Real-time vehicle orientation classification and viewpoint-aware vehicle re-identification

https://doi.org/10.2352/ISSN.2470-1173.2021.10.IPAS-234

Real-Time Vehicle Orientation Classification and Viewpoint-Aware Vehicle Re-Identification

Oliver Tamas Kocsis, Tunc Alkanat [†]***, Egor Bondarev, and Peter H.N. de With***
Eindhoven University of Technology, Department of Electrical Engineering, 5612 AP Eindhoven, The Netherlands
[†]Corresponding author: t.alkanat@tue.nl

Abstract

Vehicle re-identification (re-ID) is based on identity matching of vehicles across non-overlapping camera views. Recently, the research on vehicle re-ID attracts increased attention, mainly due to its prominent industrial applications, such as post-crime analysis, traffic flow analysis, and wide-area vehicle tracking. However, despite the increased interest, the problem remains to be challenging. One of the most significant difficulties of vehicle re-ID is the large viewpoint variations due to non-standardized camera placements. In this study, to improve re-ID robustness against viewpoint variations while preserving algorithm efficiency, we exploit the use of vehicle orientation information. First, we analyze and benchmark various deep learning architectures in terms of performance, memory use, and cost on applicability to orientation classification. Secondly, the extracted orientation information is utilized to improve the vehicle re-ID task. For this, we propose a viewpoint-aware multi-branch network that improves the vehicle re-ID performance without increasing the forward inference time. Third, we introduce a viewpoint-aware mini-batching approach which yields improved training and higher re-ID performance. The experiments show an increase of 4.0% mAP and 4.4% rank-1 score on the popular VeRi dataset with the proposed mini-batching strategy, and overall, an increase of 2.2% mAP and 3.8% rank-1 score compared to the ResNet-50 baseline.

Index Terms– Scene understanding, image retrieval, vehicle re-identification, CNN.

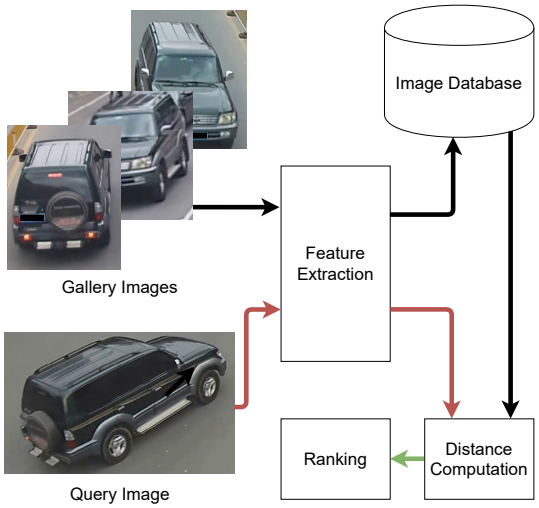


Figure 1. *Typical flow of vehicle re-ID: Images of previously encountered vehicles are passed on to the feature extraction, and a database of feature vectors is constructed (the gallery set). Then, at run-time, the feature vector for an image with an unknown vehicle (query) is extracted. Finally, a mathematical distance metric between the query and gallery feature vectors are computed and the resulting distances are ranked, where a low distance indicates a resembling match.*

Introduction

At present, public safety continues to be a major concern of society, although the safety issues are changing over the years. An important theme in safety is the tracking of specific people and vehicles. As a result, vehicle re-ID has been receiving increasing scientific attention in recent years from the computer vision community, following the increased demand for its practical applications. Vehicle re-identification (re-ID) aims to find an identity-sensitive correspondence between vehicle images which are taken from non-overlapping camera views. In other words, vehicle re-ID enables recognition of vehicles that re-appear at different geographical locations.

Vehicle re-ID enables intelligent surveillance video analysis that targets numerous use cases. For instance, it is possible to combine intra-camera tracking with vehicle re-ID over a multi-camera network to reveal long trajectories of individual vehicles. This valuable information can be used by traffic anomaly detection systems that decrease the response time of authorities, thereby increasing efficiency and safety of transportation. Similarly, vehicle re-ID enables wide-area tracking, which can be used for post-crime analysis. In addition, vehicle re-ID enhances traffic-flow analysis, which provides valuable statistics for road-utilization assessment and leads to efficient designs of public roads and intersections [1, 2]. Additional use cases of vehicle re-ID include congestion pricing, tolling on highways, and logistic applications with autonomous tracking of products transported by trucks in shipyards.

Despite the increased scientific attention and recent advancements, the vehicle re-ID problem is still far from being solved, mainly due to diverse and difficult challenges inherent to the problem. Low-resolution images, environmental occlusions, variations in lighting conditions, and low intra-class or high inter-class similarities are difficult challenges that negatively affect the performance of many algorithms. Another particularly notorious challenge is the broad variability in viewpoints. Depending on vehicle orientation and camera placement, the appearance and the visible sides of vehicles may change drastically, resulting in significant differences in feature characteristics. For example, for a

side-view image of a passenger car, it may be possible to get discriminative features from shape and appearance of the rims. However, such a feature will simply be absent in a front- or a back-view of the same vehicle. Thus, vehicle re-ID algorithms should be able to determine the relevant features for different viewpoints. Furthermore, not only the orientation of the vehicles, but also the camera placement influences the viewpoint significantly. In an ideal multi-camera network, camera mounting height and angle of the camera field-of-view (FOV) with respect to the observed road should be nearly identical for all sensors. However in practice, this is seldom the case due to realistic constraints, which even further complicates the re-ID of vehicles.

Taking a technical perspective on computer vision for vehicle re-ID purposes, it is evident that deep learning revolutionized the technical field since it gained popularity in 2012 [3]. Today, most state-of-the-art algorithms addressing various problems utilize deep learning to efficiently model complex data structures from given annotations. This approach can also be successfully applied to the problem of vehicle re-ID. Recent best-performing algorithms employ various deep network architectures, additional semantic data, novel training methodologies and data pre-processing techniques, to increase the performance and robustness of vehicle re-ID algorithms against various challenges. Similarly, in this study, we aim to improve re-ID performance by utilizing the viewpoint data (orientation of vehicles). When supplying the feature extraction with the viewpoint data in addition to raw pixel values of a vehicle image, we hypothesize that it is possible to extract richer and highly-specialized features. In other words, our aim is to teach the network to extract viewpoint-specific features that are more descriptive with respect to the identity of vehicles. To this end, we first explore the possibility of extracting accurate viewpoint classification labels from input images bearing computational cost in mind. Then, various different popular deep learning architectures are benchmarked on their performance of vehicle orientation classification. We then propose a novel vehicle re-ID algorithm that uses the extracted orientation labels. Specifically, our algorithm is a branched network, where the branching is controlled by the viewpoint label of the input image. The algorithm does not increase the inference time of the re-ID feature extraction network and is framework-independent, which means that the proposed approach can be applied to any re-ID feature extraction network to improve performance. Further, to compensate for the reduced effective batch-size per viewpoint branch and to increase the efficiency of training, we propose a novel viewpoint-aware mini-batching strategy that ensures proper training of our architecture. To summarize, the contributions of our study are as follows.

- Benchmarking of various architectures on their performance and computational cost for classifying vehicle orientation.
- Novel viewpoint-aware branched network (VABN) architecture that takes viewpoint information into account, to yield better re-ID feature extraction while preserving a low inference time.
- Viewpoint-aware branching strategy that improves the training performance.

The remainder of this paper is organized as follows. The section on Related Work provides a short overview of vehicle re-ID methods in literature. In Methodology, the problem of viewpoint classification is discussed, and our methodology is described, explaining the VABN architecture in detail. The section on Experiments presents the quantitative experimental results for both orientation classification and re-ID. Finally, concluding remarks are given in the Conclusion section.

Related Work

Due to its valuable applications in the industry, the problem of re-ID has been studied in many forms and for multiple target entities, including the re-ID of people, vehicles, and maritime vessels. Among these, person re-ID has received particularly strong scientific attention. As a result, a vehicle variant of the re-ID problem includes many methodologies from the person re-ID literature, adapted to the vehicle case. Thus, prior to moving on to specific vehicle re-ID related work, a brief overview of person re-ID algorithms is presented.

Person Re-identification. The problem of person re-ID aims to re-identify people across non-overlapping cameras. The most common methodology of person re-ID is to find a transformation, $f : \mathbb{R}^{H\times W} \rightarrow \mathbb{R}^{N_\mathrm{f}}$, that extracts an identity-sensitive, N_f-dimensional feature vector for a given image of size $H \times W$. In a typical re-ID scenario, feature vectors for the images with known identities are computed and stored in a database (the gallery set). At runtime, a feature vector for a query image is compared against each of the gallery entries by computing a mathematical distance. Finally, computed distances are ranked, and the minimum-distance entry in the gallery is considered to be the most likely match for the given query. This process is illustrated in Fig. 1 for the vehicle re-ID problem, and is used commonly in the literature with only a few exceptions. As a result, re-ID literature relies heavily on novel feature extraction techniques and metric-learning methodologies that aim to develop improved distance measures to yield better ranking.

To the best of our knowledge, the problem of person re-ID was first defined in 2005, where Zajdel *et al.* [4] used a dynamic Bayesian network formulation with handcrafted color features to enable re-ID. Similarly, handcrafted feature extraction methodologies were used extensively in other earlier studies. For instance, in [5], Gheissari *et al.* used spatio-temporal over-segmentation to determine the viewpoint-invariant regions in a given image. Then, color features with structural information were computed and used as re-ID features. In [6], authors introduced Symmetry-Driven Accumulation of Local Features (SDALF). They extracted features by first applying foreground-background segmentation on person bounding boxes. Then, the authors detected salient regions by employing silhouette partitioning on the foreground. Final features were computed using the color and texture features from each salient region. In [7], the authors divided the input image into semantically meaningful parts such as the top, the torso, legs, the left arm and the right arm by applying a detector based on Histogram of Oriented Gradients (HOG). Then, local position, color and gradient features from each semantic region were extracted and concatenated to form the final feature vector. Finally, the authors computed distances between feature vectors by employing a pyramidal matching scheme to improve the attention of the method to small visual cues. In [8] and [9], scale-invariant feature transform (SIFT), and its modification, the speeded-up robust features (SURF) were used to extract features from person bounding boxes. Other popular feature extraction methods that have been proposed to solve

the problem of person re-ID include maximally stable color regions (MSCR) [10], center rectangular ring ratio occurrence descriptor (CRRRO) [11], Schmid and Gabor texture features [12] and Mean Riemannian Covariance Grid (MRCG) [13]. Besides handcrafted properties, learned features have been extensively used recently in person re-ID. Similar methods have been employed also for the vehicle re-ID problem. Thus, the related work utilizing deep learning methods is covered in the context of vehicle re-ID.

Vehicle Re-identification. In recent years, travel safety and efficiency have become increasingly important, thus giving more scientific attention to vehicle re-ID. Especially with the introduction of large-scale vehicle re-ID datasets [14–17], deep learning based methods have become feasible and have obtained state-of-the-art performance. In general, the majority of studies are concerned with modifications of popular CNN architectures and incorporation of auxiliary information into training and testing phases, in order to extract rich and robust feature vectors from vehicle images. For example, in [18], authors have used a MobileNet-based architecture and have reported results for various hard-triplet mining strategies, feature normalization techniques and loss functions. In [19], Liu *et al.* have proposed a multi-branch architecture with various feature pooling schemes on each branch to capture both local and global features and to improve attention to small appearance cues. Similarly, in [20], authors have developed a quadruple-pooling strategy that applies mean-pooling on intermediate feature volumes towards horizontal, vertical, diagonal and anti-diagonal axes and concatenates them to compute rich features. In [21], authors have incorporated auxiliary information to re-ID in the form of spatio-temporal constraints with a sophisticated probabilistic prediction of vehicle routes and travel times.

Various previous studies utilize vehicle viewpoint or pose information, to enhance the performance of vehicle re-ID. For instance, the authors of OIM [22] use auxiliary data to extract orientation-invariant features. For this, Wang *et al.* train a landmark regression network that aims to extract the locations of 20 pre-defined keypoints in a given vehicle image, including the wheels, lamps, headlights, license plates and logos. To train the regression network, the authors annotated the VeRi-776 dataset for these keypoints, as well as for the orientation labels of equally-spaced 8 bins around vehicles. Following the extraction of the keypoints, the OIM architecture uses them to generate orientation-based region proposals and extracts local features for these proposals. Finally, the ultimate feature vector for a given image is computed by aggregating global and local features. In [23], authors propose a 4-branch architecture, where one of the branches is trained to classify the vehicle orientation as well as the vehicle identity. Authors use the viewpoint annotations from [22] and derive a multi-task mutual learning loss function to jointly optimize network weights for identity and viewpoint classification tasks. In [24], Khorramshahi *et al.* utilize detected keypoints in their dual branch network. In their AAVER architecture, one of the branches extracts global features using the ResNet-50 backbone, while the second branch uses the estimated keypoint heatmaps and vehicle orientations to extract local features around keypoints.

In line with the popular research directions in the field of vehicle re-ID, our approach utilizes deep features from our novel branched network architecture, which is enhanced by the auxiliary vehicle viewpoint information. Taking advantage of this additional data, the approach offers efficient training and feature refinement, leading to improved feature extraction and robustness against a large variability of viewpoints.

Methodology

Our viewpoint-aware branched network (VABN) architecture relies on accurate vehicle orientation labels to improve robustness against viewpoint variations. This section first investigates the feasibility of extracting reliable orientation labels from vehicle images, and then discusses the proposed architecture.

A. Orientation Classification

Because of the availability of large-scale classification datasets, a substantial number of strong deep learning architectures with pre-trained weights are at our disposal. In order to perform an up-to-date benchmark of existing classification networks on the task of orientation classification, we selected nine popular architectures of variable complexity: ResNet-18, ResNet-34, ResNet-50, ResNet-101, ResNet-152 [25], ResNeXt-50, ResNeXt-101 [26] and two MnasNet [27] models with depth multipliers of 0.5 and 1.0. The ResNet architecture employs residual connections to alleviate the problem of vanishing gradients and therefore, it enables deeper networks to be properly trained. ResNeXt models are similar to ResNet, but they are constructed by repeating a building block that aggregates a set of transformations with the same topology. On the other hand, MnasNet is a light-weight architecture that is designed by a neural architecture search, which was jointly optimized for high performance and low computational cost. In our experiments, the only modification to the original architectures lies in the size of the final fully-connected layer, which is set to 8 for an 8-bin orientation classification.

B. Viewpoint-Aware Branched Network

To show the applicability of viewpoint embedding for vehicle re-ID, we develop a viewpoint-aware branched deep learning architecture. Fig. 2 provides the general overview of our network.

The architecture follows the popular branched style [19, 22, 28] with a shared backbone for low-level feature generation. Without loss of generality, we use a ResNet-50 backbone where the first three residual blocks are shared among all branches, and the remaining two blocks are replicated for each branch without sharing parameters. In our architecture, each branch is responsible for extracting features from vehicle images with a specific viewpoint. In order to improve robustness against mis-classified viewpoints and to increase the number of training samples per branch, we group the cross-viewpoint classes (left-back, right-front, etc.) to their nearest "front" or "back" viewpoint classes. As a result, we obtain three branches corresponding to each of the viewpoint groups. Similar to the ResNet architecture and following the backbone feature extraction, global average pooling (GAP) is applied to the feature volumes, which collapses the spatial dimensions. The global average pooled features are trained for triplet loss with the batch-hard hard-triplet mining strategy [29],

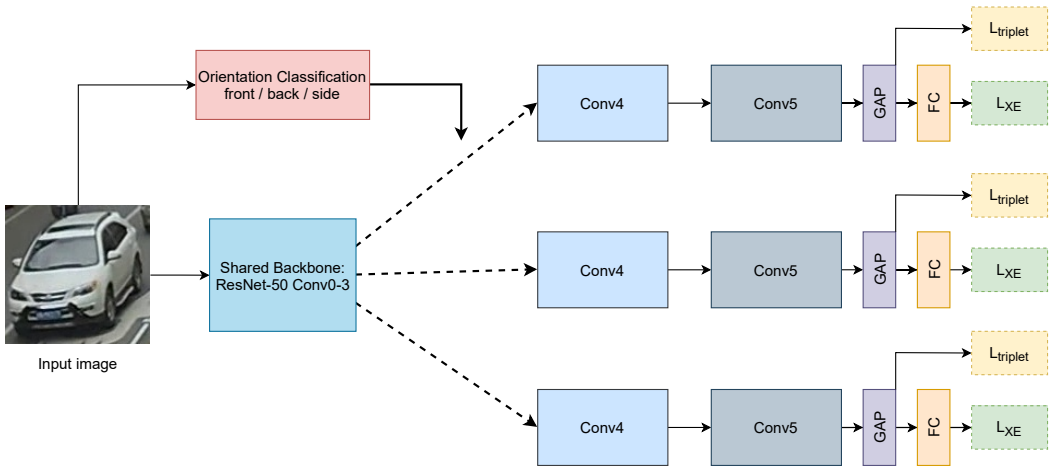


Figure 2. *Viewpoint-Aware Branched Network (VABN): First, vehicle image crops are propagated forward through the parameter-shared ResNet-50 backbone. After this, only one viewpoint-specific branch is selected according to the viewpoint of the input image to generate the final feature vector. The network is trained using the triplet loss ($L_{triplet}$) and the softmax cross-entropy loss (L_{XE}) for identity classification (FC = Fully-connected, GAP = Global average pooling).*

specified mathematically by:

$$\mathscr{L}_{\mathrm{BH}}(\theta;X)=\sum_{i=1}^{P}\sum_{a=1}^{K}\Big[m+\max_{p=1\ldots K}D\left(f_\theta(x_a^i),f_\theta(x_p^i)\right)-\min_{\substack{j=1\ldots P\\ n=1\ldots K\\ j\neq i}}D\left(f_\theta(x_a^i),f_\theta(x_n^j)\right)\Big]_+, \tag{1}$$

where X denotes an input mini-batch, θ represents the learned network weights, P and K are the number of different identities and images per each identity in a mini-batch, m is the triplet loss margin, and D denotes the distance calculation function. Following the GAP, we apply a fully-connected layer that effectively reduces the data size to the number of identities in the training dataset. Resulting vectors are then trained with the softmax cross-entropy loss function for identity classification. During testing, the feature vector for a given vehicle image is obtained from the output of the GAP layer from only one branch, which is the branch that corresponds to the input viewpoint. Our architecture includes branches strongly specialized to extract features from one viewpoint only, which yields more descriptive features. However, there are two main drawbacks of the viewpoint-guided branching. (1) Each branch is trained using only the training samples that have the corresponding viewpoint, effectively reducing the training samples per branch. (2) Randomly sampling each mini-batch leads to an unbalanced number of samples per branch and consequently reduces the efficiency of the batch-normalization layers, leading to sub-optimal training. We employ two training strategies to address both drawbacks by two measures. (A) Use a two-step training strategy, where on top of the ImageNet [30] pre-trained weights, we first train our ResNet-50 backbone on the full dataset with the same per-branch settings of our VABN architecture. (B) Develop a viewpoint-aware mini-batching strategy, to ensure that a sufficient number of samples per branch is available for each training iteration.

In previous studies, selecting suitable triplets during re-ID training was found to be of great importance [29]. Ideally, the constructed triplets should be of medium difficulty, since too easy or too difficult triplets do not contribute to and sometimes harm the overall generalization performance. As a result, various hard-triplet mining strategies emerged to ensure proper selection of triplets. Following a similar reasoning and to allow batch-normalization layers to work properly, we propose to use a viewpoint-aware mini-batching strategy. The proposed strategy is depicted in Fig. 3 for a batch size of 32 images. In the commonly used batch-hard hard-triplet mining strategy [29], a mini-batch is constructed by first randomly picking P identities and then selecting K images from each. Then, for each image in the mini-batch, the network weights are updated only once, considering the hardest positive and negative samples. Here, we extend this idea with viewpoint information. We start by picking randomly, one primary and one secondary viewpoint. Among the eligible identities that include samples from both viewpoints, we randomly select P_v identities. Then, we randomly adopt K_p and K_s samples from each identity, where K_p and K_s are the number of samples having the primary and secondary viewpoints, respectively. This sampling strategy ensures that, (1) there is at least one different viewpoint sample per identity, and (2) for every branch that is trained during a training iteration, the minimum number of samples is at least $min(K_s,K_p)\cdot P_v$, which can be further tuned to ensure proper training.

Experiments

We have conducted our experiments on the VeRi-776 dataset [14, 31], using vehicle orientation annotations from [22]. For the orientation classification experiments, we use an 80-20 training-testing set split on identities. For the re-ID experiments, we employ the readily available, disjoint query and gallery sets of the VeRi-776 dataset and the ResNet-18 generated orientation labels.

A. Discussion: Orientation Classification

We have benchmarked 9 popular models on their classification performance of vehicle orientation and their related computational cost. We have trained the ImageNet pre-trained models for 40 epochs with the Adam optimizer [32] and with a learning

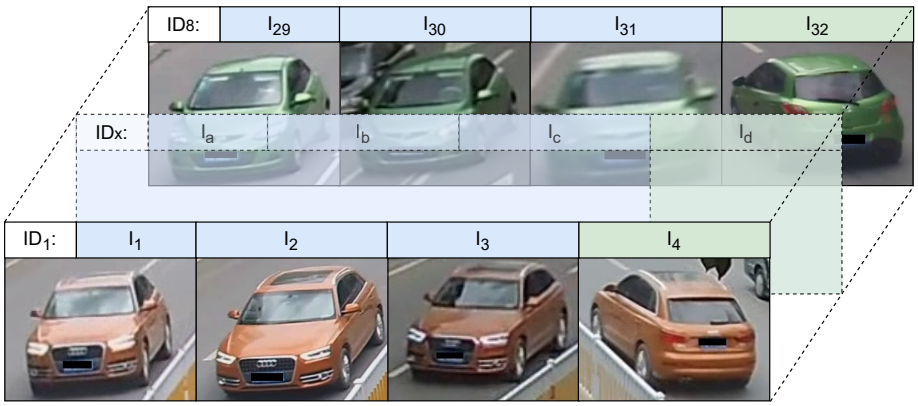


Figure 3. *Illustration of an example viewpoint-aware mini-batch for $P_v = 8$, $K_p = 3$ and $K_s = 1$. Here, the randomly selected primary viewpoint is "front" and secondary is "back". The total batch size for this example is $P_v \cdot (K_p + K_s) = 32$. In this mini-batch, the images with the primary and secondary viewpoints are denoted by blue and green colors. (Color online)*

rate of $2\text{x}10^{-4}$, which is reduced by a factor of ten at every 10^{th} epoch. We have resized the images first to a size of 256×256 pixels, and used random cropping with an output size 224×224. We have also used random erasing data augmentation [33]. The obtained results are summarized in Table 1. It should be noted that the tested models vary greatly in complexity, from 2.2 to 88.7 million parameters. Consequently, we observe a variability of inference time between 13.1 and 82.6 ms executing on a Quadro M1200 GPU.

Examining the results in Table 1, it is possible to conclude that most of the models perform similarly. One outstanding exception is MnasNet with a depth multiplier of 0.5. The accuracy of this model is around 4.5% lower than other architectures, which can be explained by the fact that this model is by far the least complex architecture with only 2.2 million parameters. For the other models, the classification accuracy saturates around 94% and only minor differences in performance occur. Results for the ResNet models clearly show that, even for the smallest ResNet-18 model, the accuracy reaches the saturation point, similar to much more complex models such as ResNet-101 and ResNet-152. Among all models, ResNeXt-50 attains top performance with an overall accuracy of 94.87%. However, compared against ResNet-18, this means an increased accuracy of only 0.27%, at the expense of a significantly higher inference time and GPU memory usage. It should be noted that when examining the common mis-classifications by the benchmarked models, we have encountered mis-labeled samples, which may explain the saturated performance of the models. Overall, the results listed in Table 1 imply that the current attainable performance and computational cost of the orientation classification is sufficient for this data to be used for improving re-ID, even in the case of strict real-time constraints and low-cost hardware.

Table 1: Performance of the benchmarked methods on the VeRi-776 dataset for the orientation classification task. MnasNet depth multipliers are given between parentheses.

Model	# Parameters	Infer. time (ms)	Acc. (%)
ResNet-18	11,689,512	**13.1**	94.60
ResNet-34	21,797,672	23.0	94.80
ResNet-50	25,557,032	29.7	94.46
ResNet-101	44,549,160	56.2	94.68
ResNet-152	60,192,808	82.6	94.58
ResNeXt-50	25,028,904	33.9	**94.87**
ResNeXt-101	88,791,336	74.7	94.25
MnasNet (0.5)	2,220,824	22.1	89.99
MnasNet (1.0)	4,383,312	22.7	94.34

B. Discussion: VABN Performance

We have evaluated the performance of our VABN architecture using the official evaluation protocol of the VeRi-776 dataset. We have used the ImageNet pre-trained ResNet-50 backbone and the 2-step training approach of first training the backbone alone for triplet and cross-entropy losses for vehicle re-ID. We have employed the AMSGrad-based [34] Adam optimizer and trained the model for 125 epochs with an initial learning rate of $2\text{x}10^{-4}$, which was reduced by a factor of ten at epochs 75 and 100. For the viewpoint-aware mini-batching, we have set $P_v = 4$, $K_p = 3$ and $K_s = 1$. During training, we have used random horizontal flipping data augmentation and set the image size to 256×256 pixels. During testing, we have exploited re-ranking with k-reciprocal neighbors [35] with parameters $k_1 = 20$, $k_2 = 6$ and $\lambda = 0.3$. Our experimental results are summarized in Table 2. Table 2 shows that when enabled by the viewpoint-aware mini-batching, our VABN architecture achieves 93.2% rank-1 re-ID score with a mean average precision (mAP) of 71.6%. These results imply a performance increase of 2.8% rank-1 score and 1.3% mAP over the ResNet-50 baseline with re-ranking, and 3.8% rank-1 score and 2.2% mAP without re-ranking. Furthermore, the viewpoint-aware mini-batching strategy improves the performance of our VABN architecture by 4.4% rank-1 and 4.0% mAP. However, on the other hand, the results for VABN without the viewpoint-aware batching are lower compared to the ResNet-50 baseline. We conjecture that this performance loss is due to the sub-optimal training in the case of random batching, which does not guarantee a sufficient amount of samples for the batch-normalization layers in every training iteration. The results suggest that the use of a viewpoint-aware mini-batching structure is an essential requirement to attain a high performance with our VABN architecture.

Table 2: Re-ID performance comparison. From top to bottom, the performance of ResNet-50 baseline, our VABN architecture with 2-step training and our architecture with 2-step training and viewpoint-aware mini-batching strategy.

Method	mAP	Rank-1	Rank-5
ResNet-50 Baseline	63.60	88.00	**95.59**
ResNet-50 Baseline + RR	70.27	90.45	93.21
VABN + 2ST	61.13	86.77	94.99
VABN + 2ST + RR	67.55	88.86	92.19
VABN + 2ST + Sampler	65.82	91.84	95.17
VABN + 2ST + Sampler + RR	**71.56**	**93.21**	94.70

Conclusion

This study presents a vehicle re-ID methodology that takes advantage of the auxiliary viewpoint information to improve the re-ID performance. The first contribution of this study is the feasibility analysis of extracting accurate orientation classification labels from given vehicle image crops. The performed benchmark of nine up-to-date classification networks from literature with respect to their accuracy, GPU memory usage and forward inference time, reveals that it is possible to attain high-quality classification accuracy even with less-demanding architectures such as ResNet-18 and MnasNet (1.0). The required forward inference time demands for a reasonable accuracy is found to be small enough to even allow real-time operation on mobile platforms. Secondly, we present our VABN architecture for the re-ID problem. The VABN architecture utilizes viewpoint information along with raw pixel values of vehicle images to extract rich and descriptive features. Adopting branching according to the viewpoint observed at the input, our network learns to extract highly-specialized, viewpoint-specific features without increasing the forward inference time. Furthermore, the proposed architecture is framework-independent, meaning that it can be applied to any backbone to improve re-ID performance. Finally, we introduce the viewpoint-aware mini-batching strategy. This technique enables efficient training of the viewpoint-aware architecture and ensures that the network is trained with triplets of desired difficulty. If the proposed architecture is combined with the viewpoint-aware mini-batching, the system outperforms the strong ResNet-50 baseline. With respect to real-time operation and efficiency, we conjecture that the results and methodology presented in this study is helpful in bridging the gap between industrial applications and the use of re-ID systems.

Acknowledgment

This research is funded by the European H2020 Interreg PASSAnT Project and Provincial Government of Noord-Brabant, The Netherlands.

References

[1] Milind Naphade, Shuo Wang, David Anastasiu, Zheng Tang, Ming-Ching Chang, Xiaodong Yang, Liang Zheng, Anuj Sharma, Rama Chellappa, and Pranamesh Chakraborty. The 4th ai city challenge, 2020.

[2] Dominik Zapletal and Adam Herout. Vehicle re-identification for automatic video traffic surveillance. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2016.

[3] Alex Krizhevsky, Ilya Sutskever, and Geoffrey Hinton. Imagenet classification with deep convolutional neural networks. *Neural Information Processing Systems*, 25, 01 2012.

[4] Wojciech Zajdel, Zoran Zivkovic, and BJA Krose. Keeping track of humans: Have i seen this person before? In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 2081–2086. IEEE, 2005.

[5] Niloofar Gheissari, Thomas B Sebastian, and Richard Hartley. Person reidentification using spatiotemporal appearance. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'06)*, volume 2, pages 1528–1535. IEEE, 2006.

[6] Loris Bazzani, Marco Cristani, and Vittorio Murino. Symmetry-driven accumulation of local features for human characterization and re-identification. *Computer Vision and Image Understanding*, 117(2):130–144, 2013.

[7] Etienne Corvee, Francois Bremond, Monique Thonnat, et al. Person re-identification using spatial covariance regions of human body parts. In *2010 7th IEEE International Conference on Advanced Video and Signal Based Surveillance*, pages 435–440. IEEE, 2010.

[8] Kai Jüngling, Christoph Bodensteiner, and Michael Arens. Person re-identification in multi-camera networks. In *CVPR 2011 WORKSHOPS*, pages 55–61. IEEE, 2011.

[9] Omar Hamdoun, Fabien Moutarde, Bogdan Stanciulescu, and Bruno Steux. Interest points harvesting in video sequences for efficient person identification. In *The Eighth International Workshop on Visual Surveillance-VS2008*, 2008.

[10] Bingpeng Ma, Yu Su, and Frédéric Jurie. Local descriptors encoded by fisher vectors for person re-identification. In *European Conference on Computer Vision*, pages 413–422. Springer, 2012.

[11] Wei-Shi Zheng, Shaogang Gong, and Tao Xiang. Associating groups of people. In *BMVC*, volume 2, pages 1–11, 2009.

[12] Chunxiao Liu, Shaogang Gong, Chen Change Loy, and Xinggang Lin. Person re-identification: What features are important? In *European Conference on Computer Vision*, pages 391–401. Springer, 2012.

[13] Slawomir Bak, Etienne Corvee, François Bremond, and Monique Thonnat. Multiple-shot human re-identification by mean riemannian covariance grid. In *2011 8th IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS)*, pages 179–184. IEEE, 2011.

[14] X. Liu, W. Liu, H. Ma, and H. Fu. Large-scale vehicle re-identification in urban surveillance videos. In *2016 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6, July 2016.

[15] X. Liu, W. Liu, T. Mei, and H. Ma. Provid: Progressive and multimodal vehicle reidentification for large-scale urban surveillance. *IEEE Transactions on Multimedia*, 20(3):645–658, March 2018.

[16] Milind Naphade, Zheng Tang, Ming-Ching Chang, David C Anastasiu, Anuj Sharma, Rama Chellappa, Shuo Wang, Pranamesh Chakraborty, Tingting Huang, Jenq-Neng Hwang, et al. The 2019 ai city challenge. In *CVPR Workshops*, 2019.

[17] Jakub Sochor, Jakub Spanhel, and Adam Herout. Boxcars: Improving fine-grained recognition of vehicles using 3-d bounding boxes in traffic surveillance. *IEEE Transactions on Intelligent Transportation Systems*, 20(1):97–108, Jan 2019.

[18] Ratnesh Kuma, Edwin Weill, Farzin Aghdasi, and Parthasarathy Sriram. Vehicle re-identification: an efficient baseline using triplet em-

bedding. In *2019 International Joint Conference on Neural Networks (IJCNN)*, pages 1–9. IEEE, 2019.

[19] Xiaobin Liu, Shiliang Zhang, Qingming Huang, and Wen Gao. Ram: a region-aware deep model for vehicle re-identification. In *2018 IEEE International Conference on Multimedia and Expo (ICME)*, pages 1–6. IEEE, 2018.

[20] Jianqing Zhu, Huanqiang Zeng, Jingchang Huang, Shengcai Liao, Zhen Lei, Canhui Cai, and Lixin Zheng. Vehicle re-identification using quadruple directional deep learning features. *IEEE Transactions on Intelligent Transportation Systems*, 21(1):410–420, 2019.

[21] Kai Lv, Weijian Deng, Yunzhong Hou, Heming Du, Hao Sheng, Jianbin Jiao, and Liang Zheng. Vehicle reidentification with the location and time stamp. In *Proc. CVPR Workshops*, 2019.

[22] 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 *Proceedings of the IEEE International Conference on Computer Vision*, pages 379–387, 2017.

[23] Aytac Kanaci, Minxian Li, Shaogang Gong, and Georgia Rajamanoharan. Multi-task mutual learning for vehicle re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

[24] Pirazh Khorramshahi, Amit Kumar, Neehar Peri, Sai Saketh Rambhatla, Jun-Cheng Chen, and Rama Chellappa. A dual-path model with adaptive attention for vehicle re-identification, 2019.

[25] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 770–778, 2016.

[26] Saining Xie, Ross Girshick, Piotr Dollár, Zhuowen Tu, and Kaiming He. Aggregated residual transformations for deep neural networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1492–1500, 2017.

[27] Mingxing Tan, Bo Chen, Ruoming Pang, Vijay Vasudevan, Mark Sandler, Andrew Howard, and Quoc V Le. Mnasnet: Platform-aware neural architecture search for mobile. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2820–2828, 2019.

[28] Hao Chen, Benoit Lagadec, and Francois Bremond. Partition and reunion: A two-branch neural network for vehicle re-identification. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, June 2019.

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

[30] Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei. Imagenet: A large-scale hierarchical image database. In *2009 IEEE conference on computer vision and pattern recognition*, pages 248–255. Ieee, 2009.

[31] Xinchen Liu, Wu Liu, Tao Mei, and Huadong Ma. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In *European conference on computer vision*, pages 869–884. Springer, 2016.

[32] Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization, 2014.

[33] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13001–13008, 2020.

[34] Sashank J Reddi, Satyen Kale, and Sanjiv Kumar. On the convergence of adam and beyond. *arXiv preprint arXiv:1904.09237*, 2019.

[35] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1318–1327, 2017.

Author Biography

Oliver Kocsis holds a Masters degree (2020) from Eindhoven University of Technology, where he conducted research in the Vehicle Re-ID field for his thesis work. Since then he is working at Silicon Labs in the field of Embedded Systems.

Tunc Alkanat received the B.S. and M.S. degrees in electrical and electronics engineering from Middle East Technical University, Ankara, Turkey, in 2013 and 2016, respectively. He is currently working toward the Ph.D. degree at the Video Coding and Architectures Group, Eindhoven University of Technology, Eindhoven, The Netherlands. His research interests include image retrieval, anomaly detection, computer vision for surveillance and computational spectral imaging.

Egor Bondarev obtained his PhD degree in the Computer Science Department at TU/e, in research on performance predictions of real-time component-based systems on multiprocessor architectures. He is an Assistant Professor at the Video Coding and Architectures group, TU/e, focusing on sensor fusion, smart surveillance and 3D reconstruction. He has written and co-authored over 50 publications on real-time computer vision and image/3D processing algorithms. He is involved in large international surveillance projects like APPS and PS-CRIMSON.

Peter H.N. de With is Full Professor of the Video Coding and Architectures group in the Department of Electrical Engineering at Eindhoven University of Technology. He worked at various companies and was active as senior system architect, VP video technology, and business consultant. He is an IEEE Fellow and member of the Royal Holland Society of Academic Sciences and Humanities, has (co-)authored over 600 papers on video coding, analysis, architectures, and 3D processing and has received multiple papers awards. He has served as a program committee member of various IEEE conferences and holds some 30 patents.

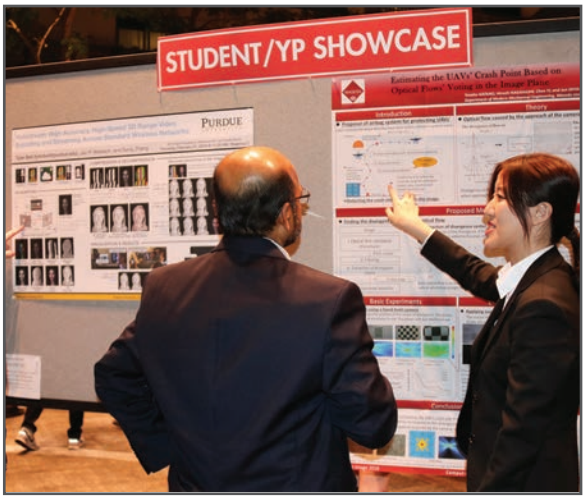
JOIN US AT THE NEXT EI!

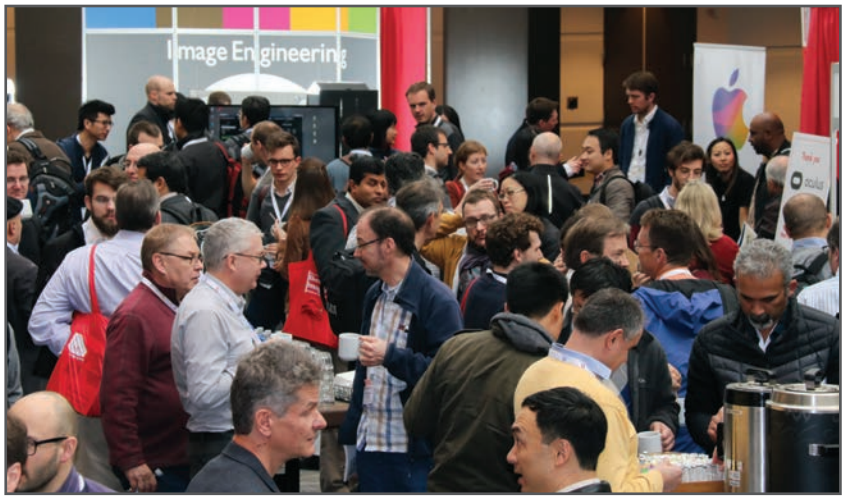
IS&T International Symposium on

Electronic Imaging

SCIENCE AND TECHNOLOGY

Imaging across applications . . . Where industry and academia meet!





• SHORT COURSES • EXHIBITS • DEMONSTRATION SESSION • PLENARY TALKS •
• INTERACTIVE PAPER SESSION • SPECIAL EVENTS • TECHNICAL SESSIONS •

www.electronicimaging.org

IS&T
imaging.org