ISSN 2616-5775 Vol. 6, Issue 11: 46-51, DOI: 10.25236/AJCIS.2023.061107

Multi-branch mutual learning net for vehicle re-identification

Xin Bai*, Ming Gao, Xin Zheng

School of Information Science and Engineering, Chongqing Jiaotong University, Chongqing, China
Corresponding author

Abstract: *To address the problem that the global feature extraction ability of single-branch network is low and does not have the sensitivity of specific scenes due to the interference of factors such as occlusion caused by fixed camera shooting and scale scaling caused by vehicle driving from far and near in vehicle re-identification, ML Net (Mutual Learning Net) is proposed. Firstly, the multi-branch structure of the network consists of one master branch and three slave branches, the master branch is responsible for the feature learning task in general complex scenes, and the three slave branches are responsible for the feature learning task under image occlusion, scale scaling, and vehicle color change, so that the network can be sensitive to specific scenes. Finally, the models are trained in each branching stage jointly with Label Smoothing Cross Entropy Loss, Triplet Loss and KL (Kullback-Leibler Divergence) Loss. The experimental results show that the proposed ML Net network achieves advanced experimental results on two publicly available datasets, VeRi776 and VehicleID.*

Keywords: *vehicle re-identification; multi-branch; mutual learning*

1. Introduction

In recent years, with the development of computer deep visual technology, vehicle re-identification technology has been increasingly applied in fields such as assisting criminal investigation, intelligent transportation, and cross-border tracking. Vehicle re-identification refers to finding a given query vehicle in a gallery image.

Previous work on vehicle re-identification mainly relied on global feature-based single-branch networks. Liu et al. [1] utilized multimodal data for coarse-to-fine global feature domain search and proposed the progressive vehicle re-identification framework (PROVID). Zheng et al. [2] treated the re-identification task as a multi-classification task, using deep neural networks to learn complex global image features and proposed the classic IDE (ID-discriminative Embedding) model. Wang et al. [3] enhanced the representation ability of the re-identification network by extracting multi-scale features in a single-branch network. He et al. [4] proposed using global features for vehicle identity mining and adopted a multi-domain learning approach to improve the retrieval accuracy of vehicle re-identification. Liao et al. [5] from JD AI Research Institute integrated a highly modular and scalable FastReID model only using global features and various tuning strategies to provide researchers with a faster and more convenient toolbox. Huynh et al. [6] improved the global representation ability of vehicle re-identification through multi-head attention mechanism and domain-difference generalization methods in the fifth AI City Challenge. Qian et al. [7] designed a Transformer-based unstructured feature decoupling vehicle re-identification method to eliminate the impact of unaligned clusters of abnormal features in semantics caused by a single-branch network.

However, single-branch networks based on global features lack the discriminative ability for detailed features, and some researchers have started to study single-branch vehicle re-identification networks based on local features. Meng et al. [8] used the U-Net network to extract vehicle local perspective information and embedded it into a single-branch network, proposing the Parsing-based View-aware Embedding Network (PVEN). Jiang et al. [9] respectively used an object detection network and a pose perception network to extract vehicle part features and pose features, and introduced them as prior information into a single-branch network, achieving advanced experimental results. However, single-branch networks based on local features require additional annotations such as keypoint labels, viewpoint labels, and vehicle part labels. Meanwhile, multi-branch networks, including online knowledge distillation and deep mutual learning methods [12], have been widely applied in image classification, object detection, and other fields.

ISSN 2616-5775 Vol. 6, Issue 11: 46-51, DOI: 10.25236/AJCIS.2023.061107

Therefore, how to utilize multi-branch networks to solve difficult tasks in re-identification such as occlusion and illumination variation has become a research focus. Portello et al. [10] proposed a multi-branch network based on view knowledge distillation (VKD), using the visual diversity information of teacher networks to supervise the online learning of student networks to improve the retrieval accuracy of re-identification tasks. Zhao et al. [11] proposed a multi-branch re-identification network based on saliency-guided asymmetric mutual hashing, using saliency maps generated by teacher networks to guide student networks to learn high-quality hash codes. Zang et al. [13] utilized scene discriminative features to enhance the representation ability of multi-branch re-identification networks.

Based on this, this paper proposes a Mutual Learning (ML Net), which consists of four branches: global branch, occlusion branch, scale variation branch, and color transformation branch. These four branches mutually learn knowledge. Finally, the features of the four branches are concatenated to evaluate the performance of the network.

2. ML Net

The ML Net network structure proposed in this paper is shown in Figure 1. Firstly, the global branch, occlusion branch, scale branch, and color transformation branch of the network handle data from different tasks separately. Then, the model optimization is performed using the combination of identity loss, triplet loss, and KL loss. Finally, the features from the four branches are concatenated and used for testing.

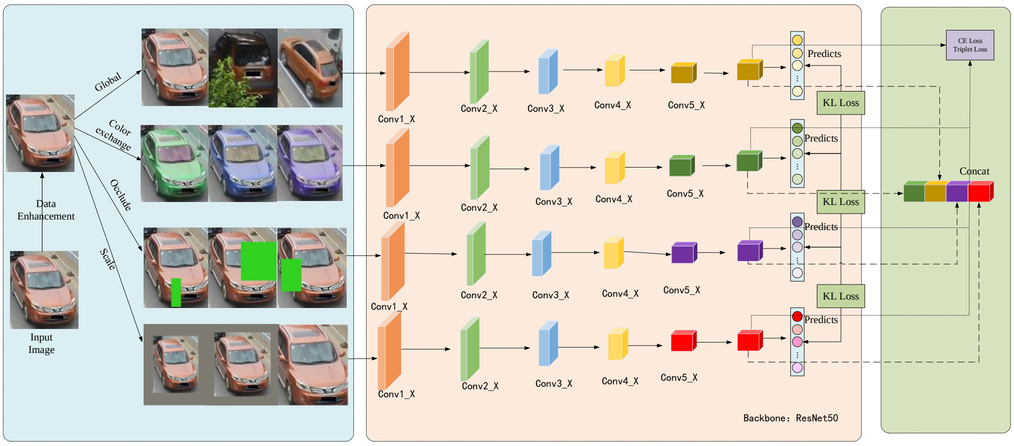
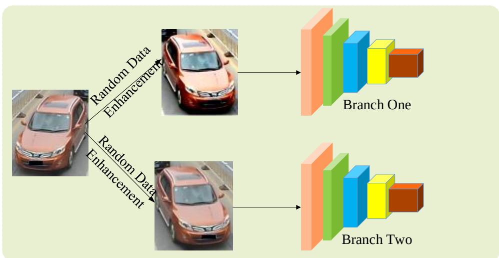


Figure 1: ML Net network structure diagram

2.1 Mutual Learning Net

2.1.1 Image alignment and primary branch network



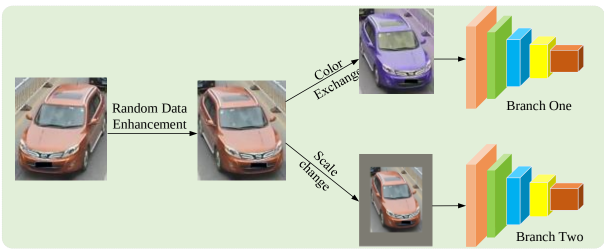


Figure 2: Heterogenous input *Figure 3: Homologous input*

In response to the common issue of image misalignment in multi-branch networks, where various data augmentation techniques such as random flip and random lighting enhancement can lead to misaligned data being received by different branch networks, as shown in Figure 2. This can result in image misalignment between different branches, thereby affecting the performance testing of the final network.

Therefore, it is proposed to first apply various data augmentation techniques in the multi-branch

ISSN 2616-5775 Vol. 6, Issue 11: 46-51, DOI: 10.25236/AJCIS.2023.061107

network and then feed the augmented data with the same distribution into the multi-branch network, as shown in Figure 3, in order to reduce the instability in performance testing caused by image misalignment.

In the global branch, serving as the primary branch network, it can provide foundational knowledge for complex scenes to the sub-branch networks. At the same time, it can also learn different knowledge from the sub-branches, thus improving the robustness of the primary branch network.

2.1.2 Slave branch network

The sub-branch networks designed in this paper include the occlusion branch, scale branch, and color transformation branch.

Figure 4: Masking branch

In the occlusion branch, the random erasing method is applied to the dataset. Unlike the conventional random erasing method, this paper sets the random erasing probability to 100% to ensure that each image in this branch has occlusion interference, as shown in Figure 4.

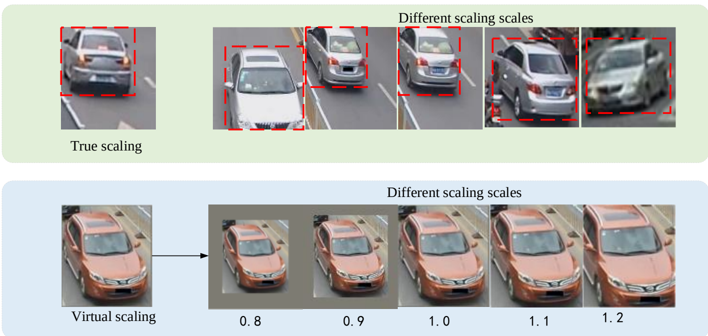


Figure 5: Scale scaling branch

In the scale branch, the average values of the RGB channels of all images are calculated separately, generating an image template. Then, the vehicle images are randomly scaled to 0.8 to 1.2 times of their original size. If the scaling value falls between the threshold of 0.8 and the threshold of 1.0, the scaled image is pasted at a random position on the image template. If the scaling value is larger than the threshold of 1.0, the scaled image is also pasted at the center of the generated image template, and the edge portions that exceed the template's boundaries are discarded. The probability of random scaling is also set to 100% to ensure that each image in this branch undergoes scale transformation, as shown in Figure 5.

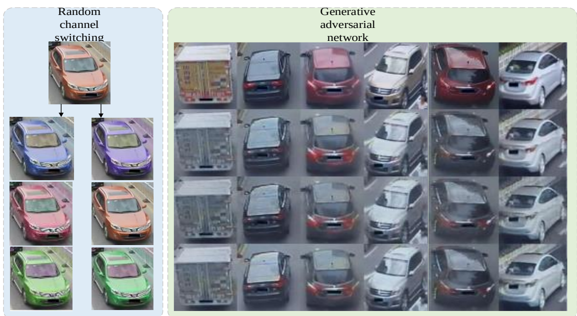


Figure 6: Color change branch

ISSN 2616-5775 Vol. 6, Issue 11: 46-51, DOI: 10.25236/AJCIS.2023.061107

In the color transformation branch, this paper randomly swaps the RGB channels' order for all images, setting the probability of random channel swapping to 100%. Considering that the color and texture information of vehicles are not strongly correlated with the edge structural information such as windows, hoods, tires, and trunks, the random channel swapping algorithm only changes the color while preserving the edge structural information. This means that by randomly swapping the RGB channels of a vehicle image, we can obtain six color variants, including the original image, without the need for time-consuming training using a Generative Adversarial Network (GAN) to alter the vehicle color. Therefore, the image random channel swapping algorithm designed in this paper is more time-efficient. As shown in Figure 6, the left section displays vehicle images generated by random channel swapping, while the right section shows vehicle images generated using the DG Net [14] adversarial network.

2.2 Joint loss function

In this paper, we employ the weighted soft-margin hard triplet loss function as the vehicle metric loss for each branch to optimize the feature distance between vehicle classes. Furthermore, the label-smoothed cross-entropy loss is employed as the vehicle identity loss for each branch to classify vehicle samples.

The respective advantages of joint metric loss and identity loss are optimized for each branch model, with λ representing the loss balancing parameter,

$$L_{part}^{\Theta} = \lambda L_{triplet}^{\Theta} + L_{cross}^{\Theta} \tag{1}$$

Where S represents the branch number. Thus, the KL loss calculation formula for each branch is as follows:

$$L_{totalkl}^{\Theta} = L_{kl}^{general} + L_{kl}^{occlude} + L_{kl}^{scale} + L_{kl}^{color} \tag{2}$$

In the end, the total loss function calculation formula for each branch model is as follows:

$$L_{total}^{\Theta} = L_{kl}^{\Theta} + L_{part}^{\Theta} \tag{3}$$

3. Experimental verification

3.1 Experimental Details

We utilized the deep learning framework PyTorch, Python 3.8, and a 3090 graphics card for model training.

3.2 Ablation Experiment

To validate the effectiveness of the proposed model, extensive ablation experiments were conducted on the VeRi776 dataset, targeting different modules, in order to explore the impact of each module on the experimental results.

3.2.1 ML Net

Table 1: Ablation experiments of ML Net (Percentage:%)

Baseline	Occluded branch	Scaling branch	Color branch	mAP	Rank-1
√				73.7	93.5
√	√			74.4	93.9
√	√	√		76.0	94.3
√	√	√	√	**76.9**	**94.8**

As shown in Table 1, ML Net (Mutual Learning Network) represents the mutual learning network. On the VeRi776 dataset, compared to the Baseline method, after adding the occlusion branch, the mAP value and Rank-1 value are improved by 0.7% and 0.3% respectively.

ISSN 2616-5775 Vol. 6, Issue 11: 46-51, DOI: 10.25236/AJCIS.2023.061107

3.3 Algorithm Comparison

To validate the effectiveness of the proposed model, Table 2 presents a comparison between MLAM Net and other recent methods on the VeRi776 dataset, including RAM, AAVER, VANet, PNVR, and PVEN. Compared to other methods, ML Net achieves superior experimental results on the VeRi776 dataset, with mAP and Rank-1 values reaching 79.6% and 95.9% respectively.

Table 2: Performance comparison of VeRi776 dataset (Percentage:%)

Method	mAP	Rank1
RAM[15]	61.5	88.6
AAVER[16]	66.3	90.1
VANet[17]	66.3	89.7
GSMC[18]	77.6	96.3
PNVR[19]	74.3	94.3
PVEN[8]	79.5	95.6
ML Net	79.6	95.9

Table 3 presents the comparison results between ML Net and other recent methods on the VehicleID dataset.

Table 3: Performance comparison of the Vehicle ID dataset (Percentage:%)

Method	Small		Medium		Large	
	Rank1	Rank5	Rank1	Rank5	Rank1	Rank5
RAM[15]	75.2	91.5	72.3	87.0	67.7	84.5
AAVER[16]	74.6	93.8	68.6	89.9	63.5	85.6
VANet[17]	88.1	97.2	83.1	95.1	80.3	92.9
GSMC[18]	83.5	96.7	79.1	94.0	75.0	90.9
PNVR[19]	78.4	92.3	75.0	88.3	74.2	86.4
ML Net	**79.4**	**93.3**	**77.6**	**91.5**	**73.1**	**86.1**

4. Conclusion

In this paper, we propose a multi-branch re-identification network with mutual learning (ML Net). Experimental results demonstrate that the proposed MLAM Net achieves mAP and Rank-1 values of 79.6% and 95.9%, respectively, on the VeRi776 dataset. On the Vehicle ID dataset, the Rank-1 and Rank-5 values on the Small test set are 79.4% and 93.3%, respectively, on the Medium test set are 77.6% and 91.5%, respectively, and on the Large test set are 73.1% and 86.1%, respectively, which outperform most existing methods.

References

[1] LIU X, LIU W, MEI T, et al. Provid: Progressive and multimodal vehicle reidentification for large-scale urban surveillance [J]. IEEE T Multimedia, 2017, 20(3): 645-658.

[2] ZHENG L, ZHANG H, SUN S, et al. Person re-identification in the wild [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017:1367-1376.

[3] WANG Y, WANG L, YOU Y, et al. Resource aware person re-identification across multiple resolutions [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 8042-8051.

[4] HE S, LUO H, CHEN W, et al. Multi-domain learning and identity mining for vehicle re-identification [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, 2020: 582-583.

[5] HE L, LIAO X, LIU W, et al. Fastreid: A pytorch toolbox for general instance re-identification [J]. arXiv Preprint arXiv:2006.02631, 2020.

[6] HUYNH, NGUYEN, et al. A strong baseline for vehicle re-identification [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021: 4147-4154.

[7] QIAN W, LUO H, PENG S, et al. Unstructured Feature Decoupling for Vehicle Re-identification [C] // Proceedings of the European Conference on Computer Vision, 2022: 336-353.

[8] MENG D, LI L, LIU X, et al. Parsing-based view-aware embedding network for vehicle re-identification [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern

ISSN 2616-5775 Vol. 6, Issue 11: 46-51, DOI: 10.25236/AJCIS.2023.061107

Recognition, 2020: 7103-7112.

[9] JIANG M, ZHANG X, YU Y, et al. Robust vehicle re-identification via rigid structure prior [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2021: 4026-4033.

[10] PORRELLO A, BERGAMINI L, CALDERARA S. Robust re-identification by multiple views knowledge distillation [C] // Proceedings of the European Conference on Computer Vision, 2020: 93-110.

[11] ZHAO C, TU Y, LAI Z, et al. Salience-guided iterative asymmetric mutual hashing for fast person re-identification [J]. IEEE Transactions on Image Processing, 2021: 7776-7789.

[12] ZHANG Y, XIANG T, HOSPEDALES T M, et al. Deep mutual learning [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2018: 4320-4328.

[13] ZANG X, LI G, GAO W, et al. Learning to disentangle scenes for person re-identification [J]. Image and Vision Computing, 2021, 116: 104330.

[14] ZHENG Z, YANG X, YU Z, et al. Joint discriminative and generative learning for person re-identification [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 2138-2147.

[15] LIU X, ZHANG S, HUANG Q, et al. Ram: a region-aware deep model for vehicle re-identification [C] // Proceedings of the IEEE International Conference on Multimedia and Expo, 2018:1-6.

[16] KHORRAMSHAHI P, KUMAR A, PERI N, et al. A dual-path model with adaptive attention for vehicle re-identification [C] // Proceedings of the IEEE International Conference on Computer Vision, 2019: 6132-6141.

[17] CHU R, SUN Y, LI Y, et al. Vehicle re-identification with viewpoint-aware metric learning [C] // Proceedings of the IEEE International Conference on Computer Vision, 2019: 8282-8291.

[18] Wang Zhenxue, Xu Zheming, Xue Yangyang, et al. Global and spatial multi-scale contexts fusion for vehicle re-identification [J]. Journal of Image and Graphics, 2023, 28(02): 471-82.

[19] HE B, LI J, ZHAO Y, et al. Part-regularized near-duplicate vehicle re-identification [C] // Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2019: 3997-4005.