Article - Engineering, Technology and Techniques

A Generated Multi Branch Feature Fusion Model for Vehicle Re-identification

Hu Zhijun[1,2*]
https://orcid.org/0000-0002-9965-4594

Raja Soosaimarian Peter Raj[3]
https://orcid.org/0000-0002-7216-2207

Sun Lilei[1]
https://orcid.org/0000-0002-7369-5494

Wu Lian[1]
https://orcid.org/0000-0003-0269-9962

Cheng Xianjing[4]
https://orcid.org/0000-0002-1644-5384

[1]Guizhou University, College of Computer Science and Technology, Guiyang, China; [2]Guangxi Normal University, College of Mathematics and Statistics, Guilin, China; [3]Vellore Institute of Technology, Department of Computer Science and Engineering, Vellore, Tamilnadu, India. [4]Zunyi Normal University, Zunyi, China.

*Correspondence: huzhijun@mailbox.gxnu.edu.cn; Tel.: +86-13788588847 (H. Z.).

HIGHLIGHTS

- A new framework is designed in which several strong complementary models are derived. On this basis, a simple multi branch network model named GMBFF is designed.
- Two different feature fusion methods namely single fusion method (SFM) and multi fusion method (MFM) are designed.

Abstract: Vehicle re-id play a very import role in recent public safety, it has received more and more attention. The local features (e.g. hanging decorations and stickers) are widely used for vehicle re-id, but the same local feature exists in one perspective, but not exactly exists in other perspectives. In this paper, we firstly use experiments to verify that there is a low linear correlation between different dimension global features. Then we propose a new technique which uses global features instead of local features to distinguish the nuances between different vehicles. We design a vehicle re-identification method named a generated multi branch feature fusion method (GMBFF) to make full use of the complementarity between global features with different dimensions. All branches of the proposed GMBFF model are derived from the same model and there are only slight differences among those branches. Each of those branches can extract highly

discriminative features with different dimensions. Finally, we fuse the features extracted by these branches. Existing research uses the fusing features for fusion and we use the global vehicle features for fusion. We also propose two different feature fusion methods which are single fusion method (SFM) and multi fusion method (MFM). In SFM, features for fusion with larger dimension occupy more weight in fused features. MFM overcomes the disadvantage of SFM. Finally, we carry out a lot of experiments on two widely used datasets which are VeRi-776 dataset and Vehicle ID dataset. The experimental results show that our proposed method is much better than the state-of-the-art vehicle re-identification methods.

Keywords: vehicle re-identification; deep learning; feature fusion; correlation coefficient matrix; global feature.

INTRODUCTION

Due to the super computing power of computers, intelligent transportation [1,2] and autonomous vehicles technology [3,4] have developed rapidly. Vehicle re-identification technology plays an increasingly important role in public safety. More and more researchers begin to focus on this topic [5-8,9,10]. Vehicle re-identification is to find the same vehicle as the query vehicle under different cameras. There are two difficulties in vehicle re-identification. First, the intra-class difference is large. Two images of the same vehicle may look very different due to the factors like different lighting, camera distance and camera perspectives. Second, the inter-class difference is small. Two different vehicles of the same color and model produced by the same manufacturer may look very similar from the same perspective. In order to solve these two difficulties, some scholars proposed the methods which use the local features of vehicles to make a more subtle distinction [11-17]. For example, the hanging ornaments, stickers or logos on the window of two vehicles are used to determine whether these two vehicles are same or not. To be specifically, different branches are used to extract global features and local features respectively. Local features and global features are integrated [11,13,14] or local features are used to suppress global features to generate better global features [12]. However, the vehicle has eight different orientations [11], which make the local features in one perspective which doesn't exist in the other perspective. So, it is little bit difficult to capture the local features of the vehicle and the local features also have certain limitations.

We realize that it is difficult for us to actively extract the desired local features, such as the pendant on the window. We can't segment the pendant on the vehicle in the original image, because most of the vehicle images have no pendant. It's meaningless to design a special segmentation method for pendant. From the references of [18,19] we can know that, if the feature map of each convolutional layer of the deep convolution network is visualized, the feature map of the high-level network contains local information that can make subtle distinction of the image. From the works [11-14], it is found that local feature has been actually used in the global feature. If we use a branch to extract global features, and use another branch to extract local feature, this is the second time this local feature to be used. This can better emphasize this local feature and make better use of this local feature for subtle distinction. We believe that when making subtle distinctions, global features can completely replace local features. We think that local features are completely available in global features, because the size of the dimension determines how much information a global feature contains. We think that with the increase of feature dimensions, the global features may get more and more information such as windows and lights. Also, it gets more local feature information about the scratches in vehicles, hanging accessories etc. It also has the coordinate information of these local features. In other words, the larger the dimension, the more detailed vehicle information (i.e. local features) can be included in the global features. However, when the dimension of the global feature is too large and if it has more distinguishing information then the learning ability of the established model is not enough to fully learn so much information. Therefore, if the feature dimension is too large then the learning ability of the model will be reduced. So, we can't completely replace local features with global features by infinitely increasing feature dimensions.

We notice that, for a given query image, there are great re-identification ranking differences between the global features with different dimensions (see Figure 2). There are two reasons for this case. First, the information concerned by these features with different dimensions is not exactly the same. Second, their ability to learn the same information is not exactly the same. So, in order to make full use of all the information contains in those global features with different dimensions, we propose a simple and novel method named feature fusion method of different dimension features (GMBFF). This method uses a multi branch network to extract features with different dimensions, we also call this multi branch network GMBFF for short. The number of branches of GMBFF network can be increased or decreased flexibly. Each branch is derived from

the same network and can be independently applied to vehicle re-identification and can extract highly discriminative feature. Each branch of our GMBFF network is trained independently. The features extracted from these branches have different dimensions and the re-identification ranking results of these branches have low correlation which is the necessary prerequisite for the effectiveness of the fusion for these features. The proposed method is simple and effective for two reasons. First, we designed branches which can extract highly discriminative features. Second, the linear correlation between each of the two branches is low.

In summary, the contributions of this paper are as follows:

1) On the basis of resnet50, a new framework is designed in which several strong complementary models are derived. On this basis, a simple multi branch network model named GMBFF is designed. Each branch of the GMBFF model is independent of each other and trained independently. In the evaluating phase, the vehicle features extracted from all branches are fused into new vehicle features for re-identification evaluation.
2) Two different feature fusion methods namely single fusion method (SFM) and multi fusion method (MFM) are designed. In SFM, features for fusion with larger dimension occupy more weight in fused features. MFM overcomes the disadvantage of SFM.
3) Lot of experiments have been done on two universal datasets which are VeRi-776 and VehicleID. The experimental results of GMBFF are better than the experimental results of state-of-the-art vehicle re-identification methods which shows the feasibility of the proposed method.

The arrangement of this paper is as follows: the second part elaborates the methods proposed in this paper, the third part introduces the dataset and configurations used in the experiment, the fourth part verifies the feasibility of the proposed method through a large number of experiments and the fifth part summarizes the proposed methods.
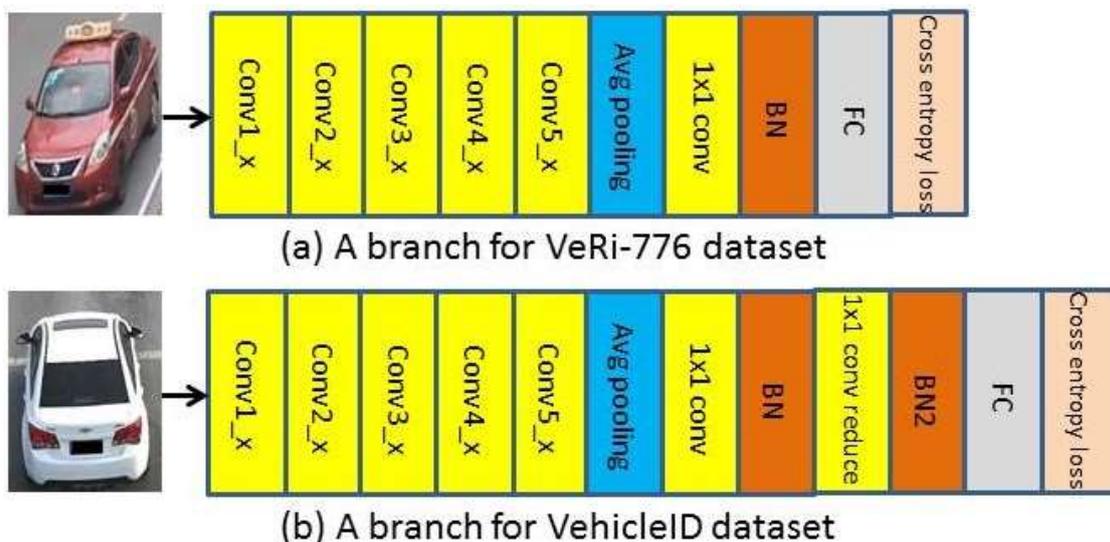

Figure 1. A branch in GMBFF model. Because the number of training IDs of the VehicleID dataset is far greater than that in the VeRi-776 dataset. In order to reduce the parameters of GMBFF model, we set up different network branches for the two datasets. (a) is a branch for the VeRi-776 dataset, and (b) is a branch for the VehicleID dataset. For (a), we add a fully-connected layer (FC1) and a batchnormal layer (BN1) to the baseline. For (b), it is compared with (a), an additional fully-connected layer (FC2) and a batchnormal (BN2) layer are added. The output size of FC2 is 256. For (a) and (b), the output of FC1 is vehicle features with dimension C. According to the different values of C, different branches are derived.

PROPOSED METHOD

The proposed method GMBFF is composed of several independent branches which are derived from the same network and all the branches only have slight differences. It can extract highly discriminative features with different dimensions. The low linear correlation between these features ensures the effectiveness of fusion. The number of branches of GMBFF network can be increased or decreased flexibly.

We also designed two different feature fusion methods to ensure that we can make full use of the information contained in these features. This section mainly introduces the branch generation method, feature fusion method and the correlation analysis between features extracted from each branch.

Branch generation method

In order to use global features to replace local features for obtaining more fine-grained information, features with larger dimensions contain more information is preferred. However, since the established model is not enough to learn so much information, it is not possible to completely replace local features with global features by infinitely increasing feature dimensions. Because features with different dimensions contain different information, and the re-identification ranking results of features with different dimensions are also very different (see Figure 2). Our purpose is to generate multiple branches which can extract highly discriminative features with different dimensions, and finally fuse these features extracted by these branches to make full use of the complementarity of the information contained in these features.
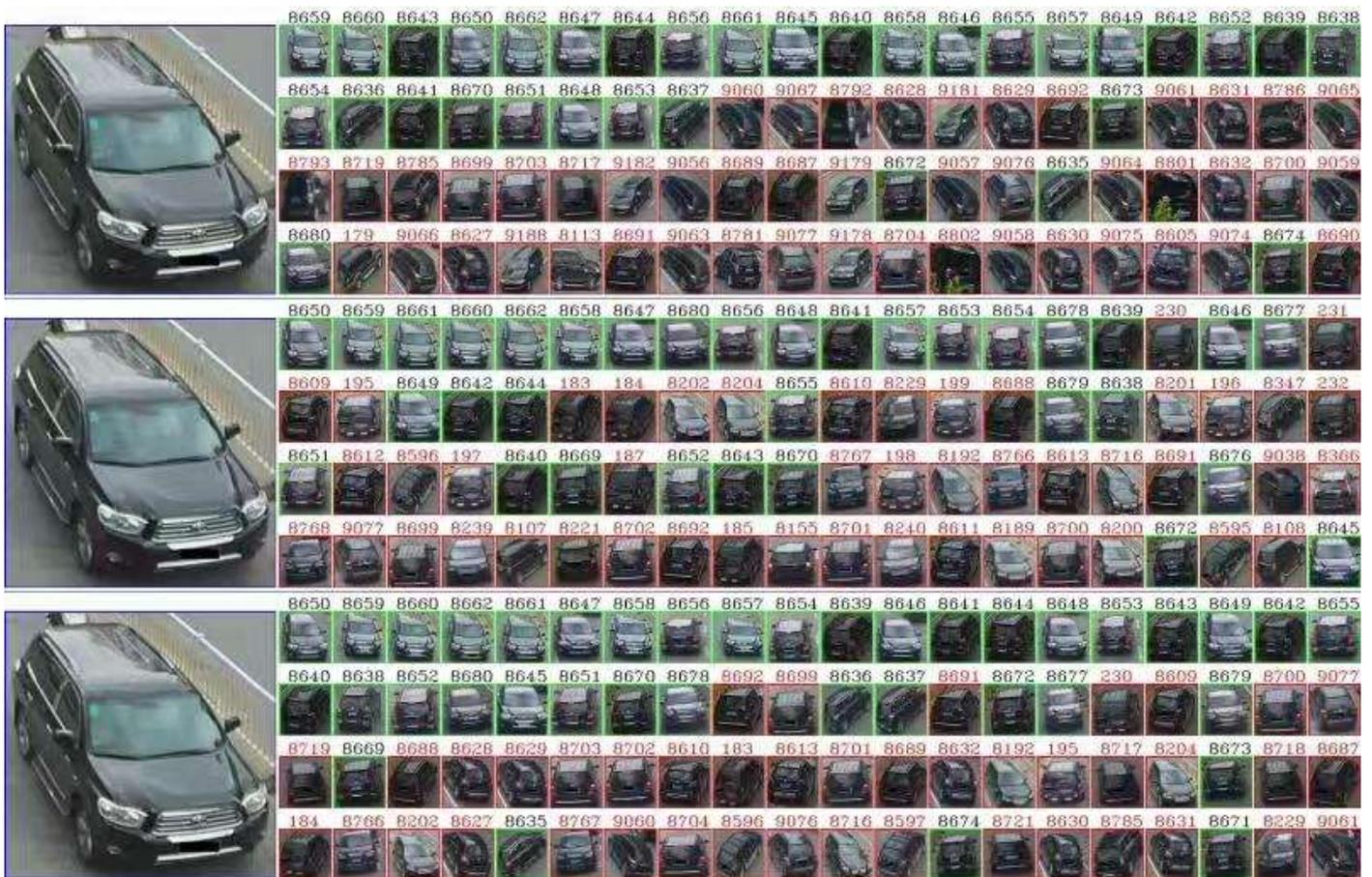
Figure 2. Comparison of re-identification results between different features with different dimensions. The vehicle image with the same vehicle ID and camera ID as the query image is removed. The top, middle and bottom are the re-identification results of features with dimensions of 512, 2048 and the fusion of these two features, respectively. They are ranked according to the Euclidean distance from small to large, and each ranking result only displays the top 80 matching images. On the left is the same query image, and on the right is the query results. The green boxes are the positive images of the query image, and the red boxes are the negative images of the query image. The number above given in each small image is the original number of the vehicle image in the gallery set.

Because resnet 50 network is widely used in person re-identification [20-23] and vehicle re-identification [11,12,24,25], it shows that resnet50 is a very efficient network in the field of re-identification, so we choose resnet50 as the baseline of this paper. However, the average pooling layer of resnet50 network can only extract the features with fixed dimension (2048), so the network must be changed. Considering that fully-connected layer can not only change the feature dimension, but also can re-integrate the information of each channel, so we add a fully-connected layer (FC1 layer in Figure 1) behind the average pooling layer of resnet50, and extracted different dimensional features by changing the output channels. Let the output size

of the FC1 layer be C, where C is the value of its channels. We set C to 128, 256, 384, 512, 768, 1024, 1536, 2048, 3072 & 4096, and then we derive ten networks that can extract features with different dimensions. We use these ten networks as the branches of GMBFF (some branches can be reduced according to the actual traffic conditions). In order to get more discriminative features, we continue to add a batchnormal layer (the BN1 layer in Figure 1) behind the FC1 layer to normalize the features which makes the features as Gaussian distributed and easier to learn the vehicle information. Finally, we use the output of the BN layer as the features extracted by these branches, and the corresponding dimensions of these features are 128, 256, 384, 512 768, 1024, 1536, 2048, 3072 and 4096, respectively. In the training phase, the ten branches are trained separately. It should be emphasized that due to the different value of C, these ten branches have different learning ability, so they must be set with different learning rate adjustment strategies. From Table 2, it can be seen that with the increase of C, the faster the branch learning speed, the more sensitive the learning rate adjustment, and the earlier the learning rate adjustment.

For the dataset with a large number of training vehicle IDs, we add a FC1 layer and a BN1 layer according to the above method. When the value of C is large, the number of parameters of the subsequent linear classification layer (FC1 layer) will be large. At this time, we can continue to add a fully-connected layer to reduce the number of parameters. Because the number of training IDs in the VehicleID dataset is 13164, which is far greater than 576 (the number of training IDs in the VeRi-776 dataset). So for this dataset, we continue to add a fully-connected layer (FC2 in Figure 1 (b) with an output dimension of 256) and a batchnormal layer (the BN2 layer in Figure 1) behind the BN layer.

In the phrase of inference, we fuse all the features extracted by these branches, then we treat the fused feature as the final feature extracted by the model. We extract the fused feature of all query images and gallery images, and use the fused feature to do the similarity search between query images and gallery images. Further more, we design two feature fusion methods, let us begin to introduce them in the following section.

Feature fusion method

Because the dimensions of the features extracted by the branches of our proposed GMBFF network are different, these features contain a lot of complementary information. We know that the importance of all the ten features are almost the same in the fused feature, but if we simply concatenate them by the dimension, the F_{4096} will certainly occupy the largest weight in the fused feature, and the F_{128} will occupy the smallest weight in the fused feature (see Figure 3). The purpose of fusing these features is to maximize the use of the supplementary information between features of different dimensions, so as to make full use of the local feature information hidden in these global features. It is difficult to determine which global feature should take a greater weight in the fused feature, so we design two feature fusion methods. The first is called single fusion method (SFM) and the other is called multi fusion method (MFM), as shown in Figure 3.
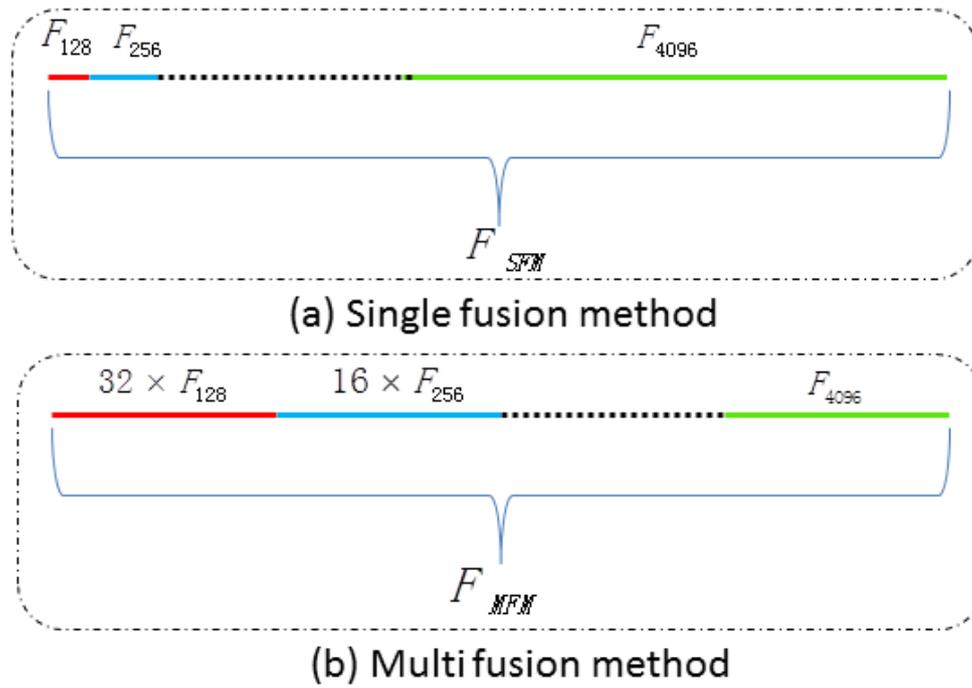
Figure 3. Two feature fusion methods. From the single fusion method (SFM), we can see that the weight of F_{4096} in the fused feature F_{SFM} is significantly greater than that of F_{128}, F_{256}, etc, but in the fused feature F_{MFM}, their weights are almost the same. The idea of designing F_{MFM} is to make the weight of each feature in the fused feature the same.

Feature fusion method

Assume the ten branches are $branch_{128}$, $branch_{256}$,···, $branch_{4096}$, and the extracted features of these branches are F_{128}, F_{256},···, F_{4096}, respectively, and let the fused feature of the method SFM is F_{SFM}, and that of MFM is F_{MFM}. Where

$$F_{SFM} = concatenate(F_{128}, F_{256}, \cdots, F_{4096}) \tag{1}$$

$$F_{MFM} = concatenate(32 \times F_{128}, 16 \times F_{256}, \cdots, F_{4096}) \tag{2}$$

Where, $32 \times F_{128}$ means thirty-two F_{128} s are concatenated, and so on. That is to say, F_{SFM} simply concatenates the features extracted by all the branches. The idea of designing F_{MFM} is to make the weight of each feature in the fused feature the same, so that the roles of all branches are balanced. The numbers of F_{128}, F_{256}, F_{384}, F_{512}, F_{768}, F_{1024}, F_{1536}, F_{2048}, F_{3072}, F_{4096} used in the concatenation of F_{MFM} is 32, 16, 11, 8, 6, 4, 3, 2, 1 and 1, respectively. MFM method makes up for the disadvantage that the feature with larger dimension used for fusion in SFM method has larger weight in the fused feature F_{SFM}. The reason is that, as shown in the top of Figure 3, in the SFM method, the dimension of F_{4096} is 4096, while the dimension of F_{128} is 128, so in F_{SFM}, the weight of F_{4096} is 32 times of that of F_{128}, but before fusing the features, for the two features, we don't know which feature should take up the greater weight in the fused feature.

Table 1. The influence of the dimensions of fused feature on FPS. The dimensions of F_{128} and $\overline{F}$ and $\overline{\overline{F}}$ are 128, 13824 and 41088, respectively, but they are little influence on FPS.

Method	F_{128}	$\overline{F}$	$\overline{\overline{F}}$
FPS	212.34	210.92	208.31

Timeliness analysis for SFM and MFM

For the two fused features F_{SFM} and F_{MFM}, the dimension of F_{SFM} is 13824, while that of F_{MFM} is 41088. Will these high dimensions affect the timeliness of re-identification? In order to study the influence of this high dimension on FPS (number of matching images with query image per second), we did the following experiments, using the branch of our GMBFF network with the feature dimension of 128 to do the re-identification experiment, we transform F_{128} into $\overline{F}$ and $\overline{\overline{F}}$, where

$$\overline{F} = concatenate(108 \times F_{128})$$ (3)

$$\overline{\overline{F}} = concatenate(321 \times F_{128})$$ (4)

$\overline{F}$ and $\overline{\overline{F}}$ are concanated by one hundred and eight F_{128}s and three hundred and twenty-one F_{128}, respectively, and their dimensions are 13824 (equals to the dimensions of F_{SFM} and 41088 (equals to the dimensions of F_{MFM}), respectively. For this branch, we use F_{128}, $\overline{F}$ and $\overline{\overline{F}}$ as feature to carry out re-identification experiments to observe the influence of dimensions on FPS. The experiment accuracy of these three features is the same. From Table 1, we can see that although the feature dimension of $\overline{\overline{F}}$ is three hundred and twenty-one times of the feature dimension of F_{128}, the increase of this dimension has little impact on FPS. The results show that the time consumption of the re-identification task is mainly focused on feature extraction rather than similarity matching, so our SFM and MFM fusion methods are feasible.

Correlation analysis

For the ten branches of proposed GMBFF network, we hope that, firstly, each branch can extract highly discriminative global features. Secondly, the results of re-identification between each branch have low linear correlation, because only in the case that the re-identification ranking results of the different branches are very different. Also, the fused feature can be guaranteed to have a higher discrimination than those used for fusion.

Correlation analysis method

The method is to use the re-identification ranking results of each query image to analyze the correlation between ten branches, and then it averages the correlation of all the query images.

To be specific, let $Q = \left[q_1, q_2, \cdots, q_{N_1}\right]$ (N_1 is the number of vehicle images of the query set) which denotes the query set, for a query vehicle image q_i in query set Q we use the k-th branch to carry out the single branch re-identification experiment, in the gallery set $G = \left[g_1, g_2, \cdots, g_{N_2}\right]$ (N_2 is the number of vehicle images of the gallery set), the ranking vector of re-identification results is

$$P_i^k = \left[p_{i,1}^k, p_{i,2}^k, \cdots, p_{i,N_2}^k\right]^T$$ (5)

Where, $p_{i,j}^k$ means, for the query image q_i, after being re-identified by k-th branch, the position of the j-th image g_j in the gallery set G after being ranked from small to large by the Euclidean distance. Note

$$\overline{p_i^k} = \frac{p_{i,1}^k + p_{i,2}^k + \cdots + p_{i,N_2}^k}{N_2}$$ (6)

and

$$\overline{P_i^k} = \left[\overline{p_i^k}, \overline{p_i^k}, \cdots, \overline{p_i^k}\right]^T$$ (7)

Then, for query image q_i, the m-th branch and n-th branch are used to do the re-identification experiment in gallery set G, respectively. The correlation of the ranking results between the two branches is as follows:

$$r_i^{m,n} = \frac{\sum (P_i^m - \overline{P_i^m})(P_i^n - \overline{P_i^n})}{\sqrt{\sum (P_i^m - \overline{P_i^m})^2} \sqrt{\sum (P_i^n - \overline{P_i^n})^2}}$$

$$= \frac{\sum_{j=1}^{N_2} (p_{i,j}^m - \overline{p_{i,j}^m})(p_{i,j}^n - \overline{p_{i,j}^n})}{\sqrt{\sum_{j=1}^{N_2} (p_{i,j}^m - \overline{p_{i,j}^m})^2} \sqrt{\sum_{j=1}^{N_2} (p_{i,j}^n - \overline{p_{i,j}^n})^2}} \qquad (8)$$

For query image q_i, the correlation coefficient matrix between all ten branches of proposed GMBFF network is defined as:

$$corr_i = \begin{pmatrix} r_i^{1,1} & r_i^{1,2} & \cdots & r_i^{1,10} \\ r_i^{2,1} & r_i^{2,2} & \cdots & r_i^{2,10} \\ \cdots & \cdots & \cdots & \cdots \\ r_i^{10,1} & r_i^{10,2} & \cdots & r_i^{10,10} \end{pmatrix} \qquad (9)$$

Finally, for all query images $q_1, q_2, \cdots, q_{N_1}$ in query set Q, calculate the average value of their correlation coefficient matrix as follows:

$$\overline{corr} = \frac{\sum_{i=1}^{N_1} corr_i}{N_1} \qquad (10)$$

We call the matrix $\overline{corr}$ as the correlation coefficient matrix between all the branches of GMBFF network and also it is called as the correlation coefficient matrix between the features extracted by these ten branches.

Correlation verification

The correlation verification is carried out in two datasets. For the query images q_i of the VeRi-776 dataset, we divide the re-identification ranking result vector P_i^k (see formula 5) into positive vector $P_i^{k,pos}$ and negative vector $P_i^{k,neg}$. For a element of P_i^k, it represents the ranking result position value of a image in the gallery set, if this image has the same ID as the query image q_i, then we put it into the positive vector $P_i^{k,pos}$, otherwise we put it into the negative vector $P_i^{k,neg}$. Where

$$P_i^{k,pos} = \left[p_{i,1}^{k,pos}, p_{i,2}^{k,pos}, \cdots, p_{i,N_3}^{k,pos} \right]^T \qquad (11)$$

$$P_i^{k,neg} = \left[p_{i,1}^{k,neg}, p_{i,2}^{k,neg}, \cdots, p_{i,N_4}^{k,neg} \right]^T \qquad (12)$$

here, $N_3 + N_4 = N_2$. Then the correlation coefficient matrix $\overline{corr^{pos}}$ and $\overline{corr^{neg}}$ are calculated by formula 5-10. For the VehicleID dataset, there is only one vehicle image in the gallery set has the same vehicle ID with the query image, so only $\overline{corr}$ (without divide P_i^k into two vectors) can be obtained for this dataset. We remove all 1 in $\overline{corr^{pos}}$ and $\overline{corr^{neg}}$ calculated from the VeRi-776 dataset and all 1 in $\overline{corr}$ calculated from the VehicleID dataset, and draw the remaining data into three histograms, as shown in Figure 4.
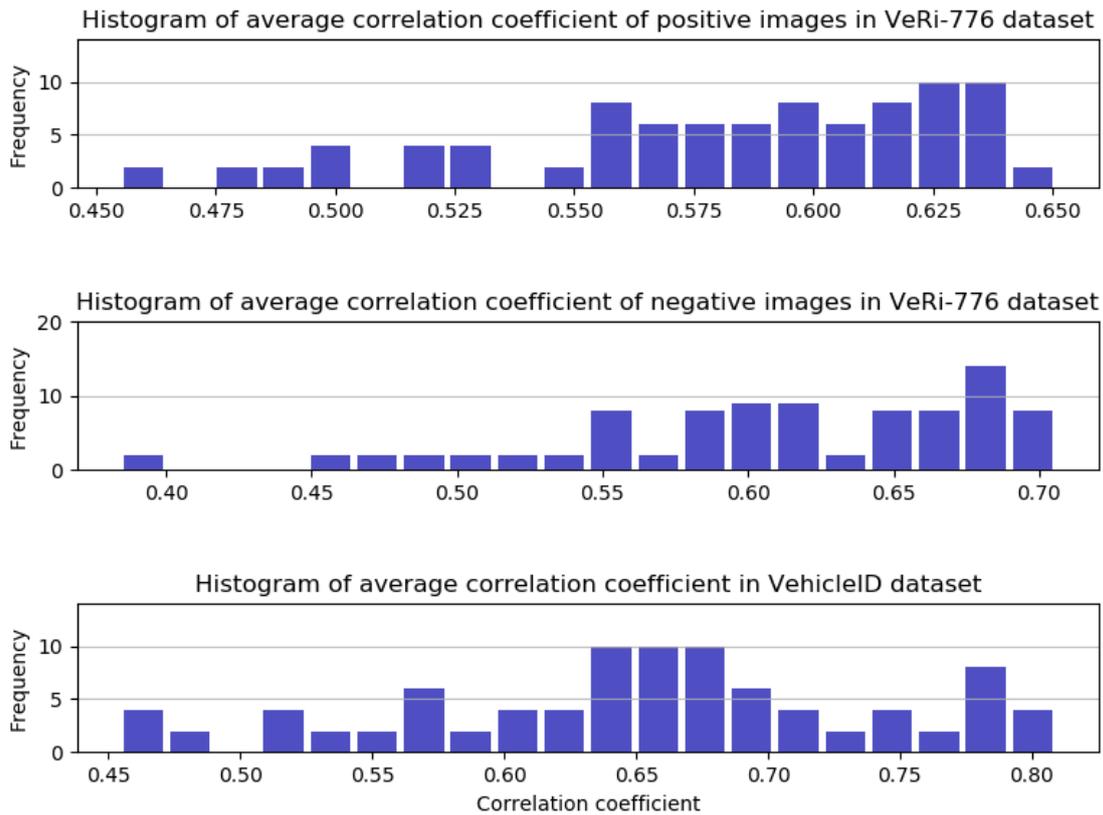
Figure 4. The histogram of average correlation coefficient between 10 branches. The average correlation coefficients between all branches are very small, which ensures that we can obtain higher discriminative fused features after the fusion of these branches.

As can be seen from Figure 4, in the VeRi-776 dataset, for the vehicle images with the same ID in the gallery set as the query image, the average correlation coefficients between all branches are mainly concentrated between 0.5 and 0.65, and for vehicle images in the gallery set with different image IDs from the query image, the average correlation coefficients between all branches are mainly concentrated between 0.55 and 0.7. For the VehicleID dataset, the correlation coefficients between all branches are mainly concentrated between 0.55 and 0.8, indicating that all the re-identification ranking results of those branches are very different (see Figure 2). If the features extracted by these branches are fused by a good fusion method then the re-identification accuracy can be improved significantly.

NUMERICAL RESULT AND DISCUSSION

Experimental tools

In this section, we first introduce two large datasets for vehicle re-identification and the evaluation protocol for the experimental results. Finally, the experimental details are presented of the proposed method.

Dataset

We use the VeRi-776 dataset and VehicleID dataset for experimental evaluation. These two datasets are widely used in vehicle re-identification.

The VeRi-776 dataset proposed by Liu and coauthors [26] is widely used in vehicle re-identification. The dataset consists of 51035 vehicle images, and the number of vehicle IDs is 776. Among them, the training set contains 37778 images, and the number of training vehicle IDs is 500. The rest is used as the test set. There are 13257 images of 200 vehicles in the test set, and the images of each vehicle are split into the query set which contains 1678 vehicle images and gallery set which contain 11579 vehicle images. Totally twenty cameras were used to complete this dataset, each image of this dataset is taken by 2-18 cameras from different perspectives. Therefore, when using a query image to match in the gallery set, the vehicle images with the same vehicle ID and camera ID as the query images are removed.

The VehicleID dataset proposed by Liu and coauthors [27] which is another widely used dataset for vehicle re-identification. This dataset has no camera ID information, and all pictures are taken from the front and rear of the vehicle. There are more vehicle IDs in this dataset than the VeRi-776 dataset. There are a total of 221,567 vehicle images of 26,328 vehicles. The number of vehicle IDs in the training set is 13,164, with a total of 113,346 vehicle images, and 108,221 vehicle images of another 13,164 vehicles form the test set. And the test set is split into several small test sets. In this paper, only three split test sets (small, medium, and large) are used. The number of vehicle IDs for the three split datasets are 800, 1600, and 2400, and the number of vehicle images is 6,493, 13,377, and 19777, respectively. Then randomly select an image from all the images of each vehicle in these three split datasets and put it into the corresponding gallery set, and the rest are used as the corresponding query set. The disadvantage of this dataset is that all the images are taken from the front and the rear, which makes that all the information of the vehicle cannot be extracted well. The advantage is that only one vehicle image is with the same vehicle ID in the gallery set as the query image, which is consistent with the actual traffic situation.

Evaluation protocol

In this paper, mean Average Precision (mAP) and Cumulative Match Characteristic (CMC) curve are used as evaluation metrics. CMC@k is defined as follows:

$$CMC @ k = \frac{\sum_{j=1}^{L} r(q_j, k)}{L} \tag{13}$$

where L is the total number of images in the query set. When there is a correct match in the top-k matches in the list, $r(q_i, k)$ is equal to 1, otherwise $r(q_i, k)$ is equal to 0. The average query precision is culculated as follows:

$$AP = \sum_{k=1}^{T} P(k) \Delta h(k) \tag{14}$$

where q is a query image, and $P(k)$ is the precision at a cutoff of k images, $\Delta h(k)$ is the change in recall that happened between cutoff $k-1$ and cutoff k, and T is the number of images in the gallery set. So mAP is defined as

$$mAP = \frac{\sum_{q \in Q} AP(q)}{L} \tag{15}$$

where L is the number of query images, and $Q = \{q_1, q_2, \cdots, q_L\}$ represents the query set. CMC@k indicates the probability of correct matching among the top-k matches in the list sorted by matching degree. We use mAP, CMC1 and CMC5 as evaluation protocol tools for VeRi-776 dataset. Each vehicle ID in the gallery set of VehicleID dataset, we only use CMC1 and CMC5 as the evaluation protocol.

Training configuration

Table 2. The learning rate adjustment epoch of each branch in the VeRi-776 dataset. The larger the dimension value is, the faster the learning speed will be.

Method	baseline	128	256	384	512	768	1024	1536	2048	3072	4096
Lr_step	10,25	5	5	3	3	2	2	2	2	1	1

For the input images of all branches, whether in the training phase or in the evaluation phase, there are only three transformations are used: (1) adjust the size to 224×224, (2) linearly adjust the pixel value to [0,1], and (3) normalize it with the mean value of [0.485, 0.456, 0.406] and standard deviation of [0.229, 0.224, 0.225], which are the same as what are in the Imagenet [28], then without any other transformations. All branches are initialized with the pre-trained model on Imagenet [28]. Using Adam parameter optimization

strategy, the batch size is given as 64. Because the learning abilities of the ten branches and baseline are different, we adopt different training strategies for them. For the VeRi-776 dataset, the initial learning rate of each branch is set to 0.0001, and different exponential adjustment strategies for learning rate are adopted. When the training epoch is greater than lr_step (each branch has different lr_step,see Table 2), the learning rate drops to 0.00001. In addition, for this dataset, we carry out an evaluation after every training epoch, and the total training epochs is 25. For the baseline about this dataset, the initial learning rate is set to 0.0001. When the training epoch reaches 10 and 25, the learning rate drops to 0.00001 and 0.000001 respectively, and the total training epoch is 40. For the VehicleID dataset, another two layers are added, the learning speed is becoming slow. All branches adopt the following learning rate adjustment scheme:

$$learning\,rate = \begin{cases} epoch \times 10^{-5}, & if \ 1 \le epoch \le 10 \\ 10^{-5}, & if \ 10 < epoch \le 20 \\ 10^{-6}, & if \ 20 < epoch \le 30 \end{cases} \tag{16}$$

Experiment

In this section, we first discuss the performance comparison between all branches and baseline of GMBFF network, and then conduct ablation experiments on our GMBFF network. The proposed method is compared with state-of-the-art methods. Finally, we do experiment on running time analysis.

Comparison between branches and baseline

In order to observe whether the branches we derive can extract highly discriminative features, we separately carry out the re-identification experiments on these branches and baseline (resnet50), so as to facilitate the contrast between all branches and baseline. From Table 3 and Table 4, we can see that all the ten branches of GMBFF network have a large improvement compared with baseline. For the VeRi-776 dataset, all branches are having the mAP from 6.19% to 11.92% which is higher than the baseline. CMC1 and CMC5 are also greatly improved. For the small split dataset of the VehicleID dataset, the CMC1 of all ten branches is more than 10% higher than the baseline, while the CMC5 is about 15% higher than the baseline. In the other two split datasets of the VehicleID dataset, the CMC1 is about 8% higher than the baseline, and the CMC5 is about 13% higher than the baseline.

Table 3. Comparison of performance between baseline and all branches in VeRi-776 dataset. All evaluate protocol of all branches of GMBFF is higher than that of baseline.

Method	mAP	CMC1	CMC5
baseline	0.5714	0.8719	0.9416
128	0.6333	0.8771	0.9452
256	0.6659	0.9036	0.9560
384	0.6839	0.9139	0.9560
512	0.6845	0.9151	0.9663
768	**0.6906**	0.9187	0.9663
1024	0.6807	0.9181	0.9675
1536	0.6896	**0.9241**	**0.9693**
2048	0.6771	0.9229	0.9681
3072	0.6545	0.9114	0.9639
4096	0.6528	0.9169	0.9590

As mentioned in the previous part of this paper, with the increase of dimension, the global feature contains more information. When the dimension is too large, the established model is not enough to learn so much information, but will reduce the re-identification accuracy. Table 3 and Table 4 verify this conclusion. From Table 3 and Table 4, we can see that the most suitable dimension of a vehicle feature extracted by a single branch model should be between 384 and 2048.

Table 4. Comparison of performance between baseline and all branches in VehecleID dataset. All evaluate protocol of all branches of GMBFF is higher than that of baseline.

Method	Small		Medium		Large	
	CMC1	CMC5	CMC1	CMC5	CMC1	CMC5
baseline	0.6555	0.7759	0.6369	0.7551	0.6072	0.7325
128	0.7516	0.9264	0.7090	0.8948	0.6718	0.8661
256	0.7597	0.9264	0.7173	0.8644	0.6783	0.8414
384	0.7641	0.9218	0.7110	0.8770	0.6776	0.8503
512	0.7597	0.9315	0.7121	0.8812	0.6858	0.8523
768	0.7655	0.9331	0.7134	0.8847	0.6802	0.8552
1024	0.7599	0.9361	0.7113	0.8915	0.6863	0.8540
1536	0.7667	0.9187	0.7193	0.8884	0.6942	**0.8593**
2048	**0.7671**	0.9280	**0.7226**	0.8835	**0.6921**	**0.8593**
3072	0.7577	**0.9417**	0.7135	**0.8915**	0.6793	0.8543
4096	0.7579	0.9329	0.7168	0.8823	0.6802	0.8509

Ablation Experiment

The way we do ablation experiments is to increase the number of branches one by one, and use the changes of evaluation protocols when the branches increase to measure the impact of the number of branches on our proposed GMBFF network. In order to achieve the purpose of flexible selection of the number of branches, we hope that the combination of the selected branches is always the optimal combination under current number of branches. So, we use the greedy algorithm to select the branches one by one. To be specifically, first we select an optimal branch from the ten branches. The selection method of the optimal branch is to average all re-identification evaluation protocols of each branch and select the branch with the largest average value. Then it will be continued to find another branch from the remaining nine branches that can form the optimal combination with the former selected branch, and so on, until all branches are used.

It can be seen from Table 5 and Table 6 that, with the increase of the number of branches, the mAP increases all the time. According to the meaning of mAP, some of the lower ranked vehicle images with the same ID as the query image are moving forward. The lower ranking of vehicle images is mainly due to interference factors such as camera distance, lighting, or occlusion and the network has not yet obtained sufficient information about these images. However, CMC1 and CMC5 increase first and then decrease (red values in Table 5 and Table 6 indicate a decline), which indicates that these top-ranked images have been extracted enough information because of too few interference factors. Because in actual urban traffic, the number of vehicles in a single camera is much less than the number of images in the gallery set of the two datasets used in this paper. It is easy to identify vehicles with few interference factors. However, for this situation, interference factors such as camera distance, occlusion or lighting will appear more frequently. Therefore, with the increase of the number of branches, and accompanish with the continuous increase of mAP, it shows that for the actual urban traffic, the more branches, the more benefit vehicle re-identification will be (if the timeliness factor is not considered).

Table 5. Experiments in which the number of branches increases one by one in VeRi-776 dataset (the red indicates decline). With the increase of the number of branches, the CMC1 and CMC5 increase first and then decrease, but the mAP increase all the way.

Method	mAP	CMC1	CMC5
1536	0.6896	0.9241	0.9693
+512	0.7272	0.9373	0.9741
+768	0.7408	0.9416	0.9753
+2048	0.7461	0.9476	0.9771
+128	0.7533	0.9470	0.9765
+1024	0.7551	0.9476	0.9783
+4096	0.7557	0.9476	0.9777
+256	0.7579	0.9470	0.9771
+384	0.7582	0.7458	0.9777
+3072	0.7594	0.9458	0.9777

Table 6. Experiments in which the number of branches increases one by one in VehecleID dataset (the red indicates decline). With the increase of the number of branches, the CMC1 and CMC5 increase first and then decrease.

Method	Small		Medium		Large	
	CMC1	CMC5	CMC1	CMC5	CMC1	CMC5
2048	0.7671	0.9280	0.7226	0.8835	0.6921	0.8593
+128	0.7799	0.9410	0.7331	0.9072	0.7039	0.8745
+768	0.7847	0.9504	0.7423	0.9071	0.7046	0.8756
+512	0.7875	0.9521	0.7425	0.9072	0.7084	0.8750
+256	0.7870	0.9521	0.7448	0.9076	0.7090	0.8764
+384	0.7870	0.9535	0.7445	0.9081	0.7086	0.8772
+1024	0.7861	0.9525	0.7456	0.9079	0.7096	0.8759
+1536	0.7866	0.9521	0.7463	0.9098	0.7117	0.8758
+3072	0.7864	0.9537	0.7476	0.9089	0.7103	0.8741
+4096	0.7856	0.9532	0.7475	0.9094	0.7096	0.8733

Comparison with state-of-the-art methods

Table 7. GMBFF method compared with state-of-the-art methods in VeRi-776 dataset.

Method	mAP	CMC1	CMC5
RAM [13]	0.6150	0.8860	0.9400
VAMI+STR [29]	0.6132	0.8592	0.9184
AAVER [11]	0.6118	0.8897	0.9470
QD-DLF [6]	0.6183	0.8850	0.9446
MTML-OSG [24]	0.6460	0.9230	0.9570
VANet [34]	0.6634	0.8978	0.9599
HSS-GCN [32]	0.4480	0.6440	0.8610
MV-GAN [32]	0.6316	0.9106	0.9577
GMBFF+four brahches(ours)	0.7463	0.9404	0.9738
GMBFF+SFM (ours)	0.7474	**0.9440**	0.9729
GMBFF+MFM (ours)	**0.7599**	0.9434	**0.9771**

Comparison on VeRi-776 dataset

In the VeRi-776 dataset, we compare the proposed GMBFF method with RAM [13], VAMI + STR [29], MTML-OSG [24], VANet [34], AAVER [11], QD-DLF [6], HSS-GCN [32] and MV-GAN [32] methods. VAMI + STR is used to infer multi-view features from single-view input, and spatio-temporal information is added. VANet is a metric learning method based on similar viewpoint and different viewpoint learning. AAVER joins the attention model, combining local features and global features. RAM fuses attribute features, global features, global features with BN layer and local features. MTML-OSG is a fusion of different features extracted from the ID branch, multi-scale analysis branch, gray analysis branch and vehicle orientation branch. QD-DLF uses four branches to extract features from the four orientations of the vehicle, and finally fuses these features. HSS-GCN proposed a hierarchical structure graph convolutional network, the framework was composed of two branches, respectively through the global module and GCN module to extract global representation and structural features. MV-GAN proposed a multi-view generative adversarial network, using two feature extraction networks to extract features from the original query image and the generated multi-view images. Then all the features are fused into a global feature. It can be seen from Table 7 that our proposed GMBFF method far exceed these recent methods. RAM, MTML-OSG, and QD-DLF fuse the features extracted by four branches. When our GMBFF uses only four branches, for mAP, GMBFF is 13% higher than RAM, 12.8% higher than QD-DLF, and 10% higher than MTML-OSG. Both CMC1 and CMC5 are greatly improved. It can also be seen in Table 7 that the proposed MFM fusion method has a certain improvement in experimental accuracy compared with the SFM fusion method.

Comparison on VehicleID dataset

Because this dataset is not labeled with spatio-temporal information, VAMI [29] removed the spatio-temporal information. FDA-Net [30] used the generative adversarial network to online generate hard negative samples from both visual appearance and feature distance. TAMR [31] is a two-level attention network metric learning method. It can be seen from Table 8 that our GMBFF network greatly exceeds these state-of-the-

art algorithms in this dataset. Similarly, when only four branches are used, our GMBFF network has a big improvement over RAM and QD-DLF. It can also be seen in Table 8 that the MFM fusion method has a certain improvement in experimental accuracy over the SFM method.

Table 8. GMBFF method compared with state-of-the-art methods in VehicleID dataset.

Method	Small		Medium		Large	
	CMC1	CMC5	CMC1	CMC5	CMC1	CMC5
VAMI [29]	0.6312	0.8325	0.5287	0.7512	0.4734	0.7029
RAM [13]	0.7520	0.9150	0.7230	0.8700	0.6700	0.8450
FDA-Net [30]	0.6403	0.8280	0.5782	0.7834	0.4943	0.7048
TAMR [31]	0.6602	0.7971	0.6290	0.7680	0.5969	0.7387
QD-DLF [6]	0.7232	0.9248	0.7066	0.8890	0.6414	0.8337
AAVER [11]	0.7469	0.9382	0.6862	0.8995	0.6354	0.8564
HSS-GCN [32]	0.7270	0.9180	0.6790	0.8780	0.6240	0.8430
MV-GAN [32]	0.7453	0.8986	0.6794	0.8165	0.6332	0.7762
GMBFF+four brahches(ours)	**0.7859**	0.9426	0.7422	0.9058	**0.7113**	**0.8755**
GMBFF+SFM(ours)	0.7789	0.9519	0.7359	0.9010	0.7044	0.8700
GMBFF+MFM(ours)	0.7856	**0.9532**	**0.7475**	**0.9094**	0.7103	0.8741

Running time analysis

As discussed in the previous section 3.2 and [6], in the re-identification task, the feature extraction time (FET, milliseconds / image) of each image is mainly used to compare the efficiency of the re-identification algorithm. Here we are mainly compared with the three vehicle re-identification methods of RAM, MTML-OSG and QD-DLF, which all use four branches. All methods are performed on the configurations listed in Section 4.2.

Table 9. FET of single branch of GMBFF network. With the increase of feature dimension, the FET almost the same.

Branch	128	256	384	512	768	1024	1536	2048	3072	4096
FET	2.776	2.614	2.651	2.648	2.619	2.650	2.646	2.669	2.658	2.686

Table 10. FET of the combination of different numbers of branches of GMBFF network and the FET comparison with RAM, MTML-SOG and QD-DLF. FET increased linearly with the increase of branch number. The FET of GMBFF is only a little more than that of RAM, MTML-SOG and QD-DLF..

Method	FET
GMBFF+two branches	6.410
GMBFF+three branches	9.456
GMBFF+four branches	12.627
GMBFF+five branches	15.806
GMBFF+six branches	19.395
GMBFF+seven branches	22.394
GMBFF+eight branches	25.739
GMBFF+nine branches	29.164
GMBFF+ten branches	32.744
RAM [13]	10.359
MTML-OSG [24]	12.512
QD-DLF [6]	11.119

As can be seen from Table 9, the efficiency of each branch of the proposed GMBFF is almost the same, and the different dimensions of the extracted features has little impact on efficiency. As can be seen from Table 10, for our GMBFF network, as the number of branches increases, the time to extract features of an image increases linearly. When there are only four branches for the GMBFF network, the FET is only slightly higher than the four-branch models of RAM and MTML-OSG, but much less than QD-DLF. The reason is that the four branches of the RAM network have some shared convolutional layers, while all branches of the

GMBFF network are independent of each other. The input image size of one branch of MTML-OSG is 160×160, and for the rest three branch is 224×224, nevertheless, the input image size of all branches of the GMBFF network is 224×224, which makes the FET of MTML-OSG a slightly less than that of ours GMBFF. The input size of the four branches of QD-DLF are all 128×128, which makes the FET of QD-DLF smallest. Through the analysis of running time shows that our GMBFF method is feasible.

DISCUSSION

We use experiments to verify that every branch of the proposed method is much better than the baseline, and by fusing these branches, we get a very highly discriminative features. We analized the correlation between each each pair of branches, the result is that the correlation between each each pair of branches is very low, which is the reason that the fused feature can be highly discriminative. The disadvantage of the proposed method is that it is a bit time-consuming by extracting features of ten branches. But it is flexibly for us to choose some branches from the ten branches according to the traffic congestion.

CONCLUSION

In this paper, we propose a method of generated multi branch feature fusion for vehicle re-identification named GMBFF. By the GMBFF, we achieve the purpose of using global features instead of local features to obtain more fine-grained distinguishing information. GMBFF is composed by several branches, and each branch can extract highly discriminative global features and is derived from a same network. We experimentally verified that there is a low linear correlation between these branches, which is a prerequisite for these branches to effectively fuse. The experimental results show that the MFM method is better than the SFM method in re-identification accuracy. Finally, experiments show that our proposed GMBFF method is superior to state-of-the-art vehicle re-identification methods.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

1. Mao QC, Sun HM, Zuo LQ, et al. Finding every car: a traffic surveillance multi scale vehicle object detection method. Appl. Intell. 2020; 50(10): 3125-36. https://doi.org/10.1007/s10489-020-01704-5
2. Tao, H. Detecting smoky vehicles from traffic surveillance videos based on dynamic features. Appl. Intell. 2020; 50(4), 1057–72. https://doi.org/10.1007/s10489-019-01589-z
3. Fazlollahtabar H, Hassanli S. Hybrid cost and time path planning for multiple autonomous guided vehicles. Appl. Intell. 2018; 48, 482–98. https://doi.org/10.1007/s10489-017-0997-x
4. Kala, R., Warwick, K. Dynamic distributed lanes: motion planning for multiple autonomous vehicles. Appl. Intell. 2014;41(1), 260–81. https://doi.org/10.1007/s10489- 014-0517-1
5. Liu W, Liu X, Ma H, et al. Beyond human-level license plate super resolution with progressive vehicle search and domain priori GAN. Proceedings of the 25th ACM international conference on Multimedia. 2017; pp. 1618-26. https://doi.org/10.1145/3123266.3123422
6. Zhu J, Zeng H, Huang J, et al. Vehicle re-identification using quadruple directional deep learning features. IEEE Transactions on Intelligent Transportation Systems. 2019; 21(1): 410-20. https://doi.org/10.1109/TITS.2019.2901312
7. Tang Y, Wu D, Jin Z, et al. Multi-modal metric learning for vehicle re-identification in traffic surveillance environment. 2017 IEEE International Conference on Image Processing (ICIP). 2017 pp. 2254-2258. https://doi.org/10.1109/ICIP.2017.8296683
8. Zhou Y, Shao L. Cross-View GAN Based Vehicle Generation for Re-identification. BMVC. 2017; pp. 1: 1-12. https://doi.org/10.5244/C.31.186
9. Wu CW, Liu CT, Chiang CE, et al. Vehicle re-identification with the space-time prior. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. 2018; pp. 121-8. https://doi.org/10.1109/CVPRW.2018.00024

10. Wu F, Yan S, Smith JS, et al. Joint semi-supervised learning and re-ranking for vehicle re-identification. 2018 24th International   Conference   on   Pattern   Recognition   (ICPR).   2018;   pp.   278-83. https://doi.org/10.1109/ICPR.2018.8545584

11. Khorramshahi P, Kumar A, Peri N, et al. A dual-path model with adaptive attention for vehicle re-identification. Proceedings   of   the   IEEE/CVF   International   Conference   on   Computer   Vision.   2020;   6132-41. https://doi.org/10.1109/ICCV.2019.00623

12. He B, Li J, Zhao Y, et al. Part-regularized near-duplicate vehicle re-identification. Proceedings of the IEEE Conference   on   Computer   Vision   and   Pattern   Recognition.   2019;   pp.   3997-4005. https://doi.org/10.1109/CVPR.2019.00412

13. Liu X, Zhang S, Huang Q, et al. Ram: a region-aware deep model for vehicle re-identification. 2018 IEEE International   Conference   on   Multimedia   and   Expo   (ICME).   2018;   pp.   1-6. https://doi.org/10.1109/ICME.2018.8486589

14. Wang ZD, Tang LM, Liu XH, et al. Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification. In Pro ceedings of the IEEE International Conference on Computer Vision. 2017; pp. 379–387. https://doi.org/10.1109/ICCV.2017.49

15. Wang H, Peng J, Jiang G, et al. Discriminative feature and dictionary learning with part-aware model for vehicle re-identification. Neurocomputing. 2021; 438: 55-62. https://doi.org/10.1016/j.neucom.2020.06.148

16. Zheng B, Lei Z, Tang C, et al. OERFF: A Vehicle Re-Identification Method Based on Orientation Estimation and Regional Feature Fusion. IEEE Access. 2021; 9: 66661-74. https://doi.org/10.1109/ACCESS.2021.3076054

17. Wang Q, Min W, Han Q, et al. Viewpoint adaptation learning with cross-view distance metric for robust vehicle re-identification. Information Sciences. 2021 564: 71-84. https://doi.org/10.1016/j.ins.2021.02.013

18. Zeiler MD, Fergus R. Visualizing and understanding convolutional net works. In European conference on computer vision. 2014;8689:818-833. https://doi.org/10.1007/978-3-319-10590-153

19. Hohman F, Kahng M, Pienta R, et al. Visual analytics in deep learning: An interrogative survey for the next frontiers. IEEE   transactions   on   visualization   and   computer   graphics.   2018;   25(8):   2674-2693. https://doi.org/10.1109/TVCG.2018.2843369

20. Sun Y, Xu Q, Li Y, et al. Perceive Where to Focus: Learning Visibility aware Part-level Features for Partial Person Re-identification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2019; pp. 393-402. https://doi.org/10.1109/CVPR.2019.00048

21. Zhang Z, Lan C, Zeng W, et al. Densely Semantically Aligned Person Re-Identification. Proceedings of the IEEE Conference   on   Computer   Vision   and   Pattern   Recognition.   2019;   pp.   667-676. https://doi.org/10.1109/CVPR.2019.00076

22. Tay CP, Roy S, Yap KH. AANet: Attribute Attention Network for Person Re Identifications. Proceedings of the IEEE Conference   on   Computer   Vision   and   Pattern   Recognition.   2019;   pp.   7134-7143. https://doi.org/10.1109/CVPR.2019.00730

23. Yin J, Fan Z, Chen S, et al. In-depth exploration of attribute information for person re-identification. Appl. Intell. 2020; 50(11): 3607-22. https://doi.org/10.1007/s10489-020-01752-x

24. Kanaci A, Li M, Gong S, et al. Multi-Task Mutual Learning for Vehicle Re-Identification. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Work shops. 2019; pp. 62-70.

25. Lou Y, Bai Y, Liu J, et al. Embedding adversarial learning for vehicle re-identification. IEEE Transactions on Image Processing. 2019; 28(8): 3794-3807. https://doi.org/10.1109/TIP.2019.2902112

26. Liu X, Liu W, Ma H, et al. Large-scale vehicle re-identification in urban surveillance videos. 2016 IEEE International Conference on Multimedia and Expo (ICME). 2016; pp. 1-6. https://doi.org/10.1109/ICME.2016.7553002

27. Liu HY, Tian YH, Yang YW, et al. Deep relative distance learning: Tell the difference between similar vehicles. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. 2016; pp. 2167–2175. https://doi.org/10.1109/CVPR.2016.238

28. Deng J, Dong W, Socher R, et al. Imagenet: A large-scale hierarchical image database. 2009 IEEE conference on computer vision and pattern recognition. 2009; pp. 248-255. https://doi.org/10.1109/CVPR.2009.5206848

29. Zhou Y, Shao L. Aware attentive multi-view inference for vehicle re-identification. Proceedings of the IEEE Conference   on   Computer   Vision   and   Pattern   Recognition.   2018;   pp.   6489-6498. https://doi.org/10.1109/CVPR.2018.00679

30. Lou Y, Bai Y, Liu J, et al. Veri-wild: A large dataset and a new method for vehicle re-identification in the wild. in Proc. IEEE Conf. Comp. Vis. Patt. Recogn. 2019; pp. 3235–3243. https://doi.org/10.1109/CVPR.2019.00335

31. Guo H, Zhu K, Tang M, et al. Two-level attention network with multi-grain ranking loss for vehicle re-identification. IEEE Transactions on Image Processing. 2019; 28(9): 4328-4338. https://doi.org/10.1109/TIP.2019.2910408

32. Zhang F, Ma Y, Yuan G, et al. Multiview image generation for vehicle reidentification. Appl. Intell. 2021; pp. 1-18. https://doi.org/10.1007/s10489-020-02171-8

33. Chu R, Sun Y, Li Y, et al. Vehicle Re-identification with Viewpoint-aware Metric Learn ing. Proceedings of the IEEE International Conference on Computer Vision. 2019; pp. 8282-8291. https://doi.org/10.1109/ICCV.2019.00837

34. Xu Z, Wei L, Lang C, et al. HSS-GCN: A Hierarchical Spatial Structural Graph Convolutional Network for Vehicle Re-identification. Proc. ICPR's Int. Workshop on Human and Vehicle Analysis for Intelligent Urban Computing (IUC). 2021; 356-364. https://doi.org/10.1007/978-3-030-68821-9 32.