2022 IEEE International Conference on Image Processing (ICIP) | 978-1-6654-9620-9/22/$31.00 ©2022 IEEE | DOI: 10.1109/ICIP46576.2022.9897541

PARTITION AND REUNION: A VIEWPOINT-AWARE LOSS FOR VEHICLE RE-IDENTIFICATION

Haobo Chen[1] *Yang Liu*[1] *Yang Huang*[1] *Wei Ke*[3] *Hao Sheng*[1,2,3]

[1] State Key Laboratory of Virtual Reality Technology and Systems,
School of Computer Science and Engineering, Beihang University, Beijing 100191, P.R.China
[2] Beihang Hangzhou Innovation Institute Yuhang,
Xixi Octagon City, Yuhang District, Hangzhou 310023, P.R.China
[3] Faculty of Applied Sciences, Macao Polytechnic University, Macao SAR 999078, P.R.China

ABSTRACT

Vehicle Re-Identification (ReID) aims to retrieve images of vehicles with the same identity from different scenarios. It is a challenging task due to the large intra-identity discrepancy caused by viewpoint variations and the subtle inter-identity difference produced by similar appearances. In this paper, we propose a Viewpoint-Aware Loss (VAL) function to deal with these challenges. Specifically, we propose *partition and reunion* operations in VAL, which significantly shrinks the intra-identity distance and acquires viewpoint-invariant representations. In addition, we embed a multi-decision boundary mechanism in VAL. It contributes to enlarging the inter-identity distance. A comprehensive evaluation on two benchmarks shows the superiority of our method in contrast to a series of existing state-of-the-arts.

Index Terms— Vehicle re-identification, viewpoint-aware, loss function, representation learning

1. INTRODUCTION

Vehicle Re-identification (ReID) [1–3] has received a great deal of attention recently due to practical applications in urban surveillance, i.e., cross-camera tracking [4, 5] and multi-camera behavior analysis [6,7]. It aims to retrieve vehicle images in a large camera network, where the target images have the same identity and come from multiple cameras. Therefore, vehicle ReID faces two key challenges as shown in Fig. 1 (a). Similar appearance refers to the subtle inter-identity divergence of different vehicles with similar appearance under the same viewpoint. Viewpoint variations mean the large intra-identity difference of the same vehicle with different viewpoints.

To solve the above two challenges, some methods [8–10] propose to learn more robust features by focusing on loss functions. These loss functions are roughly summarized into two categories, *i.e.,* metric learning loss functions and representation learning loss functions. Metric learning losses, including Triplet loss [11], CircleLoss [12], *etc*, directly optimize the distance of features. Representation learning losses learn discriminative features by classification. Generally speaking, representation learning losses contains the softmax loss and other softmax-based losses [13–15], where Softmax has been widely used in the classification task. Although it can be utilized in vehicle ReID, features learned by Softmax occasionally have a large intra-identity distance and a small inter-identity distance as shown in Fig. 1 (c). When the appearance of the same vehicle changes dramatically under different viewpoints, the large intra-identity distance would arise. In addition, two different vehicles that have the same viewpoint may produce a small inter-identity distance. Note that when intra-identity distance is greater than inter-identity distance, the performance of vehicle ReID would drop obviously. Therefore, many softmax-based losses [14, 15] are applied to solve the problem caused by small inter-identity distance. However, these losses cannot effectively narrow the intra-identity distance, meaning that the intra-identity

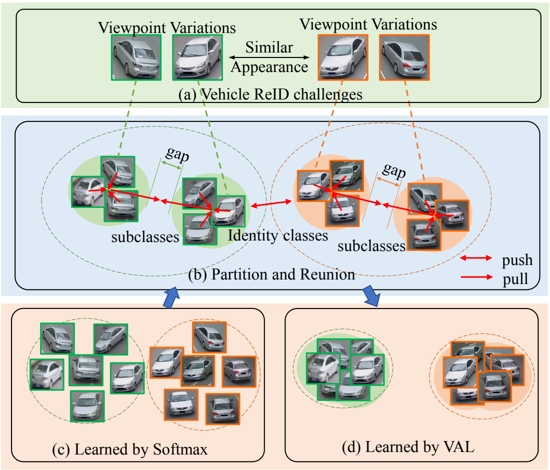

Fig. 1: Comparison of Softmax and the proposed VAL. Images with the same color belong to the identical vehicle.

distance may be greater than the inter-identity distance.

In this paper, we propose a Viewpoint-Aware Loss (VAL) function to address the problem caused by similar appearance and viewpoint variations. Our idea comes from the behavior of human beings recognizing vehicles. When humans observe an image of a vehicle, they can first distinguish its identity and viewpoint simultaneously, and then are able to learn the semantic relationship between identities and viewpoints. As shown in Fig. 1, we mimicking this process by the proposed VAL. Specifically, VAL adopts a **partition** approach to divide each identity class into various viewpoint subclasses. With subclass-based classification, we can obtain both identity and viewpoint information. Then, a **reunion** constraint is applied to VAL to group together different subclasses of the same identity, thus building semantic relationships between various subcategories. By the **partition and reunion** operations, we significantly shrink the intra-identity distance and acquire viewpoint-invariant features. Furthermore, we embed a multi-decision boundary mechanism in VAL to obtain more robust features. The decision mechanism takes different strategies to adjust the decision boundaries of intra-identity and inter-identity, respectively. While ensuring that the intra-identity distance is not increased, it effectively expands the inter-identity distance.

In summary, our main contributions include:

- We propose a Viewpoint-Aware Loss (VAL) function for vehicle ReID. By partition and reunion operations, VAL learns the relationship between identities and viewpoints and reduces intra-identity distance.
- We apply a multi-decision boundary mechanism to VAL, which employs various decision strategies and effectively increases the inter-identity distance.
- We provide reasonable theoretical analysis and experimental results on two benchmarks, demonstrating the effectiveness of our method.

2. METHODOLOGY

2.1. Motivation

We by deforming Softmax to analyze the vehicle ReID problem. The original Softmax is a widely used loss function for vehicle ReID, which is formulated as follows:

$$L_{\text{S}} = \frac{1}{N}\sum_{i=1}^{N} L_i = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{W_{y_i}^T x_i + b_{y_i}}}{\sum_{j=1}^{C} e^{W_j^T x_i + b_j}}, \tag{1}$$

where N is the batch size, $x_i \in \mathbb{R}^d$ is the deep feature and y_i is its identity class label. $W_j \in \mathbb{R}^d$ denotes the j-th column of the weight $W \in \mathbb{R}^{d\times C}$ and b_j is the bias of the identity classification layer. The number of identity classes is C.

We fix the bias $b_j = 0$, and fix the individual weight $\|W_j\| = 1$ and feature $\|x_i\| = 1$ by L_2 normalization. Then $W_j^T x_i + b_j = \gamma \cos\theta_j$, where γ is a scale factor, $\cos\theta_j = \frac{W_j^T x_i}{\|W_j^T\|\,\|x_i\|}$ and θ_j is the angle between the j-th column weight W_j and the feature x_i. The deep features of NormSoftmax are distributed on a hyper-sphere:

$$L_{\text{N-S}} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{\gamma\cos\theta_{y_i}}}{\sum_{j=1}^{C} e^{\gamma\cos\theta_j}}. \tag{2}$$

In NormSoftmax, the learned features are expected to satisfy $s_{y_i} > s_j, j \neq y_i$, which may result in small inter-class distance. Some losses [14, 15] add various margins to expand the inter-class distance. However, the problem of large intra-class distance caused by viewpoint variations still exists.

2.2. Viewpoint-Aware Loss

We propose a Viewpoint-Aware Loss (VAL) to mitigate the large discrepancy raised by viewpoint variations. By the partition and reunion operations in VAL, we remarkably narrow the intra-identity distance and acquire more robust features.

Partition. In Softmax/NormSoftmax, images of the identical vehicle belong to the same identity class and are generally captured from various viewpoints. We divide each identity class into V different viewpoint subclasses (also called identity-viewpoint classes) to mine viewpoint information. In other words, the images with the same identity and similar viewpoints are defined as the same class in VAL. Then the original C identity classes becomes $C \times V$ identity-viewpoint classes. We assigned $V = 2$ in this paper, indicating the front and rear viewpoint subclasses. The whole network structure is illustrated in Fig. 2 (a). VAL is applie to the identity-viewpoint classification layer. Original weight $W \in \mathbb{R}^{d\times C}$ is replaced by the weight of the identity-viewpoint classification layer $M \in \mathbb{R}^{d\times(C\times V)}$. In NormSoftmax, W_j is the center of the identity class j. Analogously, we reshape M to $d \times C \times V$ and $M_{j,k}$ is the center of the identity-viewpoint class (j, k). Then we obtain the partition loss:

$$L_{\text{P}} = -\frac{1}{N}\sum_{i=1}^{N} \log \frac{e^{\gamma\varphi(y_i,v_i)}}{\sum_{k}^{V} e^{\gamma\varphi(y_i,k)} + \sum_{j\neq y_i}^{C}\sum_{k}^{V} e^{\gamma\varphi(j,k)}}, \tag{3}$$

where V is the number of viewpoint subclasses and v_i is the viewpoint label of the i-th sample. $\varphi(j, k)$ calculates the cosine similarity between the feature x_i and class center $M_{j,k}$:

$$\varphi(j,k) = \cos\theta_{j,k} = \frac{M_{j,k}^T x_i}{\|M_{j,k}^T\|\,\|x_i\|}. \tag{4}$$

Based on the partition, we can simultaneously acquire the identity and viewpoint information. However, the partition destroys the integrity of the identity class, inevitably discarding some discriminative (identity-relevant) information and hampering the ReID performance. Thus, we propose a reunion approach to recover the lost information.

Reunion. The reunion builds semantic relationships between identity and viewpoint. More concretely, we design a

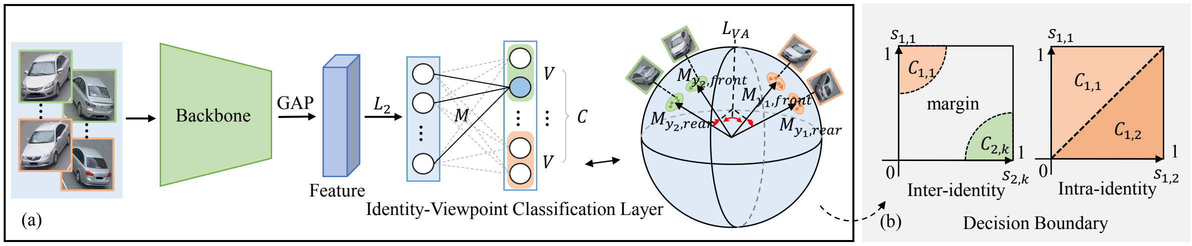

Fig. 2: The framework of our approach. (a): Images are represented as features, which are placed on the hyper-sphere by L_2 normalization and classified by the identity-viewpoint classification layer. (b): The two types of decision boundaries of VAL under two identities scenarios. The dashed line represents decision boundary, and white areas are decision margins.

reunion loss to enlarge the correlation among viewpoint subclasses of the same identity:

$$L_{\mathrm{R}} = -\frac{2}{V(V-1)} \sum_{j=1}^{C} \sum_{k=1}^{V} \sum_{l=1, l\neq k}^{V} \frac{M_{j,l}^{T} M_{j,k}}{\|M_{j,l}^{T}\| \, \|M_{j,k}\|}, \qquad (5)$$

where $M_{j,k}$ is the center of identity-viewpoint class (j, k). By minimizing the reunion loss, these subclasses are pulled closer and form more compact identity classes. Subsequently, we achieve VAL by combining the partition and reunion loss:

$$L_{\mathrm{VA}} = L_{\mathrm{P}} + \lambda L_{\mathrm{R}}, \qquad (6)$$

in which λ is a weight coefficient, indicating the degree of association between various subclasses.

2.3. Multi-decision Boundary Mechanism

In this section, we delve into the proposed VAL and analyze its decision boundary. Based on the analysis, we further propose a Multi-decision Boundary (MB) mechanism to expanding inter-identity distance.

Since the partition loss (L_{P}) determines the decision boundary of VAL, we analyze the situation where only L_{P} is used. Considering a scenario of two vehicle identities, there are totally four identity-viewpoint classes $C_{j,k}(j, k = 1, 2)$. In L_{P}, $\cos\theta_{j,k}$ denotes the cosine similarity between the learned feature and the weight of identity-viewpoint class $C_{j,k}$. We mark it as $s_{j,k}$. When classifying categories with the same identity (intra-identity), L_{P} forces $s_{1,1} > s_{1,2}$ for $C_{1,1}$, and when classifying categories with different identities (inter-identity), it forces $s_{1,1} > s_{2,k}(k = 1, 2)$ for $C_{1,1}$. In a nutshell, L_{P} applies the same constraints to all classifications between various categories, including intra-identity and inter-identity. However, we argue that inter-identity classification should be stricter. More concretely, given a feature and its ground truth class is $C_{1,1}$, we expect $s_{1,1} > s_{1,2} > s_{2,k}$.

To this end, we embed the MB mechanism in the partition loss (L_{P}). $\varphi(j, k)$ in Eq. 3 is redefined as follows:

$$\varphi(j,k) = \begin{cases} [O_{y_i} - \cos\theta_{y_i,k}]_{+}(\cos\theta_{y_i,k} - \Delta_{y_i}), & j = y_i \\ [\cos\theta_{j,k} - O_j]_{+}(\cos\theta_{j,k} - \Delta_j), & j \neq y_i \end{cases}, \qquad (7)$$

where O is optimum of $\cos\theta$ and $[\cdot]_{+}$ denotes the "cut-off at zero" operation. Δ determines the size of margins. Following [12], we set $O_{y_i} = 1 + m$, $\Delta_{y_i} = 1 - m$ for $j = y_i$ and $O_j = -m$, $\Delta_j = m$ for $j \neq y_i$, respectively.

Then we further analyze the effects of the new L_{P} (with MB). To distinguish these losses, we call the original L_{P} as L_{P} w/o MB and the original VAL as VAL w/o MB, respectively. Still considering the scenario of two vehicles, the four categories $C_{j,k}(j, k = 1, 2)$ are classified by L_{P} (with MB). In intra-identity classification, L_{P} (with MB) forces $\varphi(1, 1) > \varphi(1, 2)$, namely $s_{1,1} > s_{1,2}$ for $C_{1,1}$, which is same as before. Whereas in inter-identity classification, it makes $\varphi(1, 1) > \varphi(2, k)$, namely $(s_{1,1} - 1)^2 + (s_{2,k} - 0)^2 < 2m^2$ for $C_{1,1}$. As the decision boundary illustrated in Fig. 2 (b), we obtain a stricter classification between identities. We further consider the scenario of model trained by the new VAL (with MB). The similarity scores ideally satisfies $s_{1,1} > s_{1,2} > s_{2,k}$ for $C_{1,1}$, where L_{P} (with MB) makes $s_{2,k}$ as far as possible away rom $s_{1,1}$, and L_{R} makes $s_{1,2}$ as close as possible to $s_{1,1}$.

3. EXPERIMENTS

3.1. Datasets and Settings

We conduct experiments on two public datasets for vehicle ReID, including VeRi-776 [16] and VehicleID [17]. **VeRi-776** dataset consists of 49,357 images of 776 vehicles from 20 surveillance cameras. Its images are annotated 8 orientations in [18]. Based on the annotations, we generate the viewpoint labels for all training images. **VehicleID** is a large-scale dataset, which contains 26,267 vehicles and 221,763 images in total. Images of VehicleID only have 2 viewpoints. We annotate 300 images and train a binary classifier to infer the viewpoint labels of all training images.

We follow the commonly used metrics in ReID to quantitatively evaluate the performance by mean Average Precision (mAP) and Cumulative Matching Characteristic (CMC) curve at Top-k. We adopt ResNet50 [19] as the backbone. Input images are resized to 256 × 256 and augmented with color jittering, random flip. γ, m and λ in VAL are 48, 0.5 and 1.

Table 1: Evaluation (%) with various weights λ of VAL.

VAL	mAP	Top-1	Top-5
$\lambda = 0$	78.14	96.44	98.03
$\lambda = 0.1$	80.35	96.71	98.73
$\lambda = 1$	**81.36**	**96.72**	**98.75**
$\lambda = 10$	81.05	96.55	98.73

Table 2: Evaluation (%) with different number of viewpoints.

Number	mAP	Top-1	Top-5
$V = 2$	**81.36**	**96.72**	**98.75**
$V = 4$	80.63	96.10	98.69
$V = 8$	80.04	96.03	98.64

3.2. Ablation Study

In this section, we first analyze the weight λ in VAL. Then we delve into the number of viewpoint subclasses. Finally, we validate the effectiveness of various parts in VAL.

Effect of the weight coefficient. The reunion loss associates the identities and viewpoints, and its weight λ balances the features' identity information and viewpoint information. As shown in Tab. 1, we perform experiments with different λ on VeRi-776. When the weight is set to a small value ($\lambda = 0.1$), VAL can be better than L_{P}. Moreover, VAL gains the best performance with $\lambda = 1$.

Effect of the number of subclasses. We set V to 2, 4, 8 and conduct experiments on VeRi-776. The performance of VAL in Tab. 2 gradually declines as the number of subclasses increases. The reason for this phenomenon may be that the partition leads to the loss of identity-relevant information, and too many subcategories increase the intra-identity distance.

Effecttiveness of various parts in VAL. Method L_{P} w/o MB gets a lower performance than Softmax on mAP metric in Tab. 3, which suggests that the partition does lose the identity-relevant discriminative information. However, L_{P} w/o MB outperforms Softmax slightly on Top-1 and Top-5 metrics, indicating the partition may learn viewpoint-relevant discriminative information. When employing L_{R}, both VAL w/o MB and VAL achieve significant improvements on mAP metric. This comparison suggests that the reunion can effectively reduce the intra-identity distance. After using the MB mechanism, VAL is obviously higher than L_{P}. This phenomenon shows that MB mechanism can enhance the proposed VAL. Furthermore, when MB mechanism and the reunion loss are utilized together, VAL surpasses Softmax by a large margin.

Table 3: Performance of various loss in VAL on VeRi-776.

Method	mAP	Top-1	Top-5
Softmax	78.24	94.58	98.15
L_{P} w/o MB	78.12	95.95	98.33
L_{P}	78.14	96.44	98.03
VAL w/o MB	80.13	95.71	98.45
VAL	**81.36**	**96.72**	**98.75**

Table 4: Performance (%) comparison with state-of-the-art method on VeRi-776 and VehicleID (VeID).

Method	VeRi-776		VeID Small		VeID Medium		VeID Large	
	mAP	Top-1	Top-1	Top-5	Top-1	Top-5	Top-1	Top-5
OIFE*† [18]	51.42	68.30	-	-	-	-	67.00	82.90
VAMI*† [20]	61.32	85.92	63.12	83.25	52.87	75.12	47.34	70.29
EALN* [21]	57.44	84.39	75.10	88.10	71.80	83.90	69.30	81.40
PRNV† [22]	74.30	94.30	78.40	92.30	75.00	88.30	74.20	86.40
SAVER [23]	79.60	96.40	79.90	95.20	77.60	91.10	75.30	88.30
HCANet [1]	-	-	83.70	-	81.10	-	78.00	-
PVEN† [24]	79.50	95.60	84.70	**97.00**	80.60	94.50	77.80	92.00
Baseline	78.24	94.58	82.44	93.22	77.49	91.69	76.36	89.20
VAL	**81.36**	**96.72**	**85.69**	96.96	**81.25**	**95.02**	**78.19**	**93.02**

3.3. Comparison with the State-of-the-arts

We compare our approach with some methods on VeRi-776 and VehicleID in Tab. 4. (*) indicates the usage of viewpoint labels. (†) indicates the usage of additional annotations besides viewpoint labels (*e.g.* bounding box and segmentation mask). Our VAL can obtain the state-of-the-art performance on both VeRi-776 and VehicleID datasets, surpassing the Baseline a lot. Though VAL is slightly lower than PVEN [24] in Top-5 on small dataset of VechileID, PVEN needs extra segmentation mask, which is hard to achieve in practice. Our VAL only requires two-viewpoint labels, and we can obtain them at a low cost in real-world scenarios.

4. CONCLUSION

In this paper, we propose a Viewpoint-Aware Loss (VAL) to address the similar appearance and viewpoint variations challenges in vehicle ReID. The partition and reunion operations are implemented in VAL to reduce the intra-identity distance, which contributes to acquiring viewpoint-invariant representations. In addition, a multi-decision boundary mechanism is embedded in VAL to expand the inter-identity distance. Experiments on two benchmarks confirm the effectiveness of the proposed method.

5. ACKNOWLEDGEMENT

This study is partially supported by the National Key R&D Program of China(No.2018YFB2101100), the National Natural Science Foundation of China(No.61872025) , and the Science and Technology Development Fund, Macau SAR(File no.0001/2018/AFJ) and the Open Fund of the State Key Laboratory of Software Development Environment(No. SKLSDE-2021ZX-03) . Thank you for the support from HAWKEYE Group.

6. REFERENCES

[1] Xinze Dou, Yang Liu, Kai Lv, Zhang Xiong, Hao Sheng, and Computer Science, "High Confidence Attribute Recognition For Vehicle Re-Identification," in *ICIP*, 2021, pp. 2353–2357.

[2] Kai Lv, Hao Sheng, Zhang Xiong, Wei Li, and Liang Zheng, "Pose-based view synthesis for vehicles: A perspective aware method," *IEEE TIP*, pp. 5163–5174, 2020.

[3] Kai Lv, Heming Du, Yunzhong Hou, Weijian Deng, Hao Sheng, Jianbin Jiao, and Liang Zheng, "Vehicle re-identification with location and time stamps.," in *CVPRW*, 2019, pp. 399–406.

[4] Hao Sheng, Shuai Wang, Yang Zhang, Dongxiao Yu, Xiuzhen Cheng, Weifeng Lyu, and Zhang Xiong, "Near-online tracking with co-occurrence constraints in blockchain-based edge computing," *IEEE IoTJ*, pp. 2193–2207, 2021.

[5] Kai Lv, Hao Sheng, Zhang Xiong, Wei Li, and Liang Zheng, "Improving driver gaze prediction with reinforced attention," *IEEE TMM*, pp. 4198–4207, 2020.

[6] Hao Sheng, Kai Lv, Jiahui Chen, and Wei Li, "Robust visual tracking using correlation response map," in *ICIP*, 2016, pp. 1689–1693.

[7] Hao Sheng, Yang Zhang, Yubin Wu, Shuai Wang, Weifeng Lyu, Wei Ke, and Zhang Xiong, "Hypothesis testing based tracking with spatio-temporal joint interaction modeling," *IEEE TCSVT*, pp. 2971–2983, 2020.

[8] Yan Bai, Yihang Lou, Feng Gao, Shiqi Wang, Yuwei Wu, and Ling-Yu Duan, "Group-sensitive triplet embedding for vehicle reidentification," *IEEE TMM*, pp. 2385–2399, 2018.

[9] Ruihang Chu, Yifan Sun, Yadong Li, Zheng Liu, Chi Zhang, and Yichen Wei, "Vehicle re-identification with viewpoint-aware metric learning," in *ICCV*, 2019, pp. 8281–8290.

[10] Hao Sheng, Kai Lv, Yang Liu, Wei Ke, Weifeng Lyu, Zhang Xiong, and Wei Li, "Combining pose invariant and discriminative features for vehicle reidentification," *IEEE IoTJ*, pp. 3189–3200, 2021.

[11] Alexander Hermans, Lucas Beyer, and Bastian Leibe, "In defense of the triplet loss for person re-identification," *arXiv preprint arXiv:1703.07737*, 2017.

[12] Yifan Sun, Changmao Cheng, Yuhan Zhang, Chi Zhang, Liang Zheng, Zhongdao Wang, and Yichen Wei, "Circle loss: A unified perspective of pair similarity optimization," in *CVPR*, 2020, pp. 6397–6406.

[13] Feng Wang, Xiang Xiang, Jian Cheng, and Alan Loddon Yuille, "Normface: L_2 hypersphere embedding for face verification," in *ACM MM*, 2017, pp. 1041–1049.

[14] Feng Wang, Jian Cheng, Weiyang Liu, and Haijun Liu, "Additive margin softmax for face verification," *IEEE SPL*, pp. 926–930, 2018.

[15] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou, "Arcface: Additive angular margin loss for deep face recognition," in *CVPR*, 2019, pp. 4690–4699.

[16] Xinchen Liu, Wu Liu, Tao Mei, and Huadong Ma, "A deep learning-based approach to progressive vehicle re-identification for urban surveillance," in *ECCV*, 2016, pp. 869–884.

[17] Hongye Liu, Yonghong Tian, Yaowei Wang, Lu Pang, and Tiejun Huang, "Deep relative distance learning: Tell the difference between similar vehicles," in *CVPR*, 2016, pp. 2167–2175.

[18] Zhongdao Wang, Luming Tang, Xihui Liu, Zhuliang Yao, Shuai Yi, Jing Shao, Junjie Yan, Shengjin Wang, Hongsheng Li, and Xiaogang Wang, "Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification," in *ICCV*, 2017, pp. 379–387.

[19] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun, "Deep residual learning for image recognition," in *CVPR*, 2016, pp. 770–778.

[20] Yi Zhou and Ling Shao, "Viewpoint-aware attentive multi-view inference for vehicle re-identification," in *CVPR*, 2018, pp. 6489–6498.

[21] Yihang Lou, Yan Bai, Jun Liu, Shiqi Wang, and Ling-Yu Duan, "Embedding adversarial learning for vehicle re-identification," *IEEE TIP*, pp. 3794–3807, 2019.

[22] Bing He, Jia Li, Yifan Zhao, and Yonghong Tian, "Part-regularized near-duplicate vehicle re-identification," in *CVPR*, 2019, pp. 3997–4005.

[23] Khorramshahi Pirazh, Neehar Peri, Jun-Cheng Chen, and Rama Chellappa, "The devil is in the details: Self-supervised attention for vehicle re-identification," in *ECCV*, 2020, pp. 369–386.

[24] Dechao Meng, Liang Li, Xuejing Liu, Yadong Li, Shijie Yang, Zheng-Jun Zha, Xingyu Gao, Shuhui Wang, and Qingming Huang, "Parsing-based view-aware embedding network for vehicle re-identification," in *CVPR*, 2020, pp. 7101–7110.