Robust Vehicle Re-identification via Rigid Structure Prior

Minyue Jiang[1‡] Xuanmeng Zhang[1,2‡*] Yue Yu[1] Zechen Bai[1] Zhedong Zheng[2]
Zhigang Wang[1] Jian Wang[1] Xiao Tan[1] Hao Sun[1] Errui Ding[1] Yi Yang[2]
Baidu Inc.[1] ReLER Lab, University of Technology Sydney[2]

Abstract

Vehicle re-identification (re-id) is one of the most important components in the current intelligence transport system, benefiting both the smart traffic management and the optimal path planning. In this paper, we focus on developing a robust part-aware structure-based vehicle re-id system against the massive appearance changes due to the pose and illumination variants. Specifically, we apply the strong convolutional neural networks to extract the visual representation, which is based on the detected vehicle images. Taking one step further, we deploy a part detector to recognize different vehicle parts, such as front, back, left, and right, which explicitly introduce the prior knowledge on the structure of the rigid objective, i.e.*, vehicle. With the geometry information, we further harness different part feature extractors to filter wrong matches. By using this simple but effective strategy, we remove the hard negative candidates while maintaining high recall accuracy, combing general global-level coarse-grained re-id feature models with part-level fine-grained features. We achieved* 71.51% *mAP in the vehicle re-id track of the AI City Challenge 2021, which verified the effectiveness and scalability of the proposed structure-based method. The code will be available at* https://github.com/Xuanmeng-Zhang/AICITY2021-Track2.

1. Introduction

With the booming of artificial intelligence techniques, especially the development of deep learning, the Intelligent Transport System (ITS) is attracting public attention in recent years [42, 39, 30, 7]. With the wide deployment of sensors throughout the city, there is an immense opportunity of making use of data captured by the sensors to make transportation systems smarter. The AI City Challenge Workshop [14] focuses on different computer vision tasks about ITS, including vehicle counting, re-identification (re-id), and tracking, etc. In this paper, we mainly focus on the vehicle re-identification task. As a critical task in computer vision, vehicle re-id aims to retrieve the same vehicle of interest captured by different cameras [41]. Vehicle re-id is of vital importance for ITS in the smart city, for it can help traffic engineers understand journey times along entire corridors.

Deeply learned vehicle re-id models have shown dominant representation ability in most benchmarks, such as Veri-776 [10], VehicleID [9] and VehicleNet [41]. These methods share a lot of common points with the person re-id approaches. For instance, most of the existing works [26, 30, 7] are characterized by a Siamese network and trained over an identity classification loss and a metric learning objective (such as triplet loss [4]). Recent developments in this field usually take advantage of extra information including vehicle types and colors [37] to assist the model training and improve the model scalability in different deployment environments.

In this work, we aim to build a robust vehicle re-id system given the rigid structure prior knowledge of vehicles. From our observation, vehicle re-id in the real-world scenarios is challenging in: (1) similar car shape and background; (2) data distribution gap between training and test data caused by style, resolution, etc. Inspired by recent successful part-based re-id methods [21, 45, 25], we consider the part information and design a part-aware approach to address the challenge in distinguishing similar appearance. To make the parts well aligned, we explicitly integrate the vehicle rigid structure into the partition process as a piece of prior knowledge. To mitigate the train-test divergence, we resort to a semi-supervised approach that involves the unseen data into model optimization by assigning reliable pseudo labels. Experiment results on vehicle re-id track of AI City Challenge 2021 show the effectiveness of the proposed approach.

The rest of this paper is organized as follows. Section 2 reviews and discusses related work. Section 3 describes the proposed method in detail, followed by the experimental results and comparison in Section 4. We provide the conclusion in Section 6.

*Work done during an internship at Baidu Inc. ‡ Equal Contribution.

2. Related Work

2.1. Supervised Vehicle re-id

Vehicle re-id aims to retrieve a specific vehicle from huge galleries captured by different cameras. To achieve this goal, more and more works [11, 5, 23] use powerful convolutional neural networks (CNNs) to learn discriminative representations of vehicle images instead of handcrafted features. To further improve performance, several works [18, 6, 31] employ widely-used deep metric learning methods to pull vehicle images of the same identity close and push different vehicles far away. However, these methods still suffer from the viewpoint variation problem. To address this issue, Yao *et al*. [29] use a graphic engine to generate synthetic data with various viewpoints. Other works [47, 13] attempt to synthesize more multi-view vehicle images by the popular generative adversarial networks (GANs) [44]. Zhou *et al*. [48] propose to extract view-invariant features by directly learn a viewpoint-aware network. Besides, a few works [19, 22, 40] make use of prior Spatio-temporal knowledge to narrow down the possible search space to filter out masses of hard negative samples. Although achieving great success, these methods are difficult to distinguish subtle differences between similar vehicles.

To alleviate the aforementioned problem, many part-based methods are proposed to mine fine-grained information. Liu *et al*. [12] divide the feature map into several stripes and learn discriminative features from these stripes individually. Wang *et al*. [27] train and extract part representations based on several predefined key points, and align them for feature comparison. Zhang *et al*. [36] develop a part-guided attention network which adaptively locates import parts and combines global-local features for vehicle re-id. In this paper, we also explore the part-based features as similarity constraints.

2.2. Unsupervised Domain Adaptation Object re-id

Supervised object re-id methods always suffer from dramatic performance drop when applied to unseen scenarios. In recent years, unsupervised domain adaptation (UDA) methods have drawn increasing attention, since they can mitigate the generalization problem by transferring re-id knowledge from labeled data to unlabeled data. Usually, UDA object re-id methods adopt a two-step pipeline for training. First, training a re-id model by supervised learning on the labeled dataset. Second, predicting pseudo-labels for the unlabeled data by clustering algorithms and taking pseudo-labels as supervision signals to fine-tune the re-id model. Based on this popular fashion, Song *et al*. [20] provide more theoretical analyses. Lin *et al*. [8] proposes to leverage the attribute to help the knowledge transfer and reduce the pseudo label noise. Similarly, to resist label noise, Zhang *et al*. [34] gradually selects reliable samples for training, and Li *et al*. [28] propose an asymmetric co-teaching framework where two distinct modules generate pseudo labels for each other, Ge *et al*. [3] develops a mutual mean-teaching framework to ensemble parameters of different training state. Zhai *et al*. [32] further extend the mutual learning strategy to multiple heterogeneous networks to enhance generalization ability.The recent state-of-the-art performance[1] has been achieved in [38] by using the mean teacher strategy and uncertainty [43]. In this paper, we also follow the basic UDA object re-id pipeline to learn robust features for the unlabeled scenario.

3. Method

3.1. Overview

The framework of the proposed method is shown in Figure 1. According to training models, we train two types of models: global-level coarse-grained models and part-level fine-grained models. By aggregating two different types of model features, we use part-aware verification to filter the hard negative gallery. Then, the post-processinng method, *i.e*., part-aware verification, traclet verification and re-ranking [35] is applied to get the final result.

3.2. Global-aware Model

The global-level coarse-grained model training procedure can be divided into two stages, *i.e*., Stage-I and Stage-II. In Stage-I, following the champion solution in the 4*th* AICity Challenge [2] [40], both the synthetic and real-world training data are used for supervised training. In Stage-II, we utilize the trained models to generate pseudo labels on the unlabeled test images. Then the labeled test images are added to the training set to retrain the global-level model, With the help of pseudo labels, the model can learning the knowledge on the unseen scenarios.

3.2.1 Stage-I

The baseline model is shown in Figure 2. The whole pipeline consists of four components, *i.e*., image pre-processing, backbone, head, and loss.

Image pre-processing. To make the model get better performance, we introduce many pre-processing operations, such as random crop, random flip, random erasing [46] and Auto-augment [1].

Backbone. In order to conduct model ensemble, we adopt 5 strong model, *i.e*., HRNet [24], Res2Net[2], ResNet-

[1] https://github.com/layumi/Person_reID_baseline_pytorch/tree/master/leaderboard

[2] https://github.com/PaddlePaddle/Research/tree/master/CV/PaddleReid

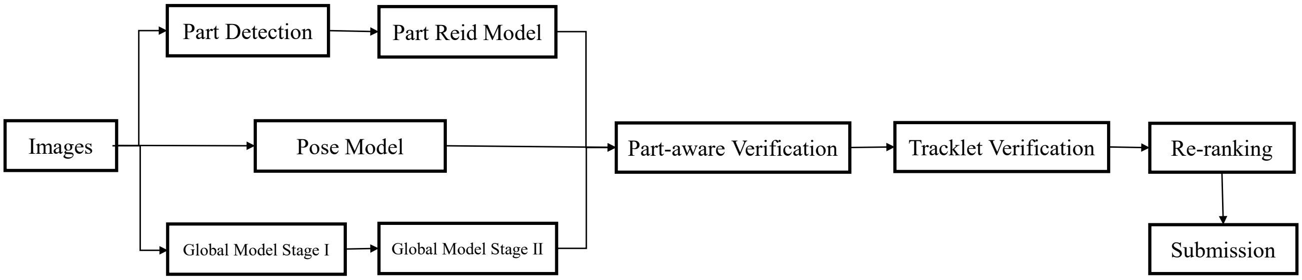

Figure 1. **An overview of the proposed framework.** The proposed method mainly contains two components: feature extraction by using ensemble models and post-processing. Global-aware features and fine-grained part-aware features are aggregated. Besides, Verification from different perspectives, including camera, part, and tracklet are used in the post-processing.

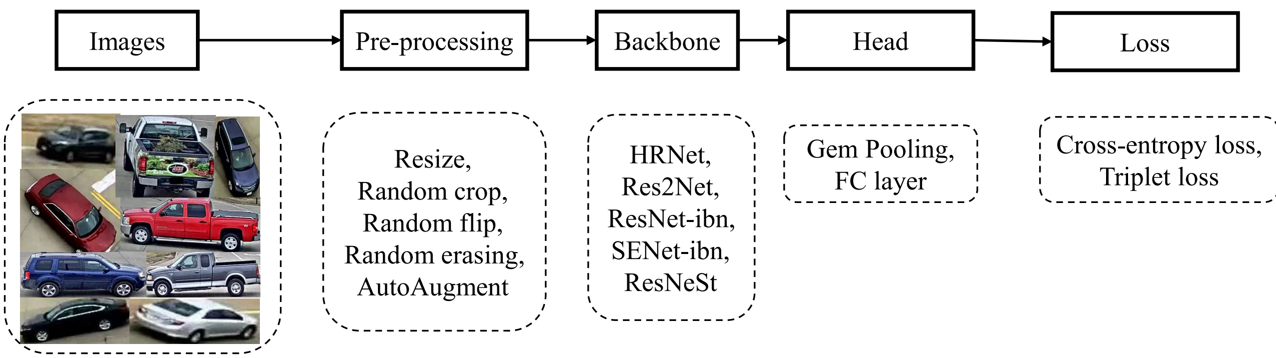

Figure 2. **The baseline model.** The whole pipeline consists of four components, *i.e.*, image pre-processing, backbone, head, and loss.

ibn [15], SENet-ibn [15] and ResNeSt [33] as the backbones.

Head. Considering the superior retrieval performance, we employ the generalized-mean(GeM) pooling [16] to aggregate the feature map into the global feature.

Loss. We deploy the two widely adopted losses, *i.e.*, the cross-entropy loss and the triplet loss to optimize the model.

3.2.2 Stage-II

We notice the models directly training on the synthetic data and real-world data perform poorly in six new test camera views, *i.e.*, camera 41-46. The reason is that these images captured by the new cameras don't appear in the training set. To bridge the data distribution gap between training and testing data, we adopt the unsupervised clustering method, *i.e.*, DBSCAN, to label the test data with pseudo labels. We train the models on the test dataset with two different pseudo labels and then perform model ensemble to get more robust features. Since we don't know how many identities are included in the new scenario, we set the parameter eps, *i.e.*, the maximum neighbor distance, to different values, aiming to generate two cluster results. As shown in Figure 3, the pseudo labels change when the parameter eps set to different values, *i.e.*, r_1 and r_2. When testing, we compute the global-level feature similarity S_{global} according to the cosine similarity of extracted feature generate from global-aware model.

3.3. Part-aware Model

Part Detection Model. In order to extract the part-level fine-grained features of the vehicle, we divided the vehicle into four parts, *i.e.*, the front part, the back part, the side part, and the top part. Following the way of object detection annotation, we select images from the training set of CityFlowV2 to annotate the bounding box of defined parts. As shown in Figure 4, we adopt YOLOv3 [17] to detect the parts of the vehicle.

Part-level Re-id Model. After getting the detected part of the vehicle, the cropped part images are used to train the corresponding part re-id models. Similar to the method of extracting the global features, we deploy the same model structure as in Figure 2 to extract the four part-level fine-grained features. Since the input of the model has changed from the entire vehicle to a part, the part model focuses more on the details of the vehicle components, such as the luggage racks and skylight. Benefit from this local informa-

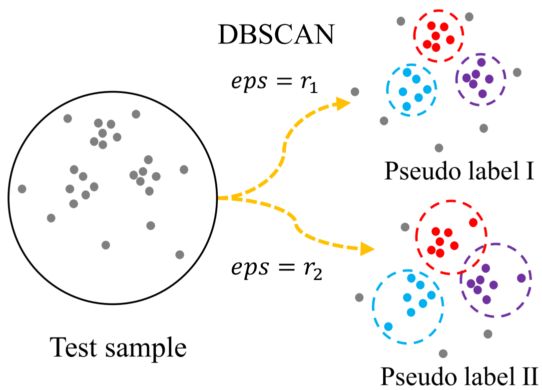

Figure 3. **The pseudo labels generation procedure**. In the semi-supervised training stage, we adopt DBSCAN clustering to generate the pseudo labels on test data. The grey vertexes represent the unlabeled samples, while the colored vertexes indicate the labeled samples. We can see the samples assigned with same labels (vertexes inside the dotted circle) change when eps set to r_1 and r_2.

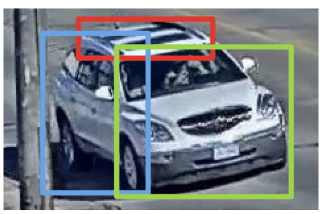

Figure 4. **The part detection results.** The red bounding box indicates the top part, and the blue box and green box indicate the side and front part respectively.

tion, the part-aware models will further enhance the overall re-identification ability.

Pose Model. However, vehicle re-identification task suffers from the similar vehicle pose, which is mainly composed of the direction and the visible-part of the vehicle. To reduce the bias of vehicle pose, we deploy two classification models, *i.e.*, direction classification model and visible-part classification model, to distinguish the vehicles with a similar pose. As shown in Figure 5, the driving direction can be divided into eight directions, *i.e.*, west, northwest, north, northwest, east, southeast, south, and southwest. We first select images from the training set of CityFLowV2 to annotate the directions. Then, a vanilla classification model, *i.e.*, ResNet101, is used to extract the direction feature. As for visible-part classification model, we take the annotations part detection as the training set. Similar to the direction model, we deploy the classification model with ResNet101 backbone to extract the visible-part feature. After that, the direction similarity $S_{direction}$ and the visible-part similarity $S_{visible-part}$ can be computed by the direction feature and visible-part feature respectively. As shown in Eq 1, we compute the pose similarity S_{pose} by adding $S_{direction}$ and

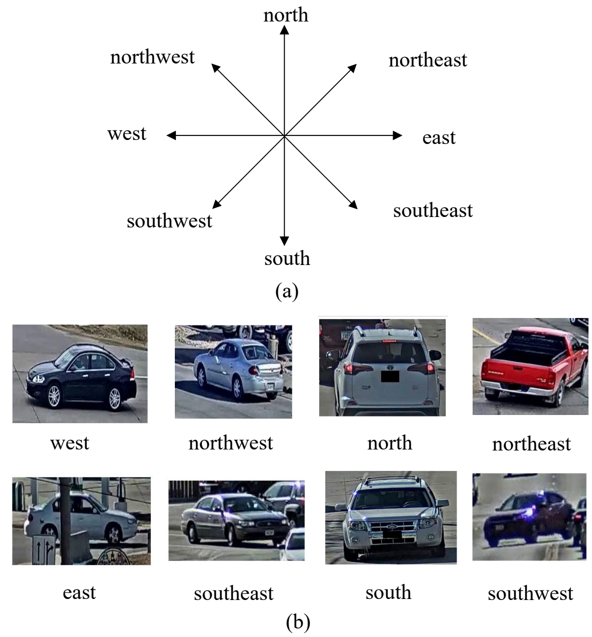

Figure 5. **The direction annotations.** In Figure (a), we define eight directions according to the driving direction of the vehicle. Figure (b) Shows examples of different directions.

$S_{visible-part}$.

$$\boldsymbol{S}_{pose} = S_{direction} + S_{visible-part} + \beta, \tag{1}$$

Here, we set the value of β to 2 to ensure S_{pose} is non-negative.

3.4. Post-processing

Part-aware Verification. In this subsection, we calculate the part similarity by the part features. We first compute the cosine similarity of four vehicle parts separately. Since the query image and gallery image may have different parts, we only calculate the cosine similarity score when the two images have the same parts, otherwise, the similarity of this part is set to be zero. The final part similarity is obtained by computing the maximum of the four part-similarities if any part of the two vehicles can be matched, otherwise, the final part similarity is the same as the global similarity:

$$S_{part} = max(S_{front}, S_{back}, S_{top}, S_{side}), \tag{2}$$

where S_{front}, S_{back}, S_{top}, S_{side} represent the front-part similarity, back-part similarity, top-part similarity, side-part similarity separately.

To punish those samples with similar poses but different details, the vehicle similarity can be denoted as follows:

$$\boldsymbol{S}_{vehicle} = \begin{cases} S_{global} - \alpha * S_{pose} * (1 - S_{part}) \text{ if } S_{part} > 0 \\ S_{global} - \alpha * S_{pose} * (1 - S_{global}) \text{ otherwise} \end{cases}, \quad (3)$$

where α is the balance weights of the global similarity and the part similarity. Empirically, we set the α as 0.1.

Tracklet Verification Based on the hypothesis that query and gallery samples are captured by different cameras, we punish the samples captured by the same cameras but in different tracklets.

Re-ranking The re-ranking [35] is applied to get the final result.

4. Experiments

In this section, we show the effectiveness and scalability of each part of our method.

4.1. Dataset

The main differences of the track2 dataset this year are two-folds. The first change is the expansion of the real data both in the training set and test set. A total of 880 vehicles are annotated. 440 vehicles are used for training. The remaining 440 vehicles are for testing. There are 85,058 images in total. 52,717 images are in the training set and 31,238 images are in the test set which nearly double the test set comparing with last year. Another difference is that new camera views, *i.e.*, camera 41 to 46, are added to the test. We notice that the images captured by these cameras have much lower resolution than others.

4.2. Implementation Details

We train the models based on the paddlepaddle and pytorch framework. The input images are resized to {384, 400, 416 } to train different resolution models. For the backbones, HRNet-W48 [24], Res2Net101[2], ResNet-101-ibn [15], SENet-101-ibn [15] and ResNeSt101 [33] are deployed for feature extraction. Besides, we also adopt some widely-used training strategy such as cosine-decay learning rate scheduler and learning rate warm-up.

4.3. Evaluation Metrics

The metric used to rank the performance of each team is the mean Average Precision (mAP) of the top-K(K=100) matches, which measures the mean of average precision (the area under the Precision-Recall curve) over all the queries. The evaluation system also provides other measures, such as the CMC-1, CMC-5 and CMC-10 hit rates, which measures the percentage of the queries that have at least one true positive result ranked within the top 1, 5 and 10 positions, respectively.

Table 1. Effect of Part-aware Verification.

	mAP(%)	CMC1(%)
w/o part-aware verification	69.74	80.69
w part-aware verification	71.51	82.05

Table 2. Effect of Semi-supervised learning.

	mAP(%)	CMC1(%)
w/o semi-supervised	63.85	74.71
w semi-supervised	67.94	78.42

Table 3. Final Public Results.

	Team Name	mAP(%)
1	DMT	74.45
2	**NewGeneration (Ours)**	71.51
3	CyberHu	66.50
4	For Azeroth	65.50
5	IDo	63.73
6	KeepMoving	63.64
7	MegVideo	62.52
8	aiem2021	62.16

4.4. Ablation Study

We mainly study the impact of part-aware verification and the semi-supervised learning method which are the two most significant components in our framework.

Part-aware Verification. As we can see from Table 1, part-aware verification can filter hard negative samples effectively and the mAP can be boost from 69.74% to 71.51%. The visual revised matching results are shown in Figure 6. Without part-aware verification, some hard negative matches appear due to pose and illumination similarity. With the help of part-aware verification, we can remove these wrong matches easily.

Semi-supervised Training. As we can see from Table 2, there remains a large performance gap between the usage of semi-supervised training (67.94%) or not (63.85%). This also demonstrates the massive influence of the poor image quality on the vehicle re-identification task.

5. Comparison with Other Teams

As shown in the leaderboard, we achieved 71.51% mAP in the vehicle re-id track of the 5th AI City Challenge, which verified the effectiveness and scalability of the proposed structure-based method. Furthermore, we visualize the retrieval results with and with post-processing respectively (see Figure 7 and Figure 8).

Figure 6. **Visualization of matching results with and without part-aware verification.** The first and the second row for each query are the matching results without and with part-aware verification respectively.

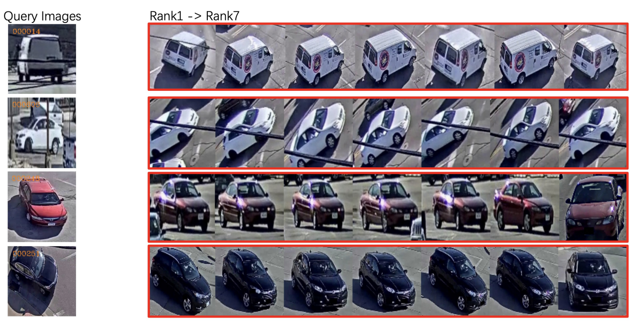

Figure 7. **Visualization of matching results before post-processing.** The first column shows the selected query images captured by different cameras, and each row shows the top 7 gallery images retrieved from left to right according to the similarity score. The images in green boxes are true positives, while the images in red boxes are false positives.

6. Conclusion

In this paper, we proposed a robust vehicle re-identification system with the rigid structure prior knowledge of vehicles. Specifically, we combine global-aware models along with the part-aware fine-grained models in a coherent manner. The experimental results on the private test set of the AI City Challenge 2021 have proved the scalability of the proposed structure-based method.

References

[1] Ekin D Cubuk, Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V Le. Autoaugment: Learning augmentation strategies from data. In *CVPR*, pages 113–123, 2019. 2

[2] Shanghua Gao, Ming-Ming Cheng, Kai Zhao, Xin-Yu Zhang, Ming-Hsuan Yang, and Philip HS Torr. Res2net: A new multi-scale backbone architecture. *IEEE transactions on pattern analysis and machine intelligence*, 2019. 2, 5

[3] Yixiao Ge, Dapeng Chen, and Hongsheng Li. Mutual mean-teaching: Pseudo label refinery for unsupervised domain

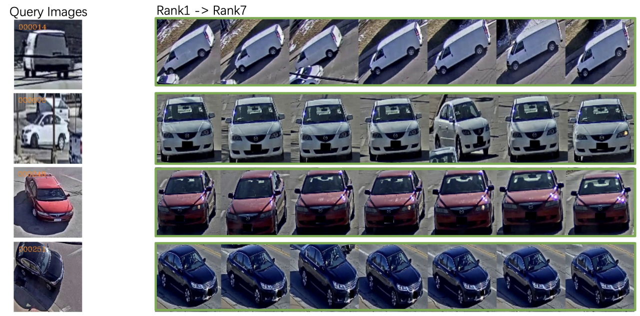

Figure 8. **Visualization of final results.** The first column shows the selected query images captured by different cameras, and each row shows the top 7 gallery images retrieved from left to right according to the similarity score. The images in green boxes are true positives, while the images in red boxes are false positives.

adaptation on person re-identification. In *International Conference on Learning Representations*, 2020. 2

[4] Alexander Hermans, Lucas Beyer, and Bastian Leibe. In defense of the triplet loss for person re-identification. *arXiv:1703.07737*, 2017. 1

[5] Aytaç Kanaci, Xiatian Zhu, and Shaogang Gong. Vehicle re-identification in context. In *GCPR*. 2

[6] Ratnesh Kumar, Edwin Weill, Farzin Aghdasi, and Parthasarathy Sriram. Vehicle re-identification: an efficient baseline using triplet embedding. In *IJCNN*, pages 1–9, 2019. 2

[7] Qingming Leng, Mang Ye, and Qi Tian. A survey of open-world person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 30(4):1092–1108, 2019. 1

[8] Yutian Lin, Liang Zheng, Zhedong Zheng, Yu Wu, Zhilan Hu, Chenggang Yan, and Yi Yang. Improving person re-identification by attribute and identity learning. *Pattern Recognition*, 95:151–161, 2019. 2

[9] Hongye Liu, Yonghong Tian, Yaowei Wang, Lu Pang, and Tiejun Huang. Deep relative distance learning: Tell the difference between similar vehicles. In *CVPR*, 2016. 1

[10] 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. 1

[11] Xinchen Liu, Wu Liu, Tao Mei, and Huadong Ma. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In *ECCV*, volume 9906 of *Lecture Notes in Computer Science*, pages 869–884, 2016. 2

[12] Xiaobin Liu, Shiliang Zhang, Qingming Huang, and Wen Gao. RAM: A region-aware deep model for vehicle re-identification. In *ICME*, pages 1–6, 2018. 2

[13] Yihang Lou, Yan Bai, Jun Liu, Shiqi Wang, and Ling-Yu Duan. Embedding adversarial learning for vehicle re-identification. *IEEE Trans. Image Process.*, 28(8):3794–3807, 2019. 2

[14] Milind Naphade, Shuo Wang, David C Anastasiu, Zheng Tang, Ming-Ching Chang, Xiaodong Yang, Liang Zheng,

Anuj Sharma, Rama Chellappa, and Pranamesh Chakraborty. The 4th ai city challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 626–627, 2020. 1

[15] Xingang Pan, Ping Luo, Jianping Shi, and Xiaoou Tang. Two at once: Enhancing learning and generalization capacities via ibn-net. In *ECCV*, pages 464–479, 2018. 3, 5

[16] Filip Radenović, Giorgos Tolias, and Ondřej Chum. Fine-tuning cnn image retrieval with no human annotation. *IEEE transactions on pattern analysis and machine intelligence*, 41(7):1655–1668, 2018. 3

[17] Joseph Redmon and Ali Farhadi. Yolov3: An incremental improvement. *arXiv preprint arXiv:1804.02767*, 2018. 3

[18] Artsiom Sanakoyeu, Vadim Tschernezki, Uta Büchler, and Björn Ommer. Divide and conquer the embedding space for metric learning. In *CVPR*, pages 471–480, 2019. 2

[19] Yantao Shen, Tong Xiao, Hongsheng Li, Shuai Yi, and Xiaogang Wang. Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals. In *ICCV*, pages 1918–1927, 2017. 2

[20] Liangchen Song, Cheng Wang, Lefei Zhang, Bo Du, Qian Zhang, Chang Huang, and Xinggang Wang. Unsupervised domain adaptive re-identification: Theory and practice. *Pattern Recognition*, 102:107173, 2020. 2

[21] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling. In *ECCV*, 2018. 1

[22] Xiao Tan, Zhigang Wang, Minyue Jiang, Xipeng Yang, Jian Wang, Yuan Gao, Xiangbo Su, Xiaoqing Ye, Yuchen Yuan, Dongliang He, Shilei Wen, and Errui Ding. Multi-camera vehicle tracking and re-identification based on visual and spatial-temporal features. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops, CVPR Workshops 2019, Long Beach, CA, USA, June 16-20, 2019*, pages 275–284, 2019. 2

[23] Yi Tang, Di Wu, Zhi Jin, Wenbin Zou, and Xia Li. Multi-modal metric learning for vehicle re-identification in traffic surveillance environment. In *ICIP*, pages 2254–2258, 2017. 2

[24] Jingdong Wang, Ke Sun, Tianheng Cheng, Borui Jiang, Chaorui Deng, Yang Zhao, Dong Liu, Yadong Mu, Mingkui Tan, Xinggang Wang, et al. Deep high-resolution representation learning for visual recognition. *IEEE transactions on pattern analysis and machine intelligence*, 2020. 2, 5

[25] Tingyu Wang, Zhedong Zheng, Chenggang Yan, Jiyong Zhang, Yaoqi Sun, Bolun Zhenga, and Yi Yang. Each part matters: Local patterns facilitate cross-view geo-localization. *IEEE Transactions on Circuits and Systems for Video Technology*, 2021. 1

[26] 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. 1

[27] 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*, pages 379–387, 2017. 2

[28] Fengxiang Yang, Ke Li, Zhun Zhong, Zhiming Luo, Xing Sun, Hao Cheng, Xiaowei Guo, Feiyue Huang, Rongrong Ji, and Shaozi Li. Asymmetric co-teaching for unsupervised cross-domain person re-identification. In *AAAI*, pages 12597–12604, 2020. 2

[29] Yue Yao, Liang Zheng, Xiaodong Yang, Milind Naphade, and Tom Gedeon. Simulating content consistent vehicle datasets with attribute descent. In *ECCV*, volume 12351, pages 775–791, 2020. 2

[30] Mang Ye, Jianbing Shen, Gaojie Lin, Tao Xiang, Ling Shao, and Steven C. H. Hoi. Deep learning for person re-identification: A survey and outlook. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2021. 1

[31] Yuhui Yuan, Kuiyuan Yang, and Chao Zhang. Hard-aware deeply cascaded embedding. In *ICCV*, pages 814–823, 2017. 2

[32] Yunpeng Zhai, Qixiang Ye, Shijian Lu, Mengxi Jia, Rongrong Ji, and Yonghong Tian. Multiple expert brainstorming for domain adaptive person re-identification. *arXiv preprint arXiv:2007.01546*, 2020. 2

[33] Hang Zhang, Chongruo Wu, Zhongyue Zhang, Yi Zhu, Zhi Zhang, Haibin Lin, Yue Sun, Tong He, Jonas Mueller, R Manmatha, et al. Resnest: Split-attention networks. *arXiv preprint arXiv:2004.08955*, 2020. 3, 5

[34] Xinyu Zhang, Jiewei Cao, Chunhua Shen, and Mingyu You. Self-training with progressive augmentation for unsupervised cross-domain person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 8222–8231, 2019. 2

[35] Xuanmeng Zhang, Minyue Jiang, Zhedong Zheng, Xiao Tan, Errui Ding, and Yi Yang. Understanding image retrieval re-ranking: A graph neural network perspective. *arXiv preprint arXiv:2012.07620*, 2020. 2, 5

[36] Xinyu Zhang, Rufeng Zhang, Jiewei Cao, Dong Gong, Mingyu You, and Chunhua Shen. Part-guided attention learning for vehicle re-identification. *CoRR*, abs/1909.06023, 2019. 2

[37] Aihua Zheng, Xianmin Lin, Chenglong Li, Ran He, and Jin Tang. Attributes guided feature learning for vehicle re-identification. *arXiv preprint arXiv:1905.08997*, 2019. 1

[38] Kecheng Zheng, Cuiling Lan, Wenjun Zeng, Zhizheng Zhan, and Zheng-Jun Zha. Exploiting sample uncertainty for domain adaptive person re-identification. *AAAI*, 2021. 2

[39] Liang Zheng, Yi Yang, and Alexander Hauptmann. Person re-identification: Past, present and future. *arXiv:1610.02984*, 2016. 1

[40] Zhedong Zheng, Minyue Jiang, Zhigang Wang, Jian Wang, Zechen Bai, Xuanmeng Zhang, Xin Yu, Xiao Tan, Yi Yang, Shilei Wen, and Errui Ding. Going beyond real data: A robust visual representation for vehicle re-identification. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition, CVPR Workshops 2020, Seattle, WA, USA, June 14-19, 2020*, pages 2550–2558, 2020. 2

[41] Zhedong Zheng, Tao Ruan, Yunchao Wei, Yi Yang, and Tao Mei. Vehiclenet: learning robust visual representation for

vehicle re-identification. *IEEE Transactions on Multimedia*, 2020. 1

[42] Zhedong Zheng, Yunchao Wei, and Yi Yang. University-1652: A multi-view multi-source benchmark for drone-based geo-localization. *ACM Multimedia*, 2020. 1

[43] Zhedong Zheng and Yi Yang. Rectifying pseudo label learning via uncertainty estimation for domain adaptive semantic segmentation. *International Journal of Computer Vision*, pages 1–15, 2021. 2

[44] Zhedong Zheng, Liang Zheng, and Yi Yang. Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. In *ICCV*, 2017. 2

[45] Zhedong Zheng, Liang Zheng, and Yi Yang. Pedestrian alignment network for large-scale person re-identification. *IEEE Transactions on Circuits and Systems for Video Technology*, 2018. 1

[46] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *AAAI*, volume 34, pages 13001–13008, 2020. 2

[47] Yi Zhou and Ling Shao. Cross-view GAN based vehicle generation for re-identification. In *Brit. Mach. Vis. Conf.* 2

[48] Yi Zhou and Ling Shao. Viewpoint-aware attentive multi-view inference for vehicle re-identification. In *CVPR*, pages 6489–6498, 2018. 2