Computers, Materials & Continua
DOI:10.32604/cmc.2021.021627
Article

Tech Science Press

A Multi-Feature Learning Model with Enhanced Local Attention for Vehicle Re-Identification

Wei Sun[1,2,*], Xuan Chen[3], Xiaorui Zhang[1,3], Guangzhao Dai[2], Pengshuai Chang[2] and Xiaozheng He[4]

[1]Jiangsu Collaborative Innovation Center of Atmospheric Environment and Equipment Technology, Nanjing University of Information Science & Technology, Nanjing, 210044, China
[2]School of Automation, Nanjing University of Information Science & Technology, Nanjing, 210044, China
[3]Engineering Research Center of Digital Forensics, Ministry of Education, Jiangsu Engineering Center of Network Monitoring, School of Computer and Software, Nanjing University of Information Science & Technology, Nanjing, 210044, China
[4]Rensselaer Polytechnic Institute, Troy, NY, 12180, USA
[*]Corresponding Author: Wei Sun. Email: sunw0125@163.com
Received: 29 June 2021; Accepted: 30 July 2021

Abstract: Vehicle re-identification (ReID) aims to retrieve the target vehicle in an extensive image gallery through its appearances from various views in the cross-camera scenario. It has gradually become a core technology of intelligent transportation system. Most existing vehicle re-identification models adopt the joint learning of global and local features. However, they directly use the extracted global features, resulting in insufficient feature expression. Moreover, local features are primarily obtained through advanced annotation and complex attention mechanisms, which require additional costs. To solve this issue, a multi-feature learning model with enhanced local attention for vehicle re-identification (MFELA) is proposed in this paper. The model consists of global and local branches. The global branch utilizes both middle and high-level semantic features of ResNet50 to enhance the global representation capability. In addition, multi-scale pooling operations are used to obtain multi-scale information. While the local branch utilizes the proposed Region Batch Dropblock (RBD), which encourages the model to learn discriminative features for different local regions and simultaneously drops corresponding same areas randomly in a batch during training to enhance the attention to local regions. Then features from both branches are combined to provide a more comprehensive and distinctive feature representation. Extensive experiments on VeRi-776 and VehicleID datasets prove that our method has excellent performance.

Keywords: Vehicle re-identification; region batch dropblock; multi-feature learning; local attention

1 Introduction

Vehicle re-identification (ReID) is a core technology of intelligent transportation systems. With the development of artificial intelligence and big data technology, vehicle re-identification has a wide range of applications in suspect tracking, unmanned parking lot management, smart logistics, and unmanned driving. It can be regarded as an image retrieval problem, aiming to retrieve images of the query vehicles from a large gallery, where the images are taken by different cameras from multi-views. Especially if the license plate is blocked, removed, destroyed, or in other special scenarios, the technology will become the only option [1–3].

Many previous studies directly adopted convolutional neural networks (CNNs) to learn the robust representation of vehicle images [4–6]. However, CNNs tend to focus on global observation while ignoring other discriminative vehicle parts. Specially, under different viewing angles, the appearances of vehicles change significantly, which leads to the instability of global features [7]. Local location features contain more stable and distinguishable information, such as windows, headlights, license plates, etc [8,9]. These features are critical for judging similar vehicles and will not change greatly with environmental changes, so the features are more robust. Hence, many studies introduce local location features to cope with illumination, posture, perspective, and occlusion [10–12]. However, these methods require additional annotations, such as direction key point annotations, local position annotations, etc. To avoid additional annotations, some studies focus on applying attention mechanisms for vehicle re-identification to increase focus on local regions [13,14], but complex attention mechanism modules need to be designed additionally. The above issues motivate us to propose a new vehicle re-identification framework that can simultaneously extract global features and strengthen local feature learning, without pre-labeling and complex attention mechanisms.

In this paper, we propose a multi-feature learning model with enhanced local attention for vehicle re-identification(MFELA). This model is composed of two branches, including global and local feature enhancement branches. The middle layer of Layer3 and last layer of Layer4 in ResNet50 [15] are used to obtain middle and high-level semantic features simultaneously, and multi-scale pooling operations of GAP and GMP are integrated to enhance the global representation of the vehicle. However, this module does not consider the subtle differences between similar cars, especially cars of the same brand, model, and color. Therefore, a method called Region Batch Dropblock (RBD) is proposed to form the local feature enhancement branch. RBD is an improvement of existing person re-recognition work BDB [16]. RBD first divides the global features into several regions which enforces the network to extract discriminative details in each region. Then RBD randomly drops the corresponding areas of multiple regions in batches, namely the same semantic car parts, to enhances the attentive feature learning of each region of the remaining area .While BDB drops the global features directly, which leads to insufficient attention to local details.

In summary, the main contributions of the study are as follows:

(1) The proposed MFELA model can obtain more representative global features. Both middle and high-level semantic features of ResNet50 are simultaneously used to extract more comprehensive global features. In addition, multi-scale pooling operations bring multi-scale information. By observing the overall appearances of the vehicles, the module attempts to maximize the separation of the identity in the feature space.
(2) RBD is proposed to learn the subtle differences between similar vehicles. RBD first adopts the idea of regional division to encourage the deep model to learn distinguishing features.

Then, RBD drops the corresponding areas in a batch during training, which further reinforces the attentive feature learning of local regions. The method is efficient and straightforward, does not require additional labels and complex attention mechanisms.

(3) The global semantic and local subtle discriminative cues are jointly learned in the final embedding of the vehicle. Ablation studies and experiments on two mainstream vehicle ReID datasets demonstrate the effectiveness of MFELA. It significantly improves ReID accuracy over the baseline and outperforms the most existing vehicle Re-ID methods.

The rest of the paper is organized as follows: Section 2 reviews the relevant works and Section 3 introduces the proposed model of vehicle Re-ID. Extensive experimental results are presented and analyzed in Section 4 and finally, the conclusions are summarized in Section 5.

2 Related Work

With the rapid development of deep learning, vehicle ReID has gradually become a hot topic. Existing vehicle ReID methods based on deep learning can be roughly divided into the following four categories.

(1) **Vehicle ReID based on global features.** Global features only focus on the vehicle's overall appearance, such as color, model type, etc. Among the earliest attempts for vehicle ReID that involve deep learning, Liu et al. [4] proposed to extract global features by using a convolutional neural network combined with traditional methods. Wang et al. [17] extracted richer features through adequate labeling information using additional attribute information, such as specific brands and models. Zhang et al. [5] reduced the redundancy of global features by applying the SE block to automatically obtain the importance of each channel feature to improve Densenet121. However, due to the influence of low resolution, illumination variation, and cross-camera perspective, the appearance of the same vehicle is visually prone to change, which makes it difficult to obtain complete and stable global features.

(2) **Vehicle ReID based on local location features.** Recent related researches use the local location features to improve representation capabilities. The RAM model proposed by Liu et al. [10] also adopted the idea of regional division to extract local features. However, the RAM model directly learns from the local region without in-depth mining of local information. Wang et al. [11] defined 20 key points of the car body and extracted local vehicle features based on the pre-defined key points. Besides, He et al. [12] predefined local region locations and used the target detection algorithm to extract features of local regions. Although these methods can extract stable local features, they all need to annotate key points and regional regions in advance. As the datasets under traffic scenarios become larger, annotations will consume a lot of time and effort, and the accuracy cannot be guaranteed.

(3) **Vehicle ReID based on local attention mechanism.** Some researchers have begun to pay attention to the application of attention mechanisms in vehicle ReID. Khorramshahi et al. [13] proposed a dual-path attention network, which can adaptively select and focus on key feature points and azimuth information. Zhang et al. [14] adopted the attention mechanism to assign higher weights to salient local areas to achieve higher attention. Whereas most attention mechanisms are complex and difficult to come up with. Meanwhile, it increases parameters of models which further leads to the difficulty of training.

(4) **Vehicle ReID based on GAN.** Another effective strategy is to introduce generative adversarial networks (GAN) [18] into vehicle ReID. Zhou et al. [19] employed GAN to realize effec-

tive multi-view feature reasoning and generate multi-view features from single-view. Lou et al. [20] utilized two GANs to generate multi-perspective and hard samples, respectively. It must be mentioned that the overall model structure and training process are also more complicated when using GAN for image generation.

3 Methodology

The entire pipeline of the proposed MFELA consists of two main parts: global feature extraction and local feature enhancement. The network structure is illustrated in Fig. 1

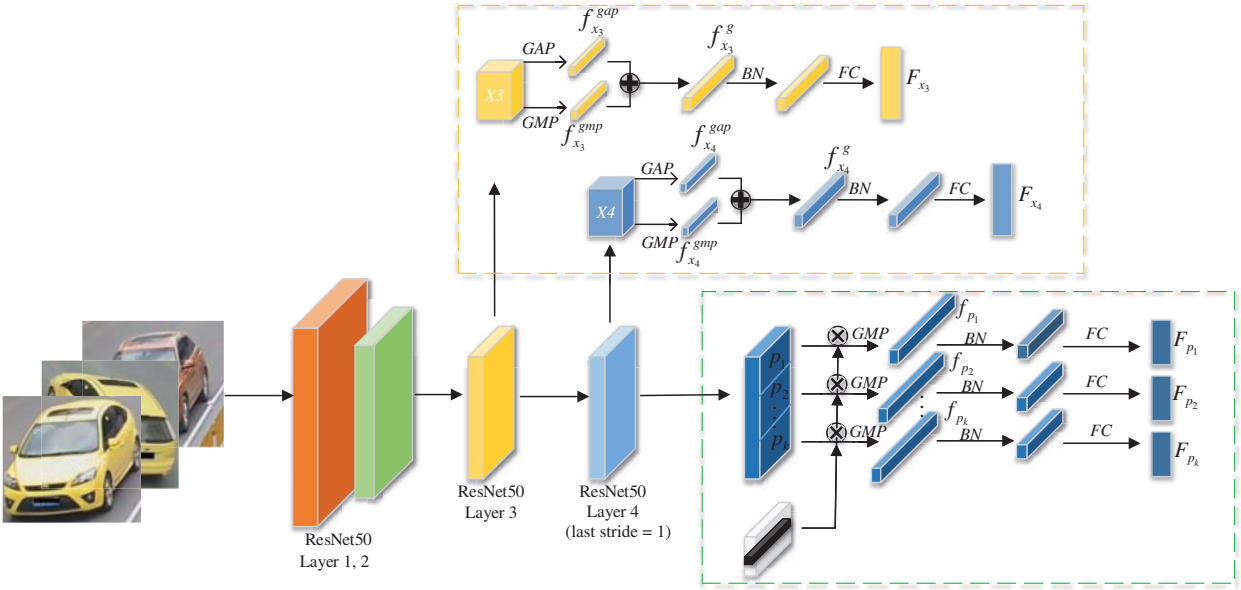

Figure 1: The entire pipeline of the proposed MFELA (yellow dotted box is the global feature extraction module to extract macroscopic appearance features, another branch enhances local features to learn the subtle differences between similar cars, and the RBD method to extract local features)

In MFELA, we use ResNet-50 as the backbone network. The purpose of global feature extraction module is to extract macroscopic appearance features. Specifically, the features obtained by RESNET-50 Layer3 and Layer4 is defined as X_3 and X_4, respectively. We respectively do global average pooling (GAP) and global max pooling (GMP) operation on X_3 to obtain $f_{x_3}{}^{gap}$ and $f_{x_3}{}^{gap}$. These two features are fused to get the first global feature $f_{x_3}{}^{g}$. Similarly, the same procedure is implemented on X_4 to get the second global feature $f_{x_4}{}^{g}$. Another branch enhances local features to learn the subtle differences between similar vehicles, and the RBD method is proposed to extract local features. Concretely, divide the feature map into k regions from top to bottom: $p_1, \ldots, p_k$, and zero out the units in the dropping area of each region to get local features $f_{p_1}, \ldots, f_{p_k}$, respectively. The Global branches and local branches are combined for multi-feature learning. The model can simultaneously learn the overall structure and fine-grained distinction information of vehicles, and enhance the discriminative ability of the model. In the process of

feature extraction, triples and softmax loss are used for training. In the following sections, we discuss the details of the network structure and model training.

3.1 Global Feature Extraction

ResNet-50 is utilized as the backbone network, we append a branch denoted as a global branch. Previous work [21] has proved that the middle layer of ResNet-50 contains more semantic information. Hence, except for the last layer, the middle layer is also employed to obtain more comprehensive representation information. We define features extracted by these two layers as X_3, X_4 and then simultaneously do GAP and GMP operation on X_3, X_4 to obtain ${f_{x_3}}^{gap}$, ${f_{x_3}}^{gmp}$, ${f_{x_4}}^{gap}$, ${f_{x_4}}^{gmp}$. Subsequently, use superposition to merge the two features ${f_{x_3}}^{gap}$ and ${f_{x_3}}^{gmp}$ to obtain the global feature ${f_{x_3}}^{g}$. Also, the feature ${f_{x_3}}^{g}$ can be expressed as ${f_{x_3}}^{g} = {f_{x_3}}^{gap} + {f_{x_3}}^{gmp}$. The same operation is done on X_4 to get the global feature ${f_{x_4}}^{g}$. Then a batch normalization (BN) layer and fully connected layer (FC) are added after ${f_{x_3}}^{g}$ and ${f_{x_4}}^{g}$, and prediction vectors F_{x_3}, F_{x_4} are obtained.

GAP preserves overall data features, while GMP acquires texture features. In addition, the fusion feature of GAP and GMP is used to obtain multi-scale information to enhance vehicle representation.

3.2 Local Feature Enhancement

If different vehicles have a similar global appearance, the differences mainly exist in local areas, such as the number and location of annual inspection signs on the windshield and window decorations. As shown in Fig. 2. Therefore, we design a local branch to enhance local features.

Figure 2: Examples of different vehicles with similar global appearance (each column shows two different vehicles, and the differences in local areas are highlighted with red circles)

Firstly, the feature maps extracted from Layer4 are divided into k non-overlapping regions vertically, and then $p_1,\ldots,p_k$ parts are obtained. Through experiments, if k is 6, the network achieves the best performance. In this way, the whole feature map is divided into regions, and each region is studied separately, which improves the learning strength of local regions. Subsequently, the RBD method randomly drops the corresponding areas of these regions in a batch. In particular, for each non-overlapping regions, there is a mask of the same size multiplied by it then the units in the dropping area of the mask are zeroed out. GMP is done on each region to get local features $f_{p1},\ldots,f_{pk}$, and the BN layer and the FC layer follow closely. Finally, the prediction vectors $F_{p_1},\ldots,F_{p_k}$ are obtained. When each region randomly drops the same area, the model pays more attention to the remaining parts to strengthen the local attention learning. A simple schematic of RBD is shown in Fig. 3.

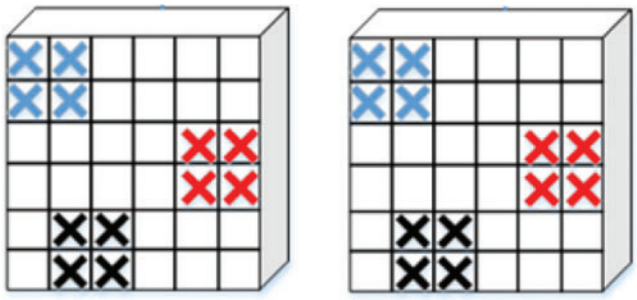

Figure 3: A simple schematic of Region Batch DropBlock (the two feature maps of the same batch are evenly divided into three parts, the same area in each region is dropped, and the crossed part in the figure indicates that it is dropped)

3.3 Training

The overall objective loss function is composed of global and local losses. Both branches are trained by the combined loss of hard triple loss [22] and cross-entropy loss.

The hard triplet loss is described as follows:

$$l_{htri}(a,p,n) = \max[\partial + max(D_{a,p}) - min(D_{a,n}), 0] \tag{1}$$

a, p and n are anchor, positive, and negative samples. ∂ is the distance margin parameter. $D_{a,p}$ and $D_{a,n}$ are the Euclidean distances calculated from the features of a and p/n in feature space.

Besides, $l_{cross}(y,\hat{y})$is cross-entropy loss:

$$l_{cross}(y,\hat{y}) = -\frac{1}{n}\sum_{i=1}^{n} y_i \log(\hat{y}_i) \tag{2}$$

y is the ground-truth vector, $\hat{y}$ is the predicted probability vector, n represents the type of sample, that is, the ID class of the vehicle.

Hard triple loss of global feature extraction branch is calculated by ${f_{x_3}}^{gap}$, ${f_{x_3}}^{gmp}$, ${f_{x_4}}^{gap}$ and ${f_{x_4}}^{gmp}$, and cross entropy loss is calculated by the prediction vector F_{x_3} and F_{x_4}. The global loss is described as follows:

$$l_{global} = l_{htri}^{f_{x_3}^{gap}} + l_{htri}^{f_{x_3}^{gmp}} + l_{htri}^{f_{x_4}^{gap}} + l_{htri}^{f_{x_4}^{gmp}} + l_{cross}^{F_{x_3}} + l_{cross}^{F_{x_4}} \tag{3}$$

Simultaneously, Hard triple loss of local feature enhancement branch is calculated using $f_{p_1},\ldots, f_{p_k}$, and cross entropy loss is calculated using the prediction vector $F_{p_1},\ldots,F_{p_k}$. The part loss is described as

$$l_{part} = \sum_{i=1}^{k} (l_{htri}^{f_{p_i}} + l_{cross}^{F_{p_i}}) \tag{4}$$

The final loss of the two-branch network is defined as

$$l = l_{global} + l_{part} \tag{5}$$

In the test stage, global features ${f_{x_3}}^g$ and ${f_{x_4}}^g$ are combined with local features $f_{p_1},\ldots,f_{p_k}$ to obtain the overall features of the vehicle. Finally, cosine distance is used to compare the features of each pair of query and test images to determine their similarity.

4 Experiments

4.1 Datasets and Evaluation Metrics

We evaluate the proposed model on two mainstream datasets, namely, VehicleID and Veri776.

VehicleID [23] is a large-scale vehicle ReID dataset. It contains 221, 763 pictures of approximately 26,267 vehicles. Images of the vehicle are taken from both front and rear perspectives. The test phase is divided into three scale datasets, namely large, medium, and small datasets. Cumulative Match Curve (CMC) for top 1 (CMC@1) and top 5 (CMC@5) matches are adopted as evaluation metrics for this dataset.

Veri776 [4] is also widely used in the vehicle Re-ID. It provides images captured from 20 cameras at different perspectives, consisting of 499,357 images of 776 different vehicles. Therefore, this dataset can reflect the actual situation of real-world traffic scenarios. Evaluation metrics employed for Veri776 are mean Average Precision (mAP) and CMC@1.

4.2 Implementation Details

Our network uses RTX2070 GPU for training, and the batch size is 32. Each batch has eight identities, so each identity contains four instance images in a batch. We use ResNet50 as the backbone CNN for training, which has been pre-trained on ImageNet [24] and the last stride is set to 1. All image sizes are cropped to 256×256 and normalized. Random flip and random erasing [25] are used for data enhancement. The model is trained 70 epochs in total. We used a warm-up strategy [26], the initial learning rate is 3.5×10^{-6} and changed to 3.5×10^{-4} in the 10th, and then drops to 3.5×10^{-5} and 3.5×10^{-6} in the 30th, 55th. Adam optimization model is adopted in the training process.

4.3 Comparison with State-of-art Methods

We compare the MFELA model against the recent state of the art methods, which includes: LOMO [27], FACT [4], AGNet [17], ODJL [5], RAM [10], OIFE [11], PRN [12], AAVER [13], PGAN [14], VAMI [19], and EALN [20]. The results of the comparison are presented in Tabs. 1 and 2.

Table 1: The mAP and CMC@1 on VeRi776

Method	VeRi-776	
	mAP	CMC@1
LOMO	9.8	23.9
FACT	18.7	51.9
AGNet	66.3	90.9
ODJL	75.5	94.8
RAM	61.5	88.6
OIFE	48.0	65.9
PRN	74.3	94.3
AAVER	58.5	88.7
PGAN	79.3	**96.5**
VAMI	50.1	77.0
EALN	57.4	84.4
MFELA	**81.9**	96.3

Table 2: The mAP and CMC@1 on VehicleID

Method	VehicleID/(Small)		VehicleID/(Medium)		VehicleID/(Large)	
	CMC@1	CMC@5	CMC@1	CMC@5	CMC@1	CMC@5
LOMO	19.7	32.1	19.0	29.5	15.3	25.6
FACT	49.5	68.0	44.6	64.2	39.9	60.50
AGNet	71.15	83.78	69.23	81.41	65.74	78.28
ODJL	81.3	94.3	78.9	92.1	76.5	89.2
RAM	75.2	91.5	72.3	87.0	67.7	84.5
OIFE	-	-	-	-	67.0	82.9
PRN	78.4	92.3	75.0	88.3	74.2	86.4
AAVER	72.5	93.2	66.9	89.4	60.2	84.9
PGAN	-	-	-	-	77.8	**92.1**
VAMI	63.1	83.3	52.9	75.1	47.3	70.3
EALN	75.1	88.1	71.8	83.9	69.3	81.4
MFELA	**85.5**	**97.0**	**80.2**	**93.9**	**78.7**	91.8

We observe that, compare with the RAM which also adopts region division, the mAP of VeRi-776 has increased by 20.4%, the CMC@1 has increased by 7.5%, and the VehicleID has also been greatly improved. The reason is that our local branch not only adopts region division but also employs RBD to increases the attention to each region. Compared to other methods, PGAN and our MFELA have achieved a significant performance improvement. It can be seen that the two methods are on the same level. Because both approaches focus on the saliency local area. Noteworthy, PGAN assigns higher weights to salient local areas to gain higher attention, which involves complex attention model design. However, the proposed MFELA achieves

excellent results without additional tags and complex attention mechanisms, which is efficient and straightforward.

4.4 Ablation Study

4.4.1 Selection of Parameter k

In this paper, the Region Batch DropBlock method is proposed to extract local features. Specifically, the feature maps with the size of 16×16 need to be divided into k regions from top to bottom. To make sure that the region is a block rather than a strip, k ranges from 1 to 8. We do value experiment on VERI-776. As shown in the Tabs. 3, if $k = 6$, the model shows the best performance. For the VehicleID dataset, we also directly use $k = 6$, and the experimental results are also satisfactory, reflecting the high robustness of the model.

Table 3: The selection of parameter k on VeRi776

MFELA	VeRi-776	
	mAP	CMC@1
k = 1	79.5	96.1
k = 2	79.9	95.8
k = 3	81.5	96.0
k = 4	80.7	96.0
k = 5	81.2	95.6
k = 6	**81.9**	**96.3**
k = 7	81.5	96.1
k = 8	81.4	96.2

4.4.2 Effectiveness of Each Branch

We conduct ablation experiments on two datasets to prove the effectiveness of each branch. The experimental results are shown in Tabs. 4 and 5. The global branch is more effective than the baseline. Because our global branch incorporates middle and high level information and leverages different multi-scale pooling operations to enhance vehicle representation. In the absence of a global branch, the local branch still performs better than the baseline. It is proved that RBD has a strong ability to enhance the attention learning of local regions. Adding the global branch could further improve the performance. This suggests that the two branches are mutually reinforcing, and both are important to the final performance. The motivation behind the two-branch structure in the MFELA Network is that it obtains more representative global and local features simultaneously.

To better understand the influence of RBD in MFELA, we visualize heatmaps of vehicle images. We use class activation mapping (CAM) [28] to visualize global features. Through CAM, we can observe where the network focuses on. In Fig. 4, it is observed that the global branch is better than the baseline. Because the global branch utilizes both Layer3 and Layer4 features and integrates GAP and GMP multi-scale pooling features to enhance the global representation of vehicles. The MELAE model is constructed based on the global branch. The Fig. 4 shows the local area of concern is covered more widely and more accurately. Thus, RBD reflects a strong focus on local regions.

Table 4: Ablation study about each branch of MFELA for VeRi776

Method	VeRi-776	
	mAP	CMC@1
Baseline	76.2	95.3
Global Branch	78.5	95.6
Local Branch	78.9	95.9
Both Branch	**81.9**	**96.3**

Table 5: Ablation study about each branch of MFELA for VehicleID

Method	VehicleID(Small)		VehicleID(Medium)		VehicleID(Large)	
	CMC@1	CMC@5	CMC@1	CMC@5	CMC@1	CMC@5
Baseline	79.9	94.8	75.9	90.3	73.2	88.1
Global Branch	82.3	95.4	77.4	92.0	75.4	89.0
Local Branch	83.1	96.0	78.6	92.9	75.7	89.6
Both Branch	**85.5**	**97.0**	**80.2**	**93.9**	**78.7**	**91.8**

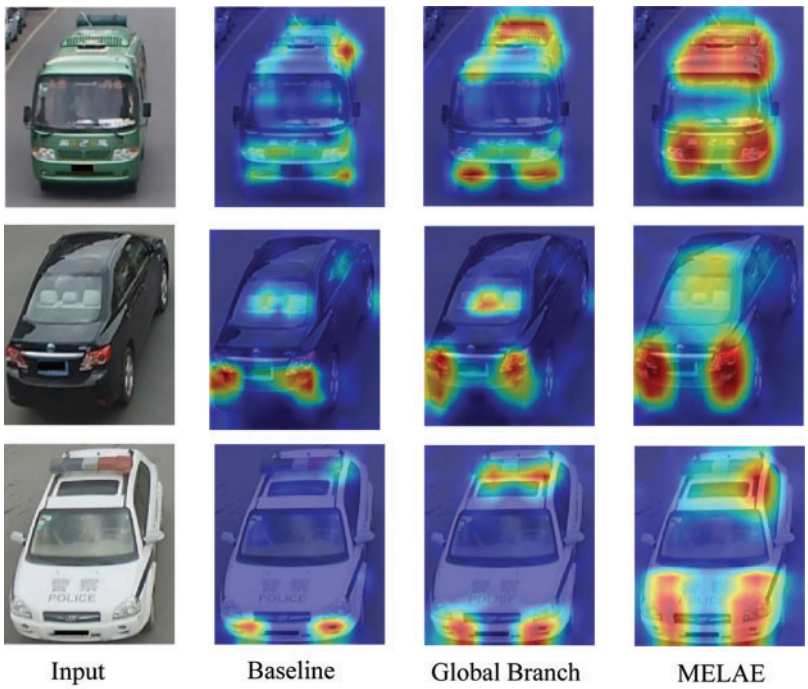

Figure 4: Class Activation Maps of different vehicles (the highlighted pixels indicate that they play a more important role in determining the similarity between vehicles)

To further illustrate the effectiveness and practicability of the proposed framework, we visualize some examples of query results on VeRi-776, as shown in Fig. 5. True matches are green, and false matches are red. The left is the query results of the baseline, and the right is the query results of the proposed method MELAE. As we can see, the number of correct images returned by the query of the MELAE model far exceeds that of the baseline, which shows that our model queries are more accurate.

Figure 5: Visualization of the ranking list on VeRi-776 (the left-hand side is retrieval results obtained from baseline, while the right-hand side is obtained by the proposed method)

5 Conclusion

In this paper, we propose a multi-feature learning model with enhanced local attention for vehicle re-identification. This model could obtain more representative global and local features at the same time. Global features incorporate both middle and last layers to extract more comprehensive global features. Besides, the global branch integrates GAP and GMP multi-scale pooling features to obtain multi-scale information. In addition, the RBD adopts regional division and regional batch dropping to strengthen the attention learning of local regions, which is efficient and straightforward. Extensive experiments are conducted to show the effectiveness of our model.

However, most vehicle images in the VeRi-776 and VehicleID datasets are complete, with fewer occluded. In the future, we will attempt to build a vehicle dataset that includes occlusion conditions and further improve the proposed model to adapt to the occlusion environment and

make the model more robust. At the same time, we will further lightweight the model to realize the re-ID in real-time traffic scenes.

Funding Statement: This work was supported, in part, by the National Nature Science Foundation of China under Grant Numbers 61502240, 61502096, 61304205, 61773219; in part, by the Natural Science Foundation of Jiangsu Province under grant numbers BK20201136, BK20191401; in part, by the Postgraduate Research & Practice Innovation Program of Jiangsu Province under Grant Numbers SJCX21_0363; in part, by the Priority Academic Program Development of Jiangsu Higher Education Institutions (PAPD) fund.

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

References

[1] S. D. Khan and H. A. Ullah, "Survey of advances in vision-based vehicle re-identification," *Computer Vision and Image Understanding,* vol. 182, pp. 50–63, 2019.

[2] F. Wu, S. Y. Yan, J. S. Smith and B. L. Zhang, "Vehicle re-identification in still images: Application of semi-supervised learning and re-ranking," *Signal Processing: Image Communication,* vol. 76, pp. 261–271, 2019.

[3] X. R. Zhang, X. Chen, W. Sun and K. Ge, "Progress of vehicle re-identification research based on deep learning," *Computer Engineering,* vol. 46, no. 11, pp. 1–11, 2020.

[4] X. C. Liu, W. Liu, H. D. Ma and H. Y. Fu, "Large-scale vehicle re-identification in urban surveillance videos," in *in Proc. ICME*, Seattle, USA, pp. 1–6, 2016.

[5] X. R. Zhang, X. Chen, W. Sun and X. Z. He, "Vehicle Re-identification model based on optimized densenet121 with joint loss," *Computers, Materials & Continua,* vol. 67, no. 3, pp. 3933–3948, 2021.

[6] W. Sun, X. R. Zhang, X. Z. He, Y. Jin and X. Zhang, "A two-stage vehicle type recognition method combining the most effective gabor features," *Computers, Materials & Continua,* vol. 65, no. 3, pp. 2489–2510, 2020.

[7] Y. T. Shen, T. Xiao, H. S. Li, S. Yi and X. G. Wang, "Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals," in *Proc. ICCV*, Venice, Italy, pp. 1900–1909, 2017.

[8] W. Sun, H. J. Du, S. B. Nie and X. Z. He, "Traffic sign recognition method integrating multi-layer features and kernel extreme learning machine classifier," *Computers, Materials & Continua,* vol. 60, no.1, pp. 147–161, 2019.

[9] M. Y. Jiang, X. M. Zang, Y. Yu, Z. C. Bai, Z. D. Zheng *et al.*, "Robust vehicle re-identification via rigid structure prior," in *Proc. CVPR*, Virtual, pp. 4026–4033, 2021.

[10] X. B. Liu, S. L. Zhang, Q. M. Huang and W. Gao, "Ram: a region-aware deep model for vehicle re-identification," in *Proc. ICME*, San Diego, USA, pp. 1–6, 2018.

[11] Z. D. Wang, L. T. Tang, X. H. Liu, Z. L. Yao, S. Yi *et al.*, "Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification," in *Proc. ICCV*, Venice, Italy, pp. 379–387, 2017.

[12] B. He, J. Li, Y. F. Zhao and Y. H. Tian, "Part-regularized near-duplicate vehicle re-identification," in *Proc. CVPR*, Long Beach, CA, pp. 3997–4005, 2019.

[13] P. Khorramshahi, A. Kumar, N. Peri, S. S. Rambhatla, J. C. Chen *et al.*, "A dual path model with adaptive attention for vehicle re-identification," in *Proc. ICCV*, Seoul, Korea, pp. 6132–6141, 2019.

[14] X. Y. Zhang, R. F. Zhang, J. W. Cao, D. Gong, M. Y. You *et al.*, "Part-guided attention learning for vehicle re-identification," arXiv preprint arXiv: 1909.06023, 2019.

[15] K. M. He, X. Y. Zhang, S. Q. Ren and J. Sun, "Deep residual learning for image recognition," in *Proc. CVPR*, Las vegas, USA, pp. 770–778, 2016.

[16] Z. Dai, M. Q. Chen, M. X. D. Gu, S. Y. Zhu and P. Tan, "Batch dropblock network for person re-identification and beyond," in *Proc. ICCV*, Seoul, Korea, pp. 3691–3701, 2019.

[17] H. B. Wang, J. Peng, D. Y. Chen, GQJ,T Zhao *et al.*, "Attribute-guided feature learning network for vehicle reidentification," *IEEE MultiMedia,* vol. 27, no. 4, pp. 112–121, 2020.

[18] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley *et al.*, "Generative adversarial networks," arXiv preprint arXiv: 1406.2661, 2014.

[19] Y. Zhou and L. Shao, "Aware attentive multi-view inference for vehicle re-identification," in *Proc. CVPR*, Salt Lake, USA, pp. 6489–6498, 2018.

[20] Y. H. Lou, Y. Bai, J. Liu, S. Q. Wang and L. Y. Duan, "Embedding adversarial learning for vehicle re-identification," *Transactions on Image Processing,* vol. 28, no. 8, pp. 3794–3807, 2019.

[21] Q. Yu, X. B. Chang, Y. Z. Song, T. Xiang and T. M. Hospedales, "The devil is in the middle: Exploiting mid-level representations for cross-domain instance matching," arXiv preprint arXiv: 1711.08106, 2017.

[22] A. Hermans, L. Beyer and B. Leibe. "In defense of the triplet loss for person re-identification," arXiv preprint arXiv: 1703.07737, 2017.

[23] H. Y. Liu, Y. H. Tian, Y. W. Yang, L. Pang and T. J. Huang, "Deep relative distance learning: Tell the difference between similar vehicles," in *Proc. CVPR*, Las vegas, USA, pp. 2167–2175, 2016.

[24] J. Deng, W. Dong, R. Socher, L. J. L. K. Li *et al.*, "Imagenet: A large-scale hierarchical image database," in *Proc. CVPR, Miami*, USA, 2009.

[25] Z. Zhong, L. Zheng, G. L. Kang, S. Z. Li and Y. Yang, "Random erasing data augmentation," in *Proc. AAAI*, New York, USA, pp. 13001–13008, 2020.

[26] P. Goyal, P. Dollár, R. Girshick, P. Noordhuis, L. Wesolowski *et al.*, "Accurate, large minibatch sgd: Training imagenet in 1 hour," arXiv preprint arXiv: 1706.02677, 2017.

[27] S. C. Liao, Y. Hu, X. Y. Zhu and S. T. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proc. CVPR*, Boston, USA, pp. 2197–2206, 2015.

[28] B. L. Zhou, A. Khosla, A. Lapedriza, A. Oliva and A. Torralba, "Learning deep features for discriminative localization," in *Proc. CVPR*, Las vegas, USA, pp. 2921–2929, 2016.