SCIENCE CHINA
Information Sciences

• **Supplementary File** •

Joint Horizontal and Vertical Deep Learning Feature for Vehicle Re-identification

Jianqing ZHU[1], Huanqiang ZENG[2*], Xin JIN[3*], Yongzhao DU[1], Lixin ZHENG[1] & Canhui CAI[1]

[1]*College of Engineering, Huaqiao University, 362021, Quanzhou* 362021*, China;*
[2]*College of Information Science and Engineering, Huaqiao University, Xiamen,* 361021*, China;*
[3]*Department of Computer Science and Technology, Beijing Electronic Science and Technology Institute, Beijing* 100070*, China*

Appendix .1 The detail parameter configuration of the proposed method

Table 1 lists the detail parameter configuration of the proposed method. The channel numbers of Conv0, SDU1, SDU2, SDU3, SDU4 and SDU5 are 64, 64, 128, 192, 256 and 320, respectively. The scope of Leaky ReLU layer of SDU5 is 0, and that of the others are 0.15. The sub-window for Conv0 and SDU represents a filter size, and that for pooling layers (i.e., MP1-MP5, HAP and VAP) means a pooling window size, and that for two spatial normalization (i.e., SN1 and SN2) layers denotes a normalization window size. As shown in Table 1, Conv0 and five SDUs use 3×3 sized filters. Five max pooling layers exploit 3×3 sized pooling windows, while HAP and VAP layers utilize 1×4 and 4×1 sized pooling windows, respectively. Two spatial normalization layers (i.e., SN1 and SN2) utilize a 4×1 sized normalization window. Moreover, only those strides on five MP layers are set as 2 pixels, and the remaining ones are set as 1 pixel.

Table 1 The parameter configuration of the proposed method.

Name	Channels	Scope of Leaky ReLU	Sub-window ($h \times w$)	Stride	Output Size
Conv0	64	0.15	3×3	1	$128 \times 128 \times 64$
SDU1	64	0.15	3×3	1	$128 \times 128 \times 64$
MP1	64	-	3×3	2	$64 \times 64 \times 64$
SDU2	128	0.15	3×3	1	$64 \times 64 \times 128$
MP2	128	-	3×3	2	$32 \times 32 \times 128$
SDU3	192	0.15	3×3	1	$32 \times 32 \times 192$
MP3	192	-	3×3	2	$16 \times 16 \times 192$
SDU4	256	0.15	3×3	1	$16 \times 16 \times 256$
MP4	256	-	3×3	2	$8 \times 8 \times 256$
SDU5	320	0	3×3	1	$8 \times 8 \times 320$
MP5	320	-	3×3	2	$4 \times 4 \times 320$
HAP	320	-	1×4	1	$4 \times 1 \times 320$
VAP	320	-	4×1	1	$1 \times 4 \times 320$
SN1	320	-	4×1	1	$4 \times 1 \times 320$
SN2	320	-	4×1	1	$4 \times 1 \times 320$
CAT	-	-	-	-	$4 \times 1 \times 640$

[*] Corresponding author (email: zeng0043@hqu.edu.cn, jinxinbesti@foxmail.com)

Appendix .2 Performance comparison

The performance comparisons among the proposed JHV-DLF and multiple state-of-art methods on VeRi database are shown in Table 2. Firstly, it can be found that the proposed JHV-DLF method acquires the highest rank-1 identification rate, 84.74%, among all the methods under comparison. Secondly, compared with three vehicle licence plate aided methods (i.e., PROVID [1], NuFACT + Plate-SNN [1] and NuFACT + Plate-REC [1]), the proposed JHV-DLF method defeats NuFACT + Plate-SNN [1] and Plate-REC [1] and is only slightly lower than PROVID [1] in MAP and rank-5 identification rate. It should be pointed out that without the aid of vehicle licence plate, the NuFACT [1] method is obviously inferior to the proposed JHV-DLF method.

Moreover, as shown in Figure 1, the proposed JHV-DLF defeats both H-DLF and V-DLF. Specifically, the MAP of JHV-DLF is 2.62% and 3.03% higher than that of H-DLF and V-DLF, respectively. Moreover, the rank-1 identification rate of JHV-DLF is 1.61% and 3.57% higher than that of H-DLF and V-DLF, respectively. This demonstrates that the proposed JHV-DLF comprehensively describing vehicle in both horizontal and vertical directions is beneficial to improve the robustness of camera viewpoint variations and so that the better performance is obtained.

Table 2 The performance (%) comparison of the proposed JHV-DLF and multiple state-of-the-art methods on VeRi [1].

Methods	MAP	Rank=1	Rank=5
JHV-DLF	53.02	**84.74**	93.15
PROVID [1]	**53.42**	81.56	**95.11**
NuFACT + Plate-SNN [1]	50.87	81.11	92.79
NuFACT + Plate-REC [1]	48.55	76.88	91.42
NuFACT [1]	48.47	76.76	91.42
LOMO [2]	9.64	25.33	46.48
BOW-SFIT [3]	1.51	1.91	4.53

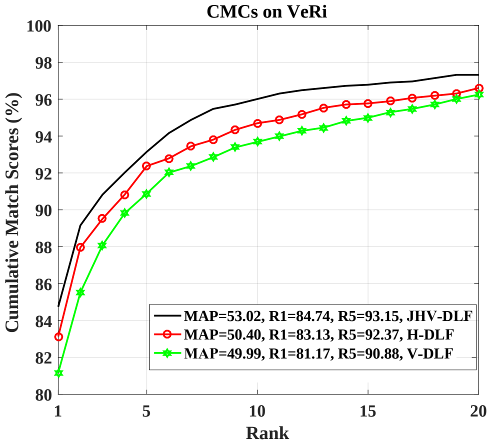

Figure 1 The performance (%) comparison of JHV-DLF, H-DFL and V-DFL on VeRi. R1 and R5 represent rank-1 and rank-5 identification rates, respectively. Moreover, the features learned with the configurations that using HV-DFLM, only using horizontal deep feature learning sub-module, only using vertical deep feature learning sub-module are denoted as JHV-DLF, H-DLF and V-DLF, respectively.

References

1 Liu X C, Liu W, Mei T, et al. Provid: Progressive and multi-modal vehicle re-identification for large-scale urban surveillance. IEEE Trans. on Multimedia, 2018, 20(3): 645–658

2 Liao S C, Hu Y, Zhu X Y, et al. Person re-identification by local maximal occurrence representation and metric learning. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Boston, Massachusetts, USA, 2015. 2197–2206

3 Zheng L, Wang S J, Zhou W G, et al. Bayes merging of multiple vocabularies for scalable image retrieval. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, Columbus, Ohio, 2014. 1963–1970