Towards Parameter-Optimized Vessel Re-identification based on IORnet

Amir Ghahremani, Yitian Kong, Egor Bondarev, Peter H.N. de With

Eindhoven University of Technology, Eindhoven, The Netherlands
Video Coding and Architectures Group (VCA)
a.ghahremani@tue.nl

Abstract. Reliable vessel re-identification would enable maritime surveillance systems to analyze the behavior of vessels by drawing their accurate trajectories, when they pass along different camera locations. However, challenging outdoor conditions and varying viewpoint appearances combined with the large size of vessels limit conventional methods to obtain robust re-identification performance. This paper employs CNNs to address these challenges. In this paper, we propose an Identity Oriented Re-identification network (IORnet), which improves the triplet method with a new identity-oriented loss function. The resulting method increases the feature vector similarities between vessel samples belonging to the same vessel identity. Our experimental results reveal that the proposed method achieves 81.5% and 91.2% on mAP and Rank1 scores, respectively. Additionally, we report experimental results with data augmentation and hyper-parameters optimization to facilitate reliable ship re-identification. Finally, we provide our real-world vessel re-identification dataset with various annotated multi-class features to public access.

Keywords: re-identification of vessels · CNNs · maritime surveillance · vessel re-identification dataset.

1 Introduction

Camera-based maritime surveillance systems monitor harbors and waterways to increase the safety and security against unknown pathless watercrafts, prevent out-of-region fishery, manage urban transportation, and control the cargo flow. Recently, vessel-behavior analysis is also an expected function for such systems, since this ability can drastically improve the efficiency of an automated surveillance system. To this end, keeping track of vessels over consecutive camera locations is of a vital importance. This requires a reliable vessel re-identification approach, which aims at the successful detection of the identity of a specific vessel at different camera locations. This concept is visualized in Fig. 1 with image samples captured by different cameras.

The outdoor maritime environment poses considerable challenges to camera-based surveillance systems by precipitation, sunshine reflection, fog, water waves, etc. In addition to these typical problems, a vessel re-identification method has

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_10

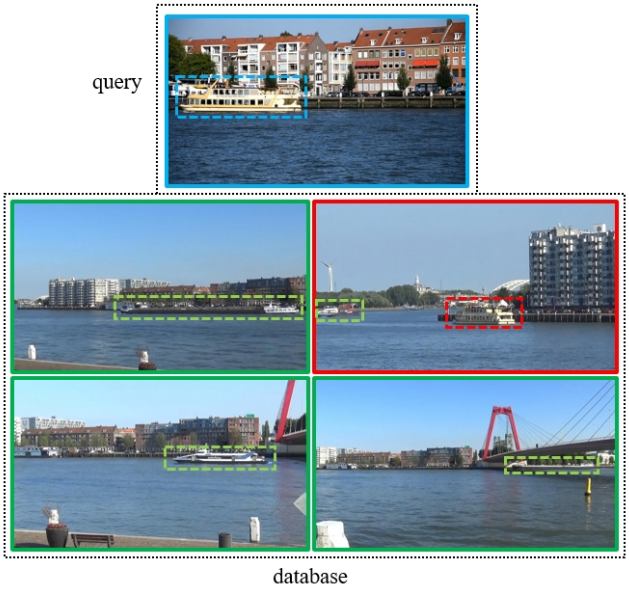


Fig. 1: Vessel re-identification is about finding the query vessel in the existing database images. In this example, the blue box in the top image represents the query image. The red box in the upper-right database image indicates the same vessel re-identified at a clearly different location.

to overcome its task-specific challenges. For instance, surveillance cameras at different locations often capture vessels from varying viewpoints. Since vessels are large objects, their appearances (including color, shape, hull textures, etc.) can be entirely different from alternative viewpoints. Moreover, vessels captured at different camera locations are surrounded by diverse types of backgrounds. Furthermore, illumination changes caused by different weather conditions and daytimes also deteriorate the vessel re-identification performance.

To the best of our knowledge, an in-depth study on the vessel re-identification problem is virtually absent in literature. However, with the emergence of Convolutional Neural Networks (CNNs), the related field of pedestrian re-identification already presents methods with promising performance [1–6]. Since vessel re-identification is conceptually similar to pedestrian re-identification, this paper addresses the vessel re-identification problem by extending a triplet-based pedestrian re-identification approach to the maritime surveillance domain. However, unlike pedestrian re-identification, in a maritime environment vessels of the same model but different identity, may still have extremely similar appearance, which makes the vessel re-identification even more challenging. Fig. 2 illustrates this problem by a few example vessel images.

In this paper, first, we attempt to solve vessel re-identification by introducing a new identity-oriented loss function for learning the vessel identity. Second,

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_10

Fig. 2: Illustration of four images of similar vessels, which belong to different vessel identities, although they are made by the same vessel manufacturing company and have the same model.

since there is no public dataset available for exploring the vessel re-identification problem, we provide our annotated vessel re-identification dataset, which was captured at various locations in several harbor cities and suburbs in the Netherlands (Amsterdam, Rotterdam), to open public access [7]. Third, this paper investigates the efficiency of high-performing human re-identification techniques for the vessel re-identification problem (e.g. data augmentation, a different number of training iterations, etc.). These experiments lead to a parameter-optimized re-identification of vessels.

The sequel of the paper is organized as follows. Section 2 provides an overview of the related work. Section 3 explains the proposed method. Section 4 presents the experimental results and validation. Section 5 concludes the paper.

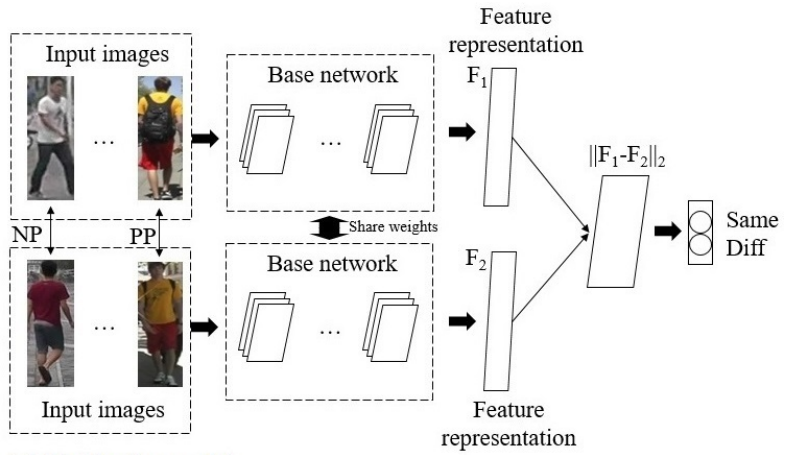
2 Related Work

As already mentioned, due to the absence of literature on vessel re-identification, we commence with addressing the widely investigated pedestrian re-identification from several research works.

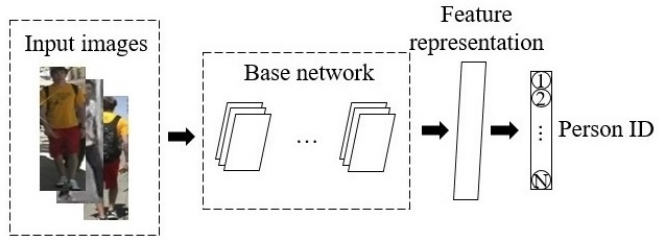
Pedestrian re-identification approaches attempt to re-identify the same person at different camera locations. These methods typically search for the best match of a query image among previously captured database images. Two common research directions for such methods are (1) to attempt to improve the image discrimination in feature space and/or (2) introduce better distance metrics [8]. Generally, pedestrian re-identification methods are divided into three main categories: (a) verification models [1, 2, 9, 10], (b) identification models [5, 6, 11–13], and (c) combinational models [8, 14, 15], which are all briefly discussed below.

Fig. 3(a) illustrates a common architecture of verification models. These models re-identify the vessel samples belonging to the same identity by assessing the feature vector similarities between their input images. The work in [1] employs a patch-matching technique, which finds the mid-level feature similarity of pairwise images, to modify a Siamese network. In [9], the method uses matching

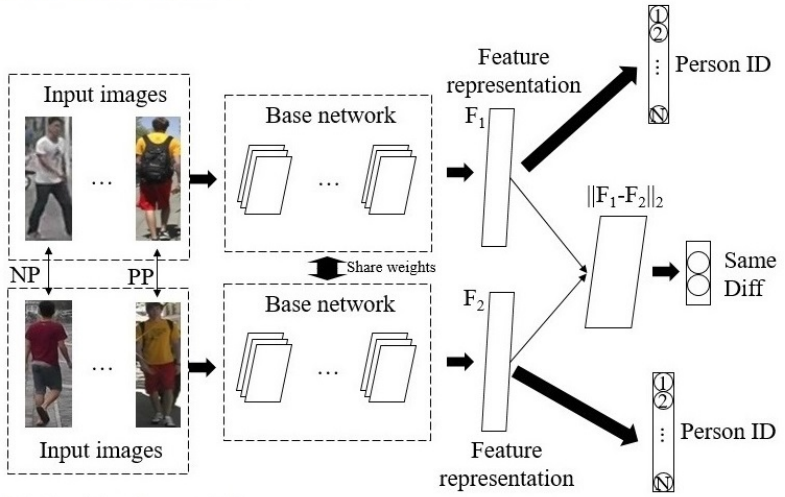
ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_10



(a) Verification model.



(b) Identification model.



(c) Combination model.

Fig. 3: Architecture of three generic pedestrian re-identification models. The numbers in the circles at the output represent different identities, where N is the total number of person identities in the database. Additionally, NP and PP represent negative and positive pairs, respectively.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_10

gates to improve a Siamese network. These gates predict the critical point of a Siamese network in higher layers by inspecting its low-level features.

A common architecture of identification models is presented in Fig. 3(b). These methods investigate a single input image to determine the person's identity. The work in [12] engages handcrafted features in a network for fine-tuning a procedure to improve the re-identification performance. The method in [6] achieves better fine-tuning performance by employing a pedestrian-attribute dataset. This work uses the data disparity between practical datasets (which have low quality) and the ImageNet [13] dataset. The work from [11] proposes to use a reliable classification model, which is obtained by combining several pedestrian re-identification datasets for person identity recognition.

Fig. 3(c) presents the architecture of a typical combinational model. These architectures incorporate both the identification and verification loss-functions to optimize the performance. In [8], the proposed method improves a Siamese architecture by comparing the feature representations of input images using a square layer. The work in [14] combines two identification subnets and one verification subnet with a Siamese network to provide robust pedestrian re-identification performance.

This paper modifies the pedestrian re-identification concept towards the vessel re-identification problem. Additionally, we introduce a new triplet-based loss function to increase the feature vector similarity between vessels belonging to the same vessel identity. Here, we focus on re-identifying vessels in harbors and waterways. Additionally, we provide an annotated vessel re-identification dataset, which includes 4,616 real-world images. These images were captured with two cameras under different weather, lighting, and timing conditions and at different locations with variable backgrounds. All annotated vessels have been labeled by a unique ID and appear in several images. Moreover, we have also annotated the bounding box, vessel type and vessel orientation of each vessel for potential further experiments.

3 Architecture Pipeline

This section describes the proposed vessel re-identification architecture, which is depicted in Fig. 4. The visualized method (IORnet) includes three modules. The first one is the feature extraction module, which receives a set of three images per vessel identity and transforms them into the feature space. The second module is the triplet subnet. This module calculates the triplet loss of the input, aiming to pull samples closer when they originate from the same vessel, while increasing the distance to different vessels/objects. The third module is the identification subnet, which increases the feature vector similarity between vessels belonging to the same vessel identity. After extracting the feature vectors from the first module, the second and third module operate in parallel to calculate the loss function. Then, the base-network weights are updated according to the calculated loss value. The following subsections discuss the three modules in detail.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_10

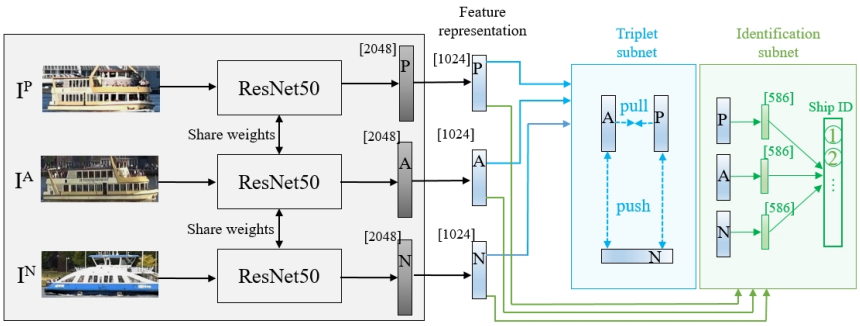


Fig. 4: Identity-Oriented Re-identification network (IORnet). This method receives three input images, which are Anchor image together with its Positive and Negative pairs. Then, the triplet and identification modules calculate the loss function and update the feature extraction CNNs.

3.1 Feature Extraction Module

This module consists of three basic CNNs to transform the input images into feature vectors. Here we employ ResNet50 [16] as the basis architecture. The extracted feature vectors have a dimension of 2,048 elements. Then, we append a batch normalization layer to speed up the convergence and optimize the deep convolution networks [17]. Finally, this module resizes the feature vectors to 1,024 elements. The three input images submitted to the basis CNNs are denoted by IA, IP, and IN, while IA and IP belong to the same vessel identity (positive pair) and IN represents another vessel identity (negative sample).

3.2 Triplet Module

Here, the conventional triplet model and its limitations are first briefly reviewed. The triplet loss was introduced in [18] to improve face re-identification performance. The objective of the triplet loss is to pull image features belonging to the same class closer to each other, while pushing the features of different image classes away from that cluster. For more clarification, we assume that A, P, and N denote the Anchor, Positive, and Negative image samples, while A and P contain the same object identity and N contains another identity. The triplet loss is then expressed by:

$$D_{AP} - D_{AN} \geq \alpha, \tag{1}$$

where D_{XY} represents the distance between images X and Y in feature space and α is the distance margin. By iteratively optimizing this process over the whole dataset, positive pairs converge into a single cluster, while distantiating that cluster from the negative samples. In this work, we employ the TriNet [4] framework as the triplet subnet. This method uses a variant of the triplet loss to

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_10

perform end-to-end deep metric learning and achieves reliable results in pedestrian re-identification.

Unfortunately and as a limitation, the triplet architecture only considers two different identities at a time. This can push a negative object sample against its cluster [19], which can lead to having dissimilar feature representations for vessel samples belonging to the same vessel identity. This drawback also increases the convergence time.

3.3 Identification Module

As just discussed, the triplet loss function may generate dissimilar feature vectors for object samples belonging to the same object identity. In order to solve this problem, we propose an Identity Oriented Re-identification network (IORnet). In IORnet, we add an identification subnet to the triplet network, as illustrated in Fig. 4. In this subnet, we consider all the samples belonging to the same identity as a unique label and perform multi-class detection learning. To this end, the feature representations extracted by the basis networks are supplied into a new fully-connected layer. Then, the softmax function is used to normalize these feature vectors. By adding the identification subnet, the final loss function can be formulated now as follows:

$$L = \gamma.L^{\text{triplet}} + (1-\gamma) \cdot L^{\text{identification}}, \tag{2}$$

where

$$L^{\text{triplet}} = \alpha + D_{AP} - D_{AN},$$

and $L^{\text{identification}}$ represents the softmax loss function. Trade-off parameter γ is defined in the unity interval. With $\gamma = 0$, the final loss becomes the identification loss function, while $\gamma = 1$ changes the equation into the pure triplet loss function. This proposed loss function restricts the whole system to provide more similar feature representations for the image samples belonging to the same vessel identity.

4 Empirical Validation

This section starts with a dataset overview and then discusses the training parameters and analyzes the performance of the proposed method.

4.1 Vessel Re-Identification Dataset

In order to train the vessel re-identification model, we have recorded several videos from various locations in the Netherlands. These videos were captured using two cameras during different daytimes. The videos contain a vast variety of different viewpoints on vessels. Additionally, several vessel types with divergent sizes and distances to the camera are found in this dataset. Finally, challenging scenarios including vessel occlusion/truncation are also annotated.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_10

The dataset contains 4,616 images with 733 different vessel identities. Each vessel identity is represented by several images. Additionally, we have labeled each vessel with a bounding box, its vessel type, and vessel orientation (i.e. the approximate positioning angle towards the camera) to facilitate future research. The vessel type range contains 10 classes: sailing vessel, container ship, passenger ship, fishing vessel, tanker, river cargo, small boat, yacht, tug, and taxi vessel. The vessel orientations are described with the following 5 orientation labels: front view, front-side view, side view, back-side view, and back view. Besides this, we have provided a unique ID to each vessel and have cropped each vessel from the whole image by an annotated bounding box. Then the dataset is split into training and test datasets. The training dataset contains 3,651 images with 586 unique vessel identities, while the test dataset includes 965 images with 147 unique vessel identities.

4.2 Training Procedure

In this work, we have employed ResNet50 [16] pre-trained on ImageNet [13] as the basis network for feature extraction. The Adam optimizer [20] is used with default hyper-parameters. We have set the initial learning rate to 0.0003, which is exponentially decayed after 35,000 iterations. The proposed re-identification CNN is trained for 50,000 iterations. We have selected 18 vessel identities and 4 images per identity to form a mini-batch of size 72. Furthermore, we have added a dropout layer [21] to reduce the risk of overfitting. The trade-off parameter γ in Equation (2) is empirically set to $\gamma = 0.6$.

4.3 Validation Results

This subsection evaluates our Identity Oriented Re-identification network (IORnet). We have trained both the state-of-the-art triplet-based network (TriNet) and the IORnet on the published training dataset. Table 1 compares the performance of these two methods on our test dataset. In this table, the methods are evaluated according to mAP (mean Average Precision), Rank1, Rank5, and Rank10 metrics. According to Table 1, TriNet provides mAP and Rank1 scores of 78.4% and 88.4%, respectively. These values indicate that extending the triplet concept to the vessel re-identification problem by training this network on an annotated vessel re-identification dataset provides robust results. Additionally, IORnet improves the mAP and Rank1 results of TriNet by approximately 3% and 2%, respectively. Evidently, the proposed loss function provides a higher performance due to increasing the similarities between feature vectors belonging to the same identity.

4.4 Discussion on Parameters and Data Augmentation

In order to explain the parameter-optimized vessel re-identification model in depth, this subsection provides additional discussion on the proposed method.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_10

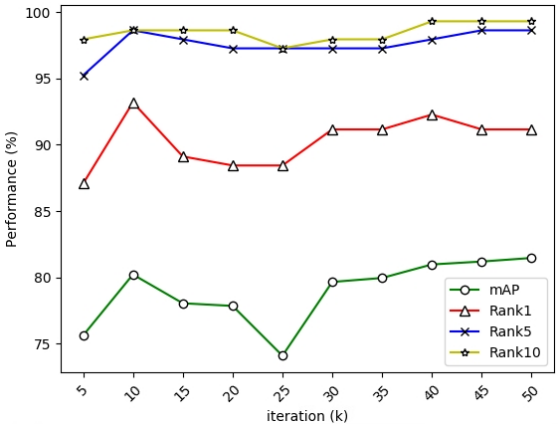


Fig. 5: Impact of training iterations on IORnet performance.

Fig. 5 illustrates the mAP scores provided by IORnet for different iterations. It can be observed that the mAP results are not stable within the first 35,000 iterations. This occurs because the learning rate is relatively large and the network skips some local optimizations. After 35,000 iterations (while decreasing the learning rate), the calculated scores tend to become more stable and also slightly improve.

Generally, data augmentation improves the performance of pedestrian re-identification methods. Therefore, we have tested this technique also on the vessel re-identification problem. To this end, our vessel images are augmented with random cropping and horizontal flipping in the training phase, similar to pedestrian re-identification work in [1, 17]. More specifically, the image size is first increased by 9/8 with the same aspect ratio. Then, the image is randomly copped to obtain the original size. We have also performed the horizontal image flipping on randomly chosen images. Fig. 6 compares the method performance with and without data augmentation. It apppears that the data augmentation technique deteriorates the vessel re-identification performance. For instance, the mAP rate is decreased from 81.46% to 76.68%. We conclude that training on random parts of vessels does not improve the re-identification model, since for large objects, random fragments do not provide a reliable statistical base for identity retrieval.

Table 1: Vessel re-identification performance.

Models	mAP	Rank1	Rank5	Rank10
TriNet	78.4%	88.4%	97.3%	98.6%
IORnet	81.5%	91.2%	98.6%	99.3%

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_10

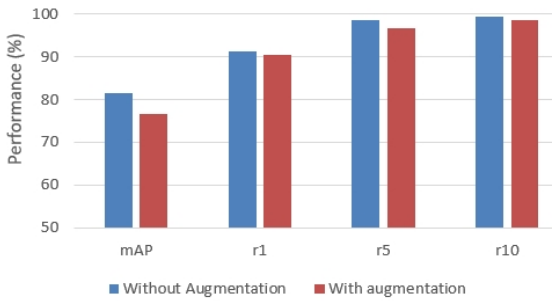


Fig. 6: Data augmentation influence on vessel re-identification performance.

Additionally, during our experiments, we have noticed that many failures in vessel re-identification are the outcome of performing re-identification on vessel samples captured from different orientations (camera viewing angle to the ship). This happens also because vessels are large objects, having very different appearances from varying orientations. To address this problem, the orientation information can be integrated into the re-identification method in our future work.

As mentioned earlier, the parameter γ was introduced to control the trade-off between the identification module and the triplet module. Tuning this parameter to the optimal value is of high importance. Here, the re-identification performance is tested with different values of the parameter to discover its influence on the perfromance. The results are illustrated in Fig. 7. It can be deduced that for $\gamma < 0.36$ and $\gamma > 0.6$ the re-identification performance deteriorates. For this reason, we have chosen $\gamma = 0.6$, since this value achieves the highest mAP and Rank1 scores.

Finally, to pursue real-time vessel re-identification, we have calculated the

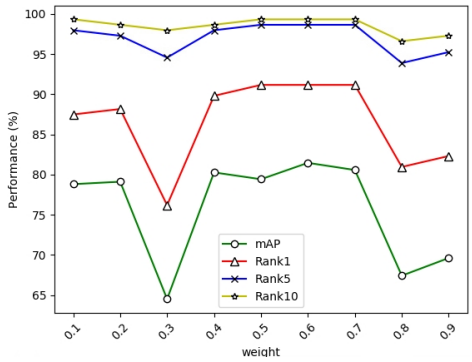


Fig. 7: Influence of γ parameter on IORnet performance.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_10

average time of identity retrieval. The tests were performed on a workstation with E5-1620 CPU, 16 GB of memory and a GTX-1080 GPU. There are 147 vessel identities in our query set and 818 images in the database. The total retrieval time measured for all 147 query images was 558.7 ms. Hence, for a single query image, it takes 3.8 ms to perform the re-identification procedure and return the ranking list, which would satisfy real-time execution.

It is also important to mention that the original TriNet decreases the feature vector size from 2,048 to 128. However, according to our experiments, this small feature vector size does not allow the identification module to achieve the desired results. Therefore, we have adopted 1,024 as the feature vector size at the output of the feature extraction module. By doing so, the training time increases from 0.3 sec/iteration to 0.56.

5 Conclusion

This paper has proposed a robust vessel re-identification method to track the identity of a specific vessel throughout a network with different camera locations. The proposed CNN-based method extends the TriNet loss function with an identification method. The improved architecture, called IORnet, concentrates on enhancing the similarities between vessel images belonging to the same vessel identity in feature space. This approach also leads to a better discrimination from other vessels. Experimental results have shown that our approach achieves 81.5% and 91.2% on mAP and Rank1 scores, respectively. Additionally, experiments were conducted for vessel re-identification using several re-identification techniques with proven value for pedestrian re-identification (like data augmentation, training parameters, etc.). This supplementary inspection has resulted into a parameter-optimized re-identification of vessels. As an important contribution, we have also developed a vessel re-identification dataset, which is annotated with bounding boxes, vessel identities, vessel categories, vessel orientations, and vessel capturing status (whether a vessel is truncated and/or occluded). This dataset includes images with annotated vessels, captured at different locations under varying weather conditions and with variable backgrounds and has become available for public access [7].

6 Acknowledgement

The authors gratefully acknowledge the project PASSANT, funded by the H2020 Interreg program, for supporting the research work.

References

1. Ejaz Ahmed, Michael Jones, and Tim K Marks: "An improved deep learning architecture for person re-identification," *CVPR*, pages 39083916, 2015.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_10

2. Shengyong Ding, Liang Lin, Guangrun Wang, and Hongyang Chao: "Deep feature learning with relative distance comparison for person re-identification," *Pattern Recognition*, 48(10):29933003, 2015.
3. Carlos Guindel, David Martin, and Jose Maria Armingol: "Joint object detection and viewpoint estimation using cnn features," *ICVES*, pages 145150. IEEE, 2017.
4. Alexander Hermans, Lucas Beyer, and Bastian Leibe: "In defense of the triplet loss for person re-identification," arXiv preprint arXiv:1703.07737, 2017.
5. Yutian Lin, Liang Zheng, Zhedong Zheng, Yu Wu, and Yi Yang: "improving person re-identification by attribute and identity learning," arXiv preprint arXiv:1703.07220, 2017.
6. Tetsu Matsukawa and Einoshin Suzuki: "Person re-identification using cnn features learned from combination of attributes," *ICPR*, pages 24282433. IEEE, 2016.
7. http://vca.ele.tue.nl/ (active from March 2019).
8. Zhedong Zheng, Liang Zheng, and Yi Yang: "A discriminatively learned cnn embedding for person reidentification," *TOMM*, 14(1):13, 2017.
9. Rahul Rama Varior, Mrinal Haloi, and Gang Wang: "Gated Siamese convolutional neural network architecture for human re-identification," *ECCV*, pages 791808. Springer, 2016.
10. De Cheng, Yihong Gong, Sanping Zhou, Jinjun Wang, and Nanning Zheng: "Person re-identification by multi-channel parts-based cnn with improved triplet loss function," *CVPR*, pages 13351344, 2016.
11. Tong Xiao, Hongsheng Li, Wanli Ouyang, and Xiaogang Wang: "Learning deep feature representations with domain guided dropout for person re-identification," *CVPR*, pages 12491258. IEEE, 2016.
12. Shangxuan Wu, Ying-Cong Chen, Xiang Li, An-Cong Wu, Jin-Jie You, and Wei-Shi Zheng: "An enhanced deep feature representation for person re-identification," *WACV*, pages 18., 2016.
13. Jia Deng, Wei Dong, Richard Socher, Li-Jia Li, Kai Li, and Li Fei-Fei.: "Imagenet: A large-scale hierarchical image database," *CVPR*, pages 248255. IEEE, 2009.
14. Mengyue Geng, Yaowei Wang, Tao Xiang, and Yonghong Tian: "Deep transfer learning for person re-identification," arXiv preprint arXiv:1611.05244, 2016.
15. Weihua Chen, Xiaotang Chen, Jianguo Zhang, and Kaiqi Huang: "A multi-task deep network for person re-identification," *AAAI*, volume 1, page 3, 2017.
16. Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun: "Deep residual learning for image recognition," *CVPR*, pages 770778, 2016.
17. Sergey Ioffe and Christian Szegedy: "Batch normalization: Accelerating deep network training by reducing internal covariate shift," arXiv preprint arXiv:1502.03167, 2015.
18. Florian Schroff, Dmitry Kalenichenko, and James Philbin: "Facenet: A unified embedding for face recognition and clustering," *CVPR*, pages 815823, 2015.
19. Yiheng Zhang, Dong Liu, and Zheng-Jun Zha.: "Improving triplet-wise training of convolutional neural network for vehicle re-identification," *ICME*, pages 13861391. 2017.
20. Diederik P Kingma and Jimmy Ba. Adam: "A method for stochastic optimization," arXiv preprint arXiv: 1412.6980, 2014.
21. Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov: "Dropout: a simple way to prevent neural networks from overfitting," *The Journal of Machine Learning Research*, 15(1):19291958, 2014.

ICCS Camera Ready Version 2019
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-22750-0_10