SCIENCE CHINA
Information Sciences

• LETTER •

July 2021, Vol. 64 179104:1–179104:2
https://doi.org/10.1007/s11432-020-3004-7

A spatial structural similarity triplet loss for auxiliary vehicle re-identification

Jianqing ZHU[1], Liu LIU[2], Xiaobin ZHU[3] & Huanqiang ZENG[4]*

[1]*College of Engineering, Huaqiao University, Quanzhou* 362021*, China;*
[2]*UBTECH Sydney AI Centre, School of Computer Science FEIT, University of Sydney, Sydney* 2008*, Australia;*
[3]*Department of Computer Science and Technology, University of Science and Technology Beijing, Beijing* 100083*, China;*
[4]*College of Information Science and Engineering, Huaqiao University, Xiamen* 361021*, China*

Received 6 May 2020/Revised 17 July 2020/Accepted 24 July 2020/Published online 7 December 2020

Citation Zhu J Q, Liu L, Zhu X B, et al. A spatial structural similarity triplet loss for auxiliary vehicle re-identification. Sci China Inf Sci, 2021, 64(7): 179104, https://doi.org/10.1007/s11432-020-3004-7

Dear editor,

Vehicle re-identification aiming to match vehicle images captured by different cameras plays an essential role in video surveillance systems for public security. However, vehicle re-identification is a challenging computer vision problem because vehicle images usually contain a list of adverse factors, such as viewpoint variations, blurs, and occlusions. Therefore, how to design an effective vehicle re-identification method has received increasing attention.

The commonly-used spatial global pooling layer is beneficial to learning viewpoint robust features for vehicle re-identification. However, the spatial global pooling layer compresses feature maps by simply calculating spatial statistics (i.e., maximum or average), restricting the spatial discrimination of feature maps. As a result, the vehicle re-identification performance is inevitably weakened.

To reserve the spatial discrimination of feature maps, the most straightforward method is to uniformly divide feature maps into several parts and then individually pool each part, as done in [1–5]. However, this method is prone to suffering from part dis-alignments since it divides parts rigidly. Although we can use part detectors to determine parts automatically, high cost is required, such as part annotations and detector executions. Therefore, how to well reserve feature maps' spatial discrimination is still an open problem for improving vehicle re-identification.

Method. As shown in Figure 1, in addition to the Euclidean distance (ED) based triplet loss and a label smooth regularized softmax (LSRS) loss, the proposed spatial structural similarity triplet loss auxiliary deep network (S^{3}ANet) applies a newly designed spatial structural similarity (S^3) triplet loss for improving vehicle re-identification.

The structural similarity (SSIM) [6] originally is used to measure similarities of images, since it can comprehensively measure the differences of luminance, contrast, and structure information. Inspired by this, the SSIM is applied to construct the S^3 triplet loss. To be more specific, the S^3 triplet loss function is formulated as

$$\begin{aligned} &L_{\mathrm{S}^3}(X_a^c, X_n^c, X_p^c) \\ &= -\log\left(1 + \mathrm{e}^{\mathrm{SSIM}(X_a^c, X_p^c) - \mathrm{SSIM}(X_a^c, X_n^c)}\right), \end{aligned} \tag{1}$$

where (X_a^c, X_n^c, X_p^c) is a training triple; (X_a^c, X_n^c) is a negative pair (i.e., two vehicle subjects of different class labels), and (X_a^c, X_p^c) is a positive pair (i.e., two vehicle subjects of the same class label); SSIM is the calculation of SSIM [6]. Furthermore, for reducing SSIM's computational complexities, each sample of a training triple is represented with a height × width × channel $= h \times w \times 1$ sized feature map. As shown in Figure 1, the $h \times w \times 1$ sized feature map is extracted by the channel global average pooling (CGAP) layer after the backbone network (i.e., ResNet-50-IBN-a [7]). The S^3 triplet loss is minimized via a stochastic gradient descent algorithm to push the SSIMs of the same subjects as large as possible and pull the SSIMs of different subjects as small as possible. Consequently, the feature maps' spatial discrimination is reserved. Appendix A provides more details of the proposed method.

Experiment. Our study presents performance comparisons between the proposed S^{3}ANet and state-of-the-art approaches (e.g., [1–4]) on VeRi776 [8] and VehicleID [9] databases. Among all the compared methods, the proposed S^{3}ANet obtains the best performance.

Keep the network architecture unchanged, the proposed S^{3}ANet with the S^3 triplet loss is superior to the one without the S^3 triplet loss on VeRi776 and VehicleID databases. For example, on the largest Test2400 subset of VehicleID, the mAP and rank-1 identification rate of the proposed S^{3}ANet with the S^3 triplet loss are individually 1.61% and 1.43% higher than those of the one without the S^3 triplet loss. All these demonstrate that the proposed S^3 triplet loss that reserves the spatial discrimination of feature maps is indeed beneficial to improving the vehicle re-identification performance.

* Corresponding author (email: zeng0043@hqu.edu.cn)

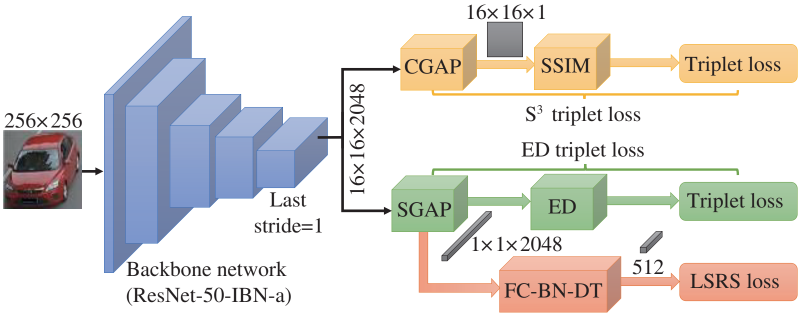

Figure 1 (Color online) The framework of the spatial structural similarity triplet loss auxiliary deep network (S^{3}ANet) for vehicle re-identification. Here, SGAP represents a spatial global average pooling layer; FC-BN-DT represents the composite of fully connected, batch normalization, and dropout layers.

Finally, it is interesting to investigate the impact of the performance using different feature map pooling strategies. For that, based on the same backbone network, two spatial division pooling strategies, i.e., vertically four quartering (V4) and horizontally four quartering (H4), by following a similar practice in [1,3,4] are implemented and compared. The proposed S^{3}ANet outperforms both V4 and H4. These results illustrate that our S^3 triplet loss auxiliary method is more effective than spatial division pooling strategies. Appendix B presents more performance comparison results.

Conclusion. This study presents a novel spatial structural similarity (S^3) triplet loss for auxiliary vehicle re-identification. Rather than tedious dividing feature maps, our method elegantly reserves the spatial discrimination of feature maps via minimizing the S^3 triplet loss. Experiment results show that our method is superior to state-of-the-art approaches.

Acknowledgements This work was supported in part by National Natural Science Foundation of China (Grant Nos. 61976098, 61871434), in part by National Key R&D Program of China (Grant No. 2018YFB0803700), in part by Natural Science Foundation of Fujian Province (Grant No. 2018J01090), in part by Key Science and Technology Project of Xiamen City (Grant No. 3502ZCQ20191005), in part by Science and Technology Bureau of Quanzhou (Grant No. 2018C115R), and in part by China Scholarship Council.

Supporting information Appendixes A and B. The supporting information is available online at info.scichina.com and link.springer.com. The supporting materials are published as submitted, without typesetting or editing. The responsibility for scientific accuracy and content remains entirely with the authors.

References

1 Chen H, Lagadec B, Bremond F. Partition and reunion: a two-branch neural network for vehicle re-identification. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, 2019. 184–192

2 Liu C H, Huynh D Q, Reynolds M. Urban area vehicle re-identification with self-attention stair feature fusion and temporal Bayesian re-ranking. In: Proceedings of International Joint Conference on Neural Networks, Budapest, 2019

3 Liu X B, Zhang S L, Huang Q M, et al. RAM: a region-aware deep model for vehicle re-identification. In: Proceedings of IEEE International Conference on Multimedia and Expo, Seattle, 2018

4 Zhu J Q, Zeng H Q, Huang J C, et al. Vehicle re-identification using quadruple directional deep learning features. IEEE Trans Intell Transp Syst, 2020, 21: 410–420

5 Zhu J Q, Zeng H Q, Jin X, et al. Joint horizontal and vertical deep learning feature for vehicle re-identification. Sci China Inf Sci, 2019, 62: 199101

6 Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity. IEEE Trans Image Process, 2004, 13: 600–612

7 Pan X G, Luo P, Shi J P, et al. Two at once: enhancing learning and generalization capacities via IBN-net. In: Proceedings of European Conference on Computer Vision, Munich, 2018. 464–479

8 Liu X C, Liu W, Mei T, et al. A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In: Proceedings of European Conference on Computer Vision, Amsterdam, 2016. 869–884

9 Liu H Y, Tian Y H, Wang Y W, et al. Deep relative distance learning: tell the difference between similar vehicles. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, 2016. 2167–2175