



An Improved AlignedReID Method for Vehicle Re-identification

Yu Deng, Jinhui Xu, Shuo Chen and Yong Song

EasyChair preprints are intended for rapid dissemination of research results and are integrated with the rest of EasyChair.

August 25, 2019

An Improved AlignedReID Method for Vehicle Re-identification ^{*}

Yu Deng^{1,2,†}, Jinhui Xu^{1,2,†}, Shuo Chen², and Yong Song¹

1.Shandong University, Weihai, Shandong Province, China
songyong@sdu.edu.cn

2.ShenZhen OEASY Limited company, ShenZhen, Guangdong Province, China

Abstract. In this paper, we propose an improved vehicle re-identification method based on the combination between the AlignedReID and the Stochastic Weight Averaging (SWA). AlignedReID extracts a global feature and local features of a vehicle’s image and performs joint learning. Local automatic alignment is achieved by computing the shortest path between the two sets of local features, so that global feature learning can benefit from local feature learning. By running an optimizer with a high constant learning rate, the SWA averages the weight of the model to ensure that a better weight combination can be found. Our improved method surpasses the most advanced methods on the VehicleID dataset and VeRi-776 dataset. In order to better solve the task of vehicle re-identification in residential area, we have made the Oeasyparking dataset and experimented with our methods, and achieved good results.

Keywords: Vehicle Re-identification · AlignedReID · SWA.

1 Introduction

Vehicle re-identification (ReID) is a technique that uses computer vision to determine whether a particular vehicle is present in an image or video sequence acquired by multiple cameras. Since there is very little difference in the camera output between the same brand and style except the number plate, the specificity of the results is far inferior to that of individual pedestrians. It is difficult to capture the clear details of the vehicle ID under the monitoring distance by multiple traffic camera, particularly since local features are not easy to extract or express clearly. Vehicle ReID research is relatively new and the accuracy of vehicle ReID systems and methods are not as good as that of a human observer [13].

^{*} Supported by the ShenZhen OEASY Limited company, the National Natural Science Foundation of China under Grants (61573213, 61673245, 61603214, 61803227), National Key Research and Development Plan of China under Grant 2017YFB1300205, Shandong Province Key Research and Development Plan under Grants (2016ZDJS02A07, 2018GGX101039), China Postdoctoral Science Foundation under Grant 2018M630778 and Independent Innovation Foundation of Shandong University under Grant 2018ZQXM005. [†] Equal contribution

According to a forecast by Gartner, Inc., there will be 20.4 billions connected smart technology in the world by 2020. Since most of the world’s population is concentrated in urban areas [1], data from traffic and surveillance cameras form an important data component of connected technology. Therefore, data from imaging systems can contain a large quantity of important information. Using this large amount of information effectively can make cities safer and smarter. Therefore, vehicle ReID plays an important role in this process.

Traditional vehicle ReID methods mostly focus on a vehicle’s low-dimensional characteristics, such as a vehicle’s shape and color [17]. In recent years, with the popularity of deep learning techniques, many researchers have focused on fine-grained image-based vehicle classification and license-plate recognition. Unfortunately, using global information about a vehicle cannot provide local differentiation information. Therefore, a vehicle cannot be well recognized in heavy rain, fog, night and in low-resolution video recordings. This means that vehicle ReID in more challenging environments is nearly impossible, e.g., when a license plate is partially or entirely hidden, when license plates change over a single trip, when annual inspection stickers change, when camera blind spots obscure a vehicle or when a vehicle has constant jitter. Thus, advanced vehicle ReID methods often still requires additional supervision and needs to combine both global and local image features.

In this paper, we use the AlignedReID [18] method, which is a method for pedestrian recognition, as a baseline and introduce an integrated SWA [9] method. In the learning stage, the AlignedReID network has two branches, global branch and local branch, which are used to learn global feature and local features simultaneously. We load two identical pre-loaded models based on AlignedReID, the first one stores the SWA average, and the second one updates the current weights of the running average model. In the local branch, AlignedReID introduce a shortest path loss to align the vehicle images, then discard local branch in the inference stage. The first model is used as the final model in the inference stage. Like AlignedReID, the mutual learning approach [19] is still works in our method.

Current vehicle ReID methods are mainly based on VeRi-776 dataset [13] and VehicleID dataset [11]. We make the following two contributions in this work: (1) We modify a pedestrian recognition model and apply it to vehicle recognition. (2) We add a SWA model-based optimization algorithm to the AlignedReID model and find that our improved method outperforms state-of-the-art methods by a large margin on VehicleID and VeRi-776 datasets. (3) We propose a dataset which servers as residential area for vehicle parking, named "Oeasy-Parking". The dataset contains all the time periods of the day.

2 Related work

Vehicle ReID methods mainly rely on finding vehicle license plate information. This is the most reliable method at present because a license plate has the most accurate identification information about a vehicle under normal circumstances,

just as the face is the most reliable identity information in pedestrian detection. Finding vehicle license plate information is not always effective, particularly when the license plate lose efficacy. For example, in due to lighting, shooting angle, fouling and other factors, the vehicle license plate recognition is wrong or unrecognizable. In some cases, the vehicle does not have a license plate or a occlusion plate or only a side image of the vehicle and so on. In this case, other identity information of the vehicle, such as appearance, color, special logo, annual inspection mark, will be a new strategy for vehicle identification. The research of vehicle ReID mainly focuses on how to extract more accurate vehicle ID feature information at present. Vehicle ReID is of great significance for solving the analysis and processing of surveillance video, especially for vehicle retrieval across cameras. Vehicle ReID will undoubtedly play an important role in intelligent management, security and automatic charging of parking lots.

In recent years, researchers have explored many aspects of vehicle ReID. Hongye Liu et al. [11] proposed a depth relative distance learning model, a new loss function Coupled Cluster Loss was put forward by improving Triplet Loss function. Zhongdao Wang et al. [16] proposed a vehicle ReID framework, including a feature embedding model with attitude invariance and a space-time regularization model. Yi Zhou et al. proposed a visual Angle perception targeting multi-view reasoning (VAMI) model [21], which only needs visual information to solve the problem of multi-view vehicle weight recognition. Ratnesh Kumar et al. [10] tackle the problem of vehicle ReID in a camera network utilizing triplet embeddings. They conducted extensive evaluations of the losses used for vehicle ReID (including comparisons or triad losses), demonstrating that best practices using learning embedding are superior to most previous methods proposed in the vehicle ReID literature. These methods almost utilize the global appearance features of vehicle images and ignore local discriminative regions.

With the re-emergence of deep learning, the Convolutional Neural Network (CNN) uses metric learning methods such as triplet loss, impoved triplet loss, and triplet hard loss to learn features in an end-to-end format. Many CNN-based methods focus on the global features of the vehicle while ignoring the spatial local information of the picture. This will cause some problems: (1) Errors appear in the vehicle detection frame, thus affecting the learning of features. (2) CNN may learn irrelevant features when vehicles are covered by external objects. (3) For two models of the same brand, or similar models of different brands, there is almost no difference except the number of the license plate. Vehicle recognition tasks are mostly carried out when license plate recognition fails. Therefore, distinguishing between two vehicles in such cases is nearly impossible by only relying on global features. Reference [12] and reference [3] combine global features and local features, but the local information is not aligned, feature learning will be affected by wrong calibration frame and external object occlusion. Although reference [16] performed local alignment by extracting key points, they failed to learn from each other with global features and could not represent local features with global features, which affected the convergence rate.

The AlignedReID method uses TriHard loss [6], a learning algorithm where a baseline input is compared to a positive and negative input, with batch hard mining as the metric loss. Then, softmax loss and metric learning loss is combined to accelerate convergence. Finally, the dynamic programming method is used to automatically align local information and use mutual learning strategies to learn global features and local features. Therefore, this method effectively solves the shortcomings of previous research.

The Nadam optimization algorithm combines two optimization methods, one by Nesterov and the other by Adam, to develop a method that has a stronger constraint on the learning rate and a more direct impact on the gradient update step [4]. By estimating the marginalized effect of label-dropout, a mechanism called the Label-Smoothing Regularization (LSR) for regularizing the classifier layer is proposed by InceptionV3 [15], which can reduce the risk of model overfitting and improve the adaptability and mathematical rigor of the models. Zhong Z et al. [20] proposed a reordering method (Re-ranking) that can further improve the accuracy of this method. Therefore, we apply Nadam optimization, LSR and Re-ranking to AlignedReID to further improve the performance of the model. We promote the benchmark model by combining the SWA method, which improves the performance and generalization ability of the model.

3 Our Approach

3.1 Benchmark model

AlignedReID uses Densenet121 (a Convolutional Neural Network) [8] to extract the feature map of the input vehicle images. The output of the last convolution layer is the feature map ($C \times H \times W$, where C is the number of channels and $H \times W$ is the spatial size). And the global feature is a C -dimension vector, which is extracted on the feature map by adopting global pooling. The local feature for each row can be extracted by applying a global pooling in the horizontal direction, and then a 1×1 convolution is applied to reduce the number of channels (from C to c). Each local feature is a c -dimension vector representing the horizontal portion of the vehicle image. Therefore, AlignedReID can represent a vehicle image with a global feature and H local features.

The difference between the images of two vehicles is the summation of the global distance and the local distance. The global distance is the L2 Euclidean distance of the global feature vector. Two images of the same vehicle share similar local features than their semantic counterparts. For local distances between feature vectors, the local components of the vector are dynamically matched element-wise, from top to bottom, to find the alignment geometry with the smallest total distance. Therefore, given local feature vectors of two images, $F=\{f_1, f_2, \dots, f_h\}$, $G=\{g_1, g_2, \dots, g_h\}$, we can use an element-wise transformation to regularize the distance to $[0, 1)$. Therefore, the distance is:

$$d_{i,j} = \frac{e^{\|f_i - g_j\|_2} - 1}{e^{\|f_i - g_j\|_2} + 1} \quad i, j \in 1, 2, 3, \dots, H \quad (1)$$

where $d_{i,j}$ represents the distance between the i -th vertical bar of the first picture and the j -th vertical bar of the second picture. A distance matrix D consists of all pairwise distances, where each element (i, j) is labeled $d_{i,j}$.

The local distance between two images is the shortest path of the total distance between the $(1, 1)$ coordinate to (H, H) coordinate in the matrix D . This distance is calculated using dynamic programming:

$$S_{i,j} = \begin{cases} d_{i,j} & i = 1, j = 1 \\ S_{i-1,j} + d_{i,j} & i \neq 1, j = 1 \\ S_{i,j-1} + d_{i,j} & i = 1, j \neq 1 \\ \min(S_{i-1,j}, S_{i,j-1}) + d_{i,j} & i \neq 1, j \neq 1 \end{cases} \quad (2)$$

where $S_{i,j}$ is the total distance between the shortest path between the $(1, 1)$ and (i, j) coordinates in matrix D , and $S_{H,H}$ is the total distance of the final shortest path (i.e. local distance) between the two images.

3.2 Improved method

AlignedReID is similar to many CNN models in that it is easy to get trapped in some a local optimal solutions during the learning phase in such a way that the loss function does not continue to decline. However, an integrated approach based on deep learning can help the model to jump out of such local optimal solutions more easily. The Snapshot Ensembling method [7] proposed by Gao Huang et al. and the Fast Geometric Ensembling (FGE) method [5] posed by Timur Garipov are the most advanced integration methods.

To benefit from Snapshot Ensembling or FGE, multiple models must be stored in a library to obtain accurate predictions based on these models. The predictions are averaged to determine the final prediction. Therefore, to improve the performance of a collection, more computation is required.

To determine the computational cost of the integration method used, and to get similar or better performance, Izmailov et al. proposed the SWA method [9]. SWA is similar to FGE method, but its calculation loss is very small. SWA defeats the current most advanced Snapshot Ensembling method, and its performance is similar to that of FGE. SWA achieves an excellent balance between performance and cost. So, we try to improve performance of our technique by introducing SWA into our benchmark model.

SWA is to begin with a pre-training solution, then run an optimization method with a constant learning rate, and finally average the weights of the model. The high constant learning rate plan ensures that optimization methods compares a set of possible solutions rather than simply converge to a certain weights combination. Izmailov et al. believe that the local minimum generated at the end of each learning rate cycle tends to accumulate in the marginal region, which has a low loss value. By averaging several such local minimum points, we can determine the global optimal solution with lower loss values, generalization, and versatility.

Unlike Snapshot Ensembling and FGE, there is no need to integrate over many models, but only to integrate over two distinct models and determine the SWA to update the equation:

$$W_{SWA} \leftarrow \frac{W_{SWA} \times n_{SWA} + W}{n_{SWA} + 1} \quad (3)$$

The first model stores the SWA average (W_{SWA}), which is used as the final model after the end of the training step for prediction. At the end of each learning rate cycle, the current weight of the second model is used to update the weights of the running average model, i.e., the weighted average of the existing average weight and the new weight generated by the second model (W in the formula). The n_{SWA} is the number of updates for the first model. Therefore, only one model needs to be trained for every two models that are stored during training. In the forecasting phase, only the model with the average weight is needed. Using this single model is much faster than using integration methods that require multiple models for prediction.

4 Experiments

4.1 Datasets

Our research is mainly based on VeRi-776 and VehicleID datasets. In this paper, we also propose a dataset which serves as residential area for vehicle parking, named "Oeasy-Parking".

Typically, previous datasets were collected by road surveillance cameras, and their datasets were mostly for the daytime scenarios. Differently, Oeasy-Parking dataset is a large-scale which contains 48,206 images with about 1/4 of the data in this dataset is in the darkness of the night. It contains all the time periods of the day. And the dataset also including a lot of glare, rainy days, foggy days conditions which obtained from the real sample of the community vehicle. Our dataset is taken from the vehicle entrance and exit of seven communities in Shenzhen, Guangdong Province, China, which is captured by 20 non-overlapping surveillance cameras. The time span is nearly nine months. We have preprocessed the image for data cleaning. In order to avoid the privacy of the owner and the neural network to learn the extra license plate information during the deep learning process, we have use Gaussian blurring on the license plates.

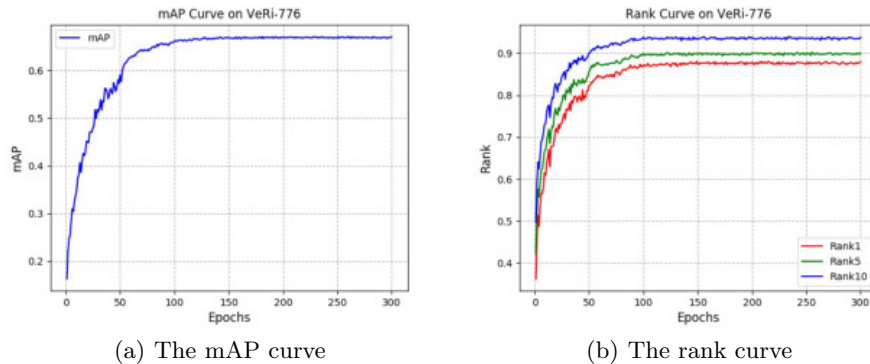
The data split statistics of VeRi-776, VehicleID and Oeasy-Parking datasets are summarised in Table 1.

4.2 Implementation details

We performed experiments on VeRi, VehicleID and Oeasy-Parking datasets using baseline and improved methods, and briefly analyzed the experimental results of the VeRi, VehicleID (small) and Oeasy-Parking datasets.

Table 1. Data split of vehicle ReID datasets evaluated in our experiments.

Dataset	Training IDs / Images	Query IDs / Images	Gallery IDs / Images
VeRi-776	576 / 37778	200 / 1678	200 / 11579
VehicleID(samll)	4416 / 37245	739 / 1678	739 / 4693
VehicleID(medium)	13164 / 113346	1464 / 3433	1464 / 9672
VehicleID(large)	13164 / 113346	2213 / 5108	2213 / 14295
Oeasy-Parking	4369 / 33360	1660 / 3007	1660 / 11839

**Fig. 1.** mAP and rank curves for baseline on VeRi-776.

Benchmark model experiments and results. In the benchmark model, we use techniques such as LSR [15] and Re-ranking [20]. We processed the images in the datasets into 256×256 to adapt to the task of vehicle ReID.

On the VeRi-776 dataset, we achieve a 67.6% mAP and an 88.3% rank-1 accuracy on the benchmark model. On the VehicleID (small) dataset, we achieve an 82.4% mAP and a 93.5% rank-1 accuracy on the benchmark model. The detailed results of baseline for the two VeRi-776, VehicleID and Oeasy-Parking datasets are shown in Table 2.

The mAP curve and the rank curve of the benchmark model are shown in Fig. 1. We can see the accuracy of the mAP and rank tend to level off when the epoch exceed 100. It indicates that the learning of benchmark model meets a bottleneck at this time, that is, the performance of the benchmark model approaches or reaches the limit.

The results of improved method. To further improve accuracy, we integrate using the SWA method for optimization, based on a benchmark model. We set SWA to start at an epoch to 161. The initial learning rate of the SWA is 0.0001. We use the learning rate change schedule in SWA. It can be seen that when epoch reaches 300, mAP still has an upward trend, so we increase the maximum value of epoch to 2000. In the VeRi-776 dataset, the mAP reaches

Table 2. Performance of baseline on VeRi-776, VehicleID and Oeasyp-Parking

Models		VeRi-776	VehicleID(small)	VehicleID(medium)	VehicleID(large)	Oeasyp-Parking
Baseline	mAP	0.676	0.824	0.801	0.775	0.966
	rank1	0.883	0.935	0.940	0.941	0.967
	rank5	0.906	0.973	0.965	0.967	0.991

77.1%, which is 9.5% higher than the highest value in the benchmark model; rank-1 accuracy reaches 94.4%, which is 6.1% higher than the baseline model. The results of the VehicleID (small) dataset obtains an 84.0% mAP, which is 1.6% higher than the highest mAP of the benchmark model; rank-1 accuracy reaches 95.1%, which is 1.6% higher than the highest rank-1 accuracy of the benchmark model. The results of the improved method on VeRi-776, VehicleID and Oeasyp-Parking are shown in Table 3.

In the benchmark model and the improved method, our Oeasyp-Parking dataset has achieved experimental results far superior to the VeRi and VehicleID datasets. The main reasons of our analysis are as follows: (1) The images in the dataset are all cropped in high-definition images, and the features are relatively clear. (2) The images of the dataset are basically the front and side images of the vehicle due to the surveillance cameras collected from the community, whose features are relatively simpler.

Table 3. Performance of improved method on VeRi-776, VehicleID and Oeasyp-Parking

Models		VeRi-776	VehicleID(small)	VehicleID(medium)	VehicleID(large)	Oeasyp-Parking
Baseline+SWA	mAP	0.771	0.840	0.812	0.791	0.969
	rank1	0.944	0.951	0.947	0.945	0.990
	rank5	0.963	0.977	0.970	0.971	0.999

Table 4. Comparison with recent works on VeRi-776

Methods	mAP	rank1	rank5
RAM [12]	0.615	0.886	0.940
Wang et al. [16]	0.514	0.894	-
AFL [17]	0.534	0.821	0.923
Siamese-CNN+Path-LSTM [14]	0.583	0.835	0.900
VAMI [21]	0.501	0.770	0.908
MoV1+BS [10]	0.676	0.902	0.929
Baseline+SWA (ours)	0.771	0.944	0.963

The mAP curves and the rank curves of improved method are shown in Fig. 2. It can be seen that when epoch reaches maximum, mAP still has an upward

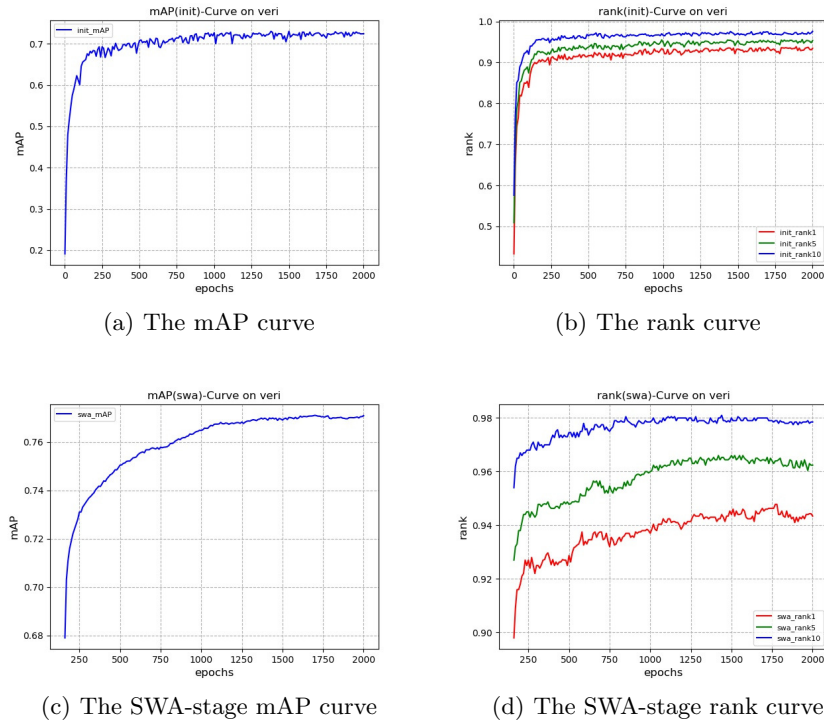


Fig. 2. The mAP and rank curves for improved method on VeRi-776.

trend. Comparing with Fig. 1, we can see that Fig. 2 has a performance curve better than the benchmark model. Therefore, SWA does improve the performance of the benchmark model and get better results.

Comparison results of the open source datasets. We studied the research results of vehicle ReID based on VeRi-776 and VehicleID in recent years. We then compared those results to our experimental results. The results are shown in Table 4 and Table 5. The MoV1+BS method of [10] obtained a 67.6% mAP on the VeRi-776 dataset with a 90.2% rank-1 accuracy. An 86.2% mAP and a 78.8% rank-1 accuracy was obtained for the VehicleID (small) dataset. Their research results, at the time, were the highest precision results based on two datasets. This result is comparable to our research results.

As shown in Table 4, on the VeRi-776 dataset, our improved method has a mAP that is 9.5% higher than MoV1+BS, and our rank-1 accuracy is 4.2% higher than MoV1+BS on the VeRi-776 dataset. As we can see from Table 5, our mAP is 2.2% lower than MoV1+BS on the VehicleID (small) dataset; our rank-1 accuracy is significantly better than that of MoV1+BS, which is greater than 16.3%. Our mAP value is lower than the mAP value for MoV1+BS on the VehicleID (small) dataset.



Fig. 3. Top-10 of the search results. The blue box indicates query images, the green boxes are correct hits, and the red boxes are incorrect hits.

Table 5. Comparison with recent works on VehicleID

Methods	VehicleID(small)			VehicleID(medium)			VehicleID(large)		
	mAP	rank1	rank5	mAP	rank1	rank5	mAP	rank1	rank5
RAM [12]	-	0.752	0.915	-	0.723	0.870	-	0.677	0.845
GSTE [2]	0.754	0.759	0.842	0.743	0.748	0.836	0.724	0.740	0.827
Wang et al. [16]	-	0.677	0.829	-	-	-	-	0.670	0.829
VAMI [21]	0.631	0.833	0.924	-	0.529	0.751	-	0.473	0.703
MoV1+BS [10]	0.862	0.788	0.962	0.817	0.734	0.926	0.782	0.693	0.895
Baseline+SWA (ours)	0.840	0.951	0.977	0.812	0.947	0.970	0.791	0.945	0.971

In order to observe the performance of the improved method visually, we show the search results of some vehicle pictures, as shown in Fig. 3. We select the picture of 10 vehicles from the query on VeRi-776. For each of the 10 vehicles, we retrieve 10 different images of the same vehicle from gallery, and rank the search results of each vehicle in query according to confidence.

These examples show that the overall performance of the improved method is good. However, in a few cases, such as the vehicle is severely occluded, the difference between the vehicles is too small, the image of the vehicle is blurred, our method still needs to be further improved.

5 Conclusion

Experiments show that AlignedReID is also well suited for vehicle ReID problems. SWA can obtain global optimal solutions with lower loss values, generalization and versatility. The improved method of combining a SWA method with the AlignedReID method achieves state-of-the-art results on VeRi-776 dataset and VehicleID dataset. Despite this, the accuracy of vehicle ReID is relatively low

compared to pedestrian ReID, and there is still a long distance from large-scale commercial use. At the same time, the real scene is more complicated. Therefore, solving the vehicle ReID is still a heavy task.

References

1. Antonio Marin-Reyes, P., Palazzi, A., Bergamini, L., Calderara, S., Lorenzo-Navarro, J., Cucchiara, R.: Unsupervised vehicle re-identification using triplet networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 166–171 (2018)
2. Bai, Y., Lou, Y., Gao, F., Wang, S., Wu, Y., Duan, L.Y.: Group-sensitive triplet embedding for vehicle reidentification. *IEEE Transactions on Multimedia* **20**(9), 2385–2399 (2018)
3. Cui, C., Sang, N., Gao, C., Zou, L.: Vehicle re-identification by fusing multiple deep neural networks. In: 2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA). pp. 1–6. IEEE (2017)
4. Dozat, T.: Incorporating nesterov momentum into adam (2016)
5. Garipov, T., Izmailov, P., Podoprikin, D., Vetrov, D.P., Wilson, A.G.: Loss surfaces, mode connectivity, and fast ensembling of dnns. In: Advances in Neural Information Processing Systems. pp. 8789–8798 (2018)
6. Hermans, A., Beyer, L., Leibe, B.: In defense of the triplet loss for person re-identification. *arXiv preprint arXiv:1703.07737* (2017)
7. Huang, G., Li, Y., Pleiss, G., Liu, Z., Hopcroft, J.E., Weinberger, K.Q.: Snapshot ensembles: Train 1, get m for free. *arXiv preprint arXiv:1704.00109* (2017)
8. Huang, G., Liu, Z., Van Der Maaten, L., Weinberger, K.Q.: Densely connected convolutional networks. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 4700–4708 (2017)
9. Izmailov, P., Podoprikin, D., Garipov, T., Vetrov, D., Wilson, A.G.: Averaging weights leads to wider optima and better generalization. *arXiv preprint arXiv:1803.05407* (2018)
10. Kumar, R., Weill, E., Aghdasi, F., Sriram, P.: Vehicle re-identification: an efficient baseline using triplet embedding. *arXiv preprint arXiv:1901.01015* (2019)
11. Liu, H., Tian, Y., Yang, Y., Pang, L., Huang, T.: Deep relative distance learning: Tell the difference between similar vehicles. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 2167–2175 (2016)
12. Liu, X., Zhang, S., Huang, Q., Gao, W.: Ram: a region-aware deep model for vehicle re-identification. In: 2018 IEEE International Conference on Multimedia and Expo (ICME). pp. 1–6. IEEE (2018)
13. Liu, X., Liu, W., Mei, T., Ma, H.: A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In: European Conference on Computer Vision. pp. 869–884. Springer (2016)
14. Shen, Y., Xiao, T., Li, H., Yi, S., Wang, X.: Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 1900–1909 (2017)
15. Szegedy, C., Vanhoucke, V., Ioffe, S., Shlens, J., Wojna, Z.: Rethinking the inception architecture for computer vision. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2818–2826 (2016)

16. Wang, Z., Tang, L., Liu, X., Yao, Z., Yi, S., Shao, J., Yan, J., Wang, S., Li, H., Wang, X.: Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification. In: Proceedings of the IEEE International Conference on Computer Vision. pp. 379–387 (2017)
17. Wu, C.W., Liu, C.T., Chiang, C.E., Tu, W.C., Chien, S.Y.: Vehicle re-identification with the space-time prior. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops. pp. 121–128 (2018)
18. Zhang, X., Luo, H., Fan, X., Xiang, W., Sun, Y., Xiao, Q., Jiang, W., Zhang, C., Sun, J.: Alignedreid: Surpassing human-level performance in person re-identification. arXiv preprint arXiv:1711.08184 (2017)
19. Zhang, Y., Xiang, T., Hospedales, T.M., Lu, H.: Deep mutual learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 4320–4328 (2018)
20. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 1318–1327 (2017)
21. Zhou, Y., Shao, L.: Aware attentive multi-view inference for vehicle re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 6489–6498 (2018)