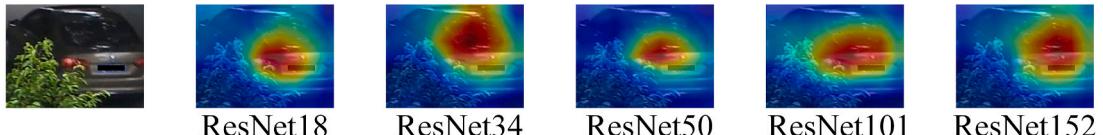
E-mail addresses: [x1y5@163.com](mailto:x1y5@163.com) (Y. Xu), [rong15305385721@163.com](mailto:rong15305385721@163.com) (L. Rong).



**Fig. 1.** The overall framework of the algorithm model for vehicle re-identification. (a) CAM represents a channel attention mechanism; (b) Illustration of multi-fine-grained feature extraction network. (Best view in color).



(a) Visualization of the feature maps of ResNet50 with different level information



(b) Visualization of the feature maps of different ResNet.

**Fig. 2.** Visualization of the feature maps of deep residual networks.

annual inspection signs, body decoration, etc.). On the other hand, it can effectively reduce the influence of obstacles and other adverse factors in the original image on vehicle feature learning. In addition, we also introduce channel attention mechanism to further strengthen the network's recognition of vehicle fine-grained features after feature map segmentation. Experiments on the mainstream open datasets (VeRi-776 [16], VehicleID [17], and VRIC [18]) show that the best result is achieved in the feature map quartering, as shown in Fig. 1(b).

The rest of this paper is organized as follows: Section 2 introduces the specific process of vehicle re-identification and provides an overview of multi-fine-grained feature extraction network and loss function. Section 3 introduces the new evaluation metric: mINP. Section 4 presents the experimental results and analysis, and Section 5 draws concluding remarks.

## 2. Proposed method

### 2.1. The specific process of vehicle re-id

Firstly, the image of the training set is imported into the multi-fine-grained feature network. After several iterations, the optimal model is

obtained. Then the model is put into the testing set (Query set and Gallery set) to calculate the Euclidean distance between the target vehicle in Query set and the vehicle to be retrieved in Gallery set, and the similarity between the vehicles is calculated and ranked by the distance. The higher the ranking, the higher the similarity. Finally, the retrieval results are printed by line. The green border shows a correct retrieval and the red border shows an incorrect retrieval.

### 2.2. Multi-fine-grained feature extraction network

In order to fully extract the fine-grained features of the vehicle and improve the accuracy of vehicle re-identification, we propose a multi-fine-grained feature extraction network (MFG-Net), as shown in Fig. 1 (b). It can be seen from Fig. 2(b) that compared with other deep residual networks [19], ResNet50 network is more targeted in extracting vehicle features. At the same time, it can be seen from Fig. 2(a) that Conv\_5 layer of the network focuses on the features of the vehicle area, so we take ResNet50 as our backbone network, and two feature extraction branches are built after the Conv\_5 layer: global branch and local branch.

The 1\*1 convolution module is added before and after the global average pooling (GAP) in the global branch. The purpose of adding 1\*1

**Table 1**

The parameter and meaning of loss function.

Parameter	Meaning
$N_i$	the number of vehicle images per batch
$N_{id}$	the number of vehicle identities
$x_j$	the output of fully connected layer for $j$ th identity
$y$	the ground truth identity of input vehicle image
$A_i$	Anchor
$P_i$	Positive
$N_j$	Negative
$\delta$	minimal margin
$\alpha$	Weight
$\beta$	Weight

convolution module before GAP is to greatly increase the non-linear characteristics while keeping the scale of the feature map unchanged, so that the network can give full play to the advantages of depth. When  $1 \times 1$  convolution module is added after GAP, the features after GAP will not go through the classification layer directly, but are fused first, and the fused features are classified. In this way, the accuracy can be greatly improved without affecting the model inference speed.

In the local branch, the vehicle feature map ( $h \times w \times c$ ) is divided into four blocks according to its height, and the size of each block is  $(h/4 \times w \times c)$ , which can effectively reduce the influence of such adverse factors as occlusion on feature extraction. At the same time, we embed the channel attention module in the global branch and the local branch respectively to make the network pay more attention to the personalized characteristics of the vehicle. The working principle of this module is shown in Fig. 1(a).

During the training stage, global branch and local branch do not share the weight and train separately. But when testing, all branch information will be assembled into a comprehensive feature to increase network performance.

### 2.3. Loss function

In the global branch of the MFG-Net network, we introduce two loss functions: hard mining triplet loss [20] and softmax cross-entropy loss; meanwhile, in the local branch, we only use softmax cross-entropy loss, and finally get a total loss by weighting.

$$L_{\text{Softmax}} = - \sum_{i=1}^{N_i} \log \left( \frac{\exp(x_i)}{\sum_{j=1}^{N_{id}} \exp(x_j)} \right) \quad (1)$$

$$L_{\text{hard mining triplet}} = \sum_{i=1}^Q \sum_{A=1}^K \left[ \underbrace{\max_{P=1, \dots, K} \|A_i - P_i\|_2}_{\text{hardest positive}} - \underbrace{\min_{\substack{N=1, \dots, K \\ j=1, \dots, Q \\ i \neq j}} \|A_i - N_j\|_2 + \delta}_{\text{hardest negative}} \right]_+ \quad (2)$$

$$L_{\text{total}} = \alpha^* L_{\text{Softmax}} + \beta^* L_{\text{hard mining triplet}} \quad (3)$$

where the meanings of the parameters of (1), (2) and (3) are listed in Table 1.

### 3. mINP: a new evaluation metric for vehicle Re-ID

Since the previous vehicle re-id model evaluation metrics mAP and Rank-n cannot objectively evaluate the retrieval ability of the model, we introduce a new model evaluation metric: mINP (mean Inverse Negative Penalty) [21], which is used to characterize the most difficult and correct retrieval ability of the network model. So far, we are the only team that uses the ability of the sample retrieval as an indicator in the field of vehicle re-identification.

$$NP_i = \frac{X_i - G_i}{X_i} \quad (i = 1, 2, \dots, Q) \quad (4)$$

$$INP_i = 1 - NP_i = \frac{G_i}{X_i} \quad (5)$$

$$mINP = \frac{1}{Q} \sum_{i=1}^Q 1 - NP_i = \frac{1}{Q} \sum_{i=1}^Q \frac{X_i - G_i}{X_i} \quad (6)$$

In vehicle re-identification,  $G$  and  $X$  are the number of target vehicles in vehicle retrieval results and the number of times to retrieve the last target vehicle, respectively;  $X-G$  is the number of interfering vehicles;  $NP$  and  $INP$  are the occupations of interfering vehicles and target vehicles in retrieval results in turn. mINP represents the mean occupancy rate of  $Q$  target vehicles in the search results.

Therefore, mINP can evaluate the performance of the model more objectively and effectively avoid the dominance of simple matching in mAP/CMC evaluation. It can not only reflect the relative performance of the vehicle re-id model but also provide a supplement to the widely used mAP and CMC metrics.

## 4. Experiment

### 4.1. Datasets and settings

We conduct extensive experiments on three public benchmarks for vehicle re-id, namely, VeRi-776, VehicleID, and VRIC. The details of the datasets are shown in Table 2.

In addition, the software tools are PyTorch, CUDA11.1, and CUDNN V8.0.4.30. The hardware device is a workstation equipped with AMD Ryzen 5 3600X CPU 32G, NVIDIA GeForce RTX 3080 and 256 GB+2 TB memory.

### 4.2. Evaluation protocols and implementation details

During the training stage, the vehicle image is resized to  $384 \times 128$ , and then enhanced by random erasure and horizontal flipping. At the same time, the Amsgrad optimizer is used to optimize our model. The

initial learning rate is 0.0003.  $\alpha$ ,  $\beta$ , and  $\delta$  are set to 1, 0.1, and 0.3.

During the testing stage, the protocol proposed in Refs. [8,16] is followed. We compute the Cumulative Matching Characteristic (CMC) curves for three datasets, and further compute the mean Average Precision (mAP) and Rank1. Moreover, we also calculate a new evaluation metric mINP on the three datasets. In this way, we can evaluate our model more objectively.

**Table 2**  
The details of the datasets.

Dataset	VeRi-776	VRIC	VehicleID	
Images	51,035	60,430	221,763	
IDs	776	5,622	26,267	
Training	37,778/	54,808/	110,178/13,134	
Set/IDs	576	2,811		
Query/IDs	1,678/	2,811/	6,532/	11,395/
	200	2,811	800	1,600
Gallery/IDs	11,579/	2,811/	800/800	1,600/
	200	2,811	1,600	2,400

**Table 3**  
The result of the horizontal division times of feature map on VeRi-776 dataset.

Horizontal division times	mAP	mINP	R1
1	71.11	25.58	93.68
2	72.05	28.45	92.60
<b>3 (Ours)</b>	<b>77.15</b>	<b>36.82</b>	<b>96.72</b>
4	73.10	30.01	94.82
5	70.88	25.32	90.98

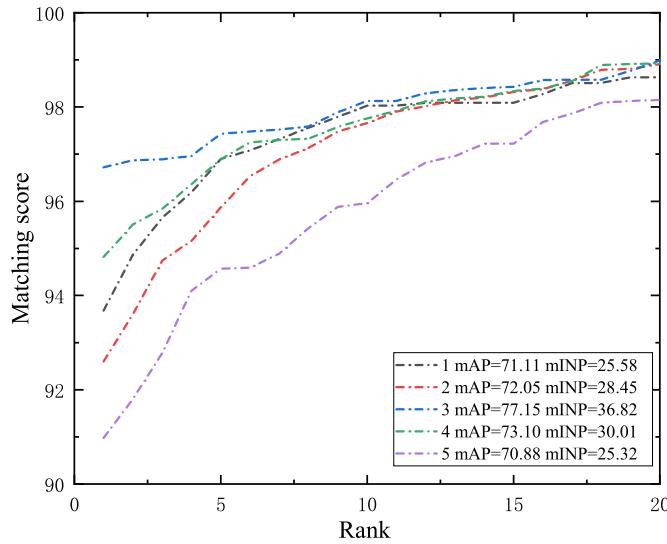


Fig. 3. The CMC curves of ablation study on VeRi-776 dataset.

**Table 4**  
The ablation studies result of the MFG-Net on VeRi-776 dataset.

Method	mAP	mINP	R1
Baseline	62.43	22.71	89.70
+ global branch (w/o $L_{hard}$ mining triplet)	68.55	27.62	91.17
+ global branch (w/ $L_{hard}$ mining triplet)	70.68	33.00	93.78
+ local branch	75.10	35.54	94.88
+CAM	<b>77.15</b>	<b>36.82</b>	<b>96.72</b>

#### 4.3. Ablation study

To investigate the effectiveness and contribution of the feature map horizontally quartered in our MFG-Net, we conducted ablation study on VeRi-776 dataset. The evaluation results are shown in Table 3. The CMC curve of the ablation study on this dataset is shown in Fig. 3, in which the effectiveness of the horizontal division times of the feature map is visually compared from Rank 1 to 20.

From Table 3 and Fig. 3 we can see, the feature map horizontally quartered (the number of horizontal divisions is equal to 3) surpasses the others. The comparison results show that the feature map horizontally quartered is more suitable for fully extracting the fine-grained features

of the vehicle in MFG-Net and can effectively improve the performance of vehicle re-id.

In addition, we choose ResNet50 with softmax cross-entropy loss as the baseline, and four vehicle re-id network frameworks are then designed based on the baseline:

- 1) + global branch (w/o  $L_{hard}$  mining triplet);
- 2) + global branch (w/  $L_{hard}$  mining triplet);
- 3) + local branch;
- 4) + CAM.

The ablation studies result of the MFG-Net on VeRi-776 dataset are shown in Table 4.

From Table 4, we can see that compared with “baseline”, “baseline + global branch (w/o  $L_{hard}$  mining triplet)” improves the 6.12% mAP, 4.91% mINP, and 1.47% Rank1 on VeRi-776 dataset. The results show that global branch effectively improve the accuracy of vehicle re-id. After applying hard mining triplet loss, “baseline + global branch (w/  $L_{hard}$  mining triplet)” outperforms “baseline + global branch (w/o  $L_{hard}$  mining triplet)” by a large margin (2.13% mAP, 5.38% mINP, and 2.61% Rank1). This result validates the effectiveness of the hard mining triplet loss to optimize the distance between positive and negative sample pairs. After adding local branch, the network improves the 4.42% mAP, 2.54% mINP, and 1.10% Rank1 on VeRi-776 dataset. We can clearly see that the feature map segmentation can reduce the interference of background information and improve the accuracy of vehicle re-id. By adding CAM, the network achieves 2.05%, 1.28% and 1.84% improvement in mAP, mINP and Rank1 on VeRi-776. This result show that CAM can enhance the feature extraction capability of network.

#### 4.4. VeRi-776

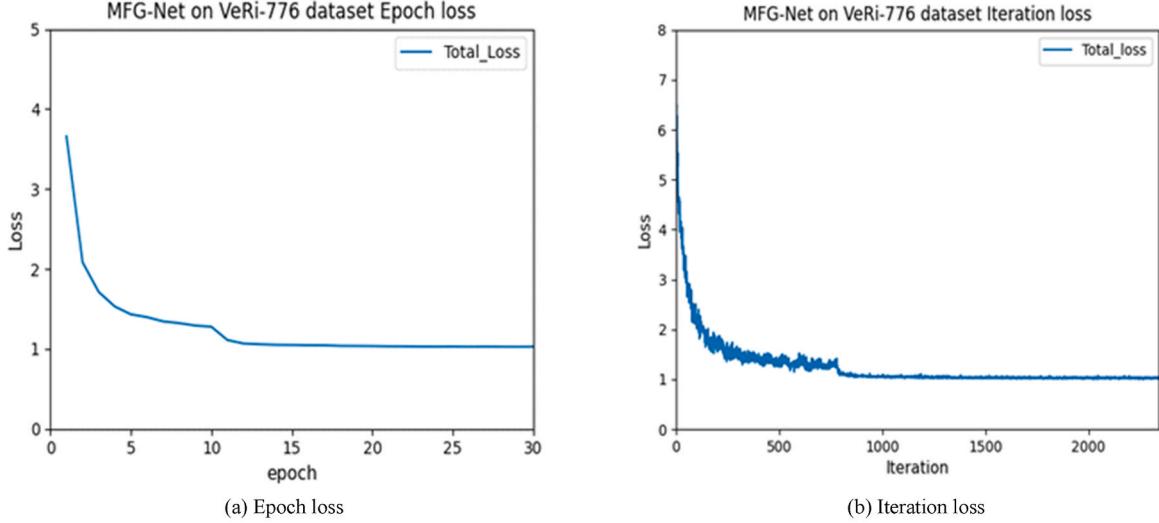
During the train stage on VeRi-776, the learning rate is decreased by a factor of 0.1 after the 10<sup>th</sup> and 20<sup>th</sup> epoch, till the end 30<sup>th</sup> epoch. And the batch-size of training and testing are both 48. The training loss of MFG-Net on VeRi-776 is shown in Fig. 4.

Our study presents performance comparisons between the proposed MFG-Net and representative state-of-the-art methods (i.e., DAVR [1], VRSDNet [22], VAMI + ST [2], RAM [12], GRF + GGL [15], BS [23], CCA [3], MRM [4], SPAN w/CPDM [24], TCL + SL [25], UMTS [26], and MsDeep [5]) on VeRi-776. According to the experimental results in Table 5, our method improves by 1.25% and 1.11% on mAP and Rank1, respectively, compared with the second-best method UMTS without using any auxiliary information (such as VAMI + ST using the vehicle’s spatio-temporal information). The new metric mINP is applied to the field of vehicle re-identification for the first time and reaches 36.82% on VeRi-776.

#### 4.5. VehicleID

During the training stage on VehicleID, the learning rate is decreased by a factor of 0.1 after the 15<sup>th</sup> and 30<sup>th</sup> epoch, till the end 40<sup>th</sup> epoch. And the batch-size of training and testing are 24 and 32, respectively. The training loss of MFG-Net on VehicleID is shown in Fig. 5.

VehicleID has three test sets, namely Test800, Test1600 and Test2400. VAMI [2], DAVR [1], RAM [12], CCA [3], MRM [4], TCL + SL [25], GRF + GGL [15], MSA [27], BS [23] and MsDeep [5] are also included in our comparison list. Table 6 shows the comparison results on VehicleID three test sets. Compared with the current best results, our model MFG-Net has improved by approximately 2.18%–3.01% on mAP. At the same time, the mINP metric reaches 68.32%, 64.38%, and 60.03% respectively. This demonstrates that MFG-Net requires less effort to find all the correct vehicle matches, verifying the ability of mINP.



**Fig. 4.** The training loss of MFG-Net on VeRi-776.(a) Epoch loss. (b) Iteration loss.

**Table 5**  
Comparison with state-of-the-art methods on VeRi-776.

Methods	mAP	mINP	Rank1
DAVR [1]	52.36	—	83.25
VRSNet [22]	53.45	—	83.49
VAMI + ST [2]	61.32	—	85.92
RAM [12]	61.50	—	88.60
GRF + GGL [15]	61.7	—	89.4
BS [23]	67.55	—	90.23
CCA [3]	68.05	—	91.71
MRM [4]	68.55	—	91.77
SPAN w/CPDM [24]	68.9	—	94.0
TCL + SL [25]	68.97	—	93.92
MsDeep [5]	74.50	—	95.10
UMTS [26]	75.9	—	95.61
<b>MFG-Net</b>	<b>77.15</b>	<b>36.82</b>	<b>96.72</b>

#### 4.6. VRIC

During the training stage on VRIC, the learning rate is decreased by a factor of 0.1 after the 15<sup>th</sup> and 30<sup>th</sup> epoch, till the end 40<sup>th</sup> epoch. And the batch-size of training and testing are both 32. The training loss of

MFG-Net on VRIC is shown in Fig. 6.

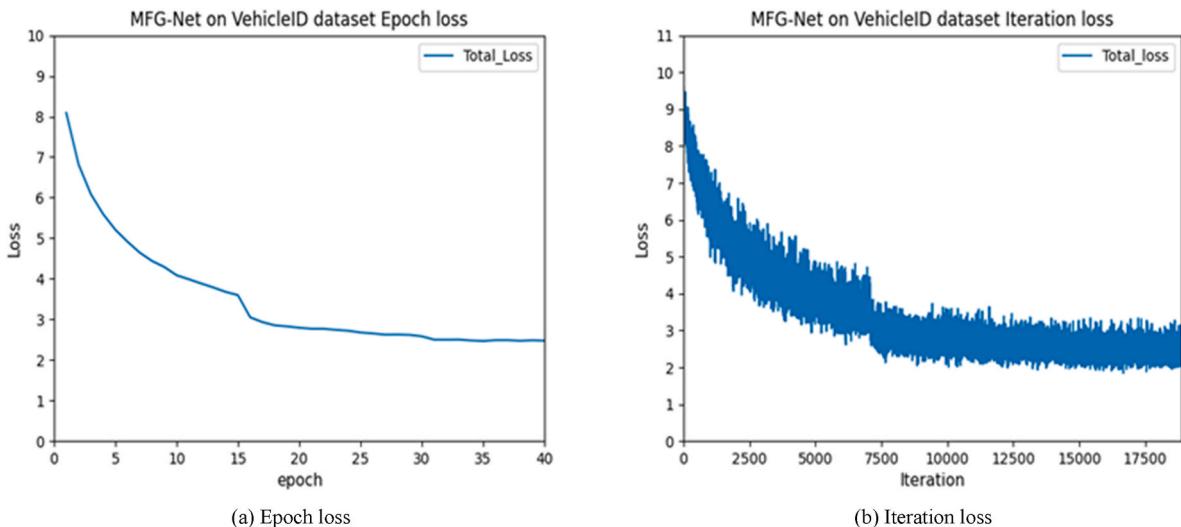
To further validate our proposed method, we carry out experiments on the VRIC dataset, which contains more challenging training examples.

It can be seen from Table 7 that the performance of our method MFG-Net is better than all other methods listed in Table 7 including MSVF [18], GLAMOR [6], BS [23] and PGAN [7] on mAP and Rank1 metric. In addition, mINP reaches 55.81% and shows that our model has better retrieval capabilities.

## 5. Discussion

In our work, we mainly use feature map segmentation combined with channel attention mechanism to extract multiple fine-grained features of vehicles to improve the accuracy of vehicle re-id. Differ from what we think [24,28,29], adopt feature alignment to adjust the image to the same scale, which is conducive to similar feature matching. Finally, these methods enhance the performance of re-id model.

In order to prove the robustness and generalization of the MFG-Net, our retrieval results are visualized on VeRi-776, VRIC and VehicleID, as shown in Fig. 7. We can see that the MFG-Net is more robust and generalized to vehicles in different poses.

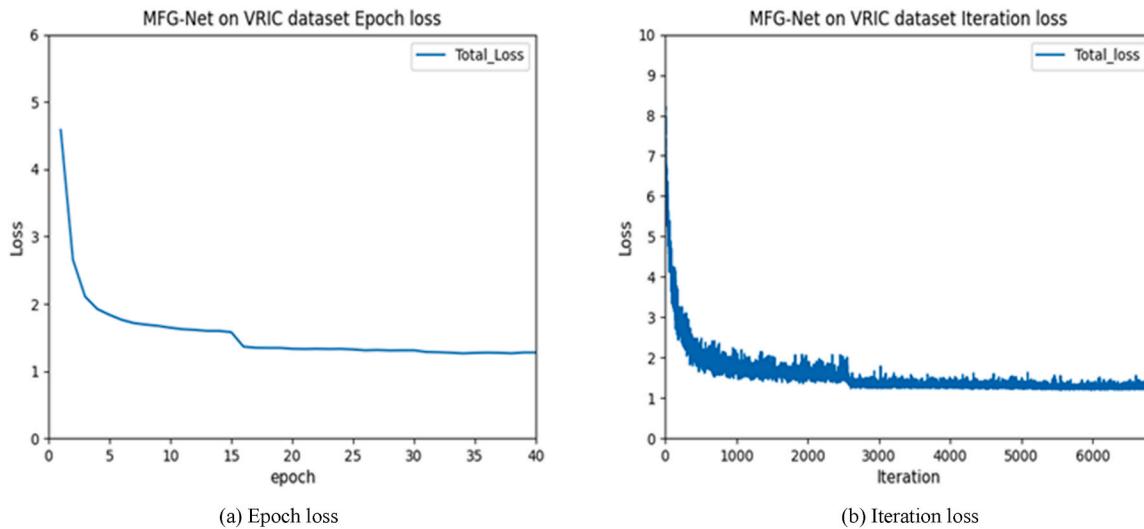


**Fig. 5.** The training loss of MFG-Net on VehicleID. (a) Epoch loss, (b) Iteration loss.

**Table 6**

Comparison with state-of-the-art methods on VehicleID.

Methods	Test800			Test1600			Test2400		
	mAP	mINP	Rank1	mAP	mINP	Rank1	mAP	mINP	Rank1
VAMI [2]	–	–	63.12	–	–	52.87	–	–	47.34
DAVR [1]	72.40	–	68.04	70.11	–	66.48	67.97	–	64.07
RAM [12]	–	–	75.20	–	–	72.30	–	–	67.70
CCA [3]	78.89	–	75.51	76.53	–	73.60	73.11	–	70.08
MRM [4]	80.02	–	76.64	77.32	–	74.20	74.02	–	70.86
TCL + SL [25]	80.13	–	74.97	77.26	–	72.84	75.25	–	71.20
GRF + GGL [15]	–	–	77.1	–	–	72.7	–	–	70.0
MSA [27]	80.31	–	77.55	77.11	–	74.41	75.55	–	72.91
BS [23]	86.19	–	78.80	81.69	–	73.41	78.16	–	69.33
MsDeep [5]	84.30	–	81.20	81.00	–	78.00	78.60	–	75.60
MFG-Net	<b>88.37</b>	<b>68.32</b>	<b>82.02</b>	<b>84.70</b>	<b>64.38</b>	<b>78.69</b>	<b>81.01</b>	<b>60.03</b>	<b>75.16</b>



**Fig. 6.** The training loss of MFG-Net on VRIC. (a) Epoch loss. (b) Iteration loss.

**Table 7**

Comparison with state-of-the-art methods on VRIC.

Methods	mAP	mINP	Rank1
MSVF [18]	47.50	—	46.61
GLAMOR [6]	76.48	—	75.58
BS [23]	78.55	—	69.09
PGAN [7]	84.80	—	78.00
<b>MFG-Net</b>	<b>84.86</b>	<b>55.81</b>	<b>79.56</b>

## 6. Conclusion

This paper presents an effective multi-fine-grained feature extraction network for vehicle re-identification. Using the proposed MFG-Net, the fine-grained features of the vehicle can be fully utilized, which provides robustness against vehicle obstructions. Our experiments show that the proposed MFG-Net is superior to multiple state-of-the-art vehicle re-identification methods on VeRi-776, VehicleID and VRIC datasets. In addition, we introduce a new evaluation metric: mINP. The experimental verification not only confirms the retrieval ability of our model, but also verifies the effectiveness of the new metric.

In the future, we will not only enrich the types of vehicles, but also

try to combine pedestrians and vehicles for re-identification. This idea will provide solid technical support for the construction of smart cities.

## **Author contribution statement**

Yan Xu: Writing – review & editing.; Leilei Rong: Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing.; Xiaolei Zhou: Investigation, Writing – review & editing.; Xuguang Pan: Validation, Formal analysis, Visualization, Supervision.; Xianglan Liu: Validation, Resources, Visualization, Supervision.

## **Declaration of competing interest**

The authors declare that they have no known competing financial interests or personal relationships that could be perceived as influencing the work reported in this paper.

This work was supported by the Natural Science Foundation of China (11547037, 11604181), Shandong Province Postgraduate Education Quality Curriculum Project (SDYKC19083), Shandong Province Post-graduate Education Joint Training Base Project (SDYJD18027), Hisense Group research and development Center Project, and the Scholarship Fund of SDUST.



**Fig. 7.** Visualization of MFG-Net retrieval results on VeRi-776, VRIC and VehicleID. The green and red boxes represent correct matching vehicles and wrong matching vehicles, respectively.

## References

- [1] Peng J, Wang H, Xu F, Fu X. Cross domain knowledge learning with dual-branch adversarial network for vehicle re-identification. Neurocomputing 2020;401: 133–44. <https://doi.org/10.1016/j.neucom.2020.02.112>.
- [2] Zhou Y, Shao L. Viewpoint-aware attentive multi-view inference for vehicle Re-identification. In: IEEE comput soc conf comput vision pattern recognit; 2018. p. 6489–98. <https://doi.org/10.1109/CVPR.2018.00679>.
- [3] Peng J, Jiang G, Chen D, Zhao T, Wang H. Eliminating cross-camera bias for vehicle re-identification. Multimed Tool Appl 2020;1–17. <https://doi.org/10.1007/s11042-020-09987-z>.
- [4] Peng J, Wang H, Zhao T, Fu X. Learning multi-region features for vehicle re-identification with context-based ranking method. Neurocomputing 2019;359: 427–37. <https://doi.org/10.1016/j.neucom.2019.06.013>.
- [5] Cheng Y, Zhang C, Gu K, Qi L, Gan Z. Multi-scale deep feature fusion for vehicle Re-identification. In: IEEE int conf acoust speech signal process proc; 2020. p. 1928–32. <https://doi.org/10.1109/ICASSP40776.2020.9053328>.
- [6] Suprem A, Pu C. Looking GLAMORous: vehicle Re-id in heterogeneous cameras networks with global and local attention. arXiv preprint arXiv:2002.02256, <https://arxiv.org/abs/2002.02256>; 2020.
- [7] Zhang X, Zhang R, Cao J, Gong D, You M. Part-guided attention learning for vehicle re-identification. 2019. arXiv preprint arXiv:1909.06023.
- [8] Shen Y, Xiao T, Li H, Yi S. Learning deep neural networks for vehicle re-id with visual-spatiotemporal path proposals. IEEE Int Conf Comput Vision 2017;1900–9. <https://doi.org/10.1109/ICCV.2017.210>.
- [9] Jiang N, Xu Y, Zhou Z, Wu W. Multi-attribute driven vehicle re-identification with spatial-temporal re-ranking. In: IEEE int conf image process; 2018. p. 858–62. <https://doi.org/10.1109/ICIP.2018.8451776>.
- [10] Jin W. Multi-camera vehicle tracking from end-to-end based on spatial-temporal information and visual features. ACM Int Conf Proc Ser 2019:227–32. <https://doi.org/10.1145/3374587.3374629>.
- [11] Gu J, Jiang W, Luo H, Yu H. An efficient global representation constrained by Angular Triplet loss for vehicle re-identification. Pattern Anal Appl 2021;24(1): 367–79. <https://doi.org/10.1007/s10044-020-00900-w>.
- [12] Liu X, Zhang S, Huang Q, Gao W. Ram: a region-aware deep model for vehicle re-identification. In: IEEE int conf on multimedia and expo (ICME); 2018. p. 1–6. <https://doi.org/10.1109/ICME.2018.8486589>.
- [13] 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; 2019. p. 3997–4005. <https://doi.org/10.1109/CVPR.2019.00412>.
- [14] Khorramshahi P, Kumar A, Peri N, Rambhatla SS, Chen JC, Chellappa R. A dual-path model with adaptive attention for vehicle re-identification. In: Proceedings of the IEEE/CVF international conference on computer vision; 2019. p. 6132–41. <https://doi.org/10.1109/ICCV.2019.00623>.
- [15] Liu X, Zhang S, Wang X, Hong R. Group-group loss-based global-regional feature learning for vehicle re-identification. IEEE Trans Image Process 2019;29:2638–52. <https://doi.org/10.1109/TIP.2019.2950796>.
- [16] Liu X, Liu W, Mei T, Ma H. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In: European conference on computer vision; 2016. p. 869–84. [https://doi.org/10.1007/978-3-319-46475-6\\_53](https://doi.org/10.1007/978-3-319-46475-6_53).
- [17] 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; 2016. p. 2167–75. <https://doi.org/10.1109/CVPR.2016.238>.
- [18] Kanaci A, Zhu X, Gong S. Vehicle re-identification in context. In: German conference on pattern recognition; 2018. p. 377–90. [https://doi.org/10.1007/978-3-030-12939-2\\_26](https://doi.org/10.1007/978-3-030-12939-2_26).

- [19] 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; 2016. p. 770–8. <https://doi.org/10.1109/CVPR.2016.90>.
- [20] Hermans A, Beyer L, Leibe B. In defense of the triplet loss for person re-identification. arXiv preprint arXiv:1703.07737. 2017, <https://arxiv.org/abs/1703.07737>.
- [21] Ye M, Shen J, Lin G, Xiang T, Shao L, Hoi SC. Deep learning for person re-identification: a survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*; 2021. <https://doi.org/10.1109/TPAMI.2021.3054775>.
- [22] Zhu J, Du Y, Hu Y, Zheng L, Cai C. VRSDNet: vehicle re-identification with a shortly and densely connected convolutional neural network. *Multimed Tool Appl* 2019;78(20):29043–57. <https://doi.org/10.1007/s11042-018-6270-4>.
- [23] Kumar R, Weill E, Aghdasi F, Sriram P. A strong and efficient baseline for vehicle re-identification using deep triplet embedding. *J Artif Intell Soft Comput Res* 2020; 10(1):27–45. <https://doi.org/10.2478/jaiscr-2020-0003>.
- [24] Meng D, Li L, Liu X, Li Y, Yang S, Zha ZJ, 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; 2020. p. 7103–12. <https://doi.org/10.1109/CVPR42600.2020.00713>.
- [25] He X, Zhou Y, Zhou Z, Bai S, Bai X. Triplet-center loss for multi-view 3d object retrieval. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 1945–54. <https://doi.org/10.1109/CVPR.2018.00208>.
- [26] Jin X, Lan C, Zeng W, Chen Z. Uncertainty-aware multi-shot knowledge distillation for image-based object re-identification. *Proc AAAI Conf Artif Intell* 2020;34(7): 11165–72. <https://doi.org/10.1609/aaai.v34i07.6774>.
- [27] Zheng A, Lin X, Dong J, Wang W, Tang J, Luo B. Multi-scale attention vehicle re-identification. *Neural Comput Appl* 2020;32(23):17489–503. <https://doi.org/10.1007/s00521-020-05108-x>.
- [28] Zheng Z, Zheng L, Yang Y. Pedestrian alignment network for large-scale person re-identification. *IEEE Trans Circ Syst Video Technol* 2018;29(10):3037–45. <https://doi.org/10.1109/TCSVT.2018.2873599>.
- [29] Liu X, Liu W, Zheng J, Yan C, Mei T. Beyond the parts: learning multi-view cross-part correlation for vehicle re-identification. In: Proceedings of the 28th ACM international conference on multimedia; 2020. p. 907–15. <https://doi.org/10.1145/3394171.3413578>.