A Shortly and Densely Connected Convolutional Neural Network for Vehicle Re-identification

Jianqing Zhu*, Huanqiang Zeng†, Zhen Lei‡, Shengcai Liao‡, Lixin Zheng* and Canhui Cai*
*College of Engineering, Huaqiao University, 269 Chenghua North Road, Quanzhou, Fujian, China
Email: {jqzhu, chcai, zlx}@hqu.edu.cn
†College of Information Science and Engineering, Huaqiao University, 668 Jiemei Avenue, Xiamen, Fujian, China
Email: zeng0043@hqu.edu.cn
‡National Laboratory of Pattern Recognition, Institute of Automation, Chinese Academy of Sciences,
95 Zhongguancun East Road, 100190, Beijing, China
Email: {scliao, zlei}@nlpr.ia.ac.cn

***Abstract*—In this paper, we propose a shortly and densely connected convolutional neural network (SDC-CNN) for vehicle re-identification. The proposed SDC-CNN mainly consists of short and dense units (SDUs), necessary pooling and normalization layers. The main contribution lies at the design of short and dense connection mechanism, which would effectively improve the feature learning ability. Specifically, in the proposed short and dense connection mechanism, each SDU contains a short list of densely connected convolutional layers and each convolutional layer is of the same appropriate channels. Consequently, the number of connections and the input channel of each convolutional layer are limited in each SDU, and the architecture of SDC-CNN is simple. Extensive experiments on both VeRi and VehicleID datasets show that the proposed SDC-CNN is obviously superior to multiple state-of-the-art vehicle re-identification methods.**

I. Introduction

Vehicles, such as cars, buses, and trucks, have been an indispensable part of human life, thus vehicle re-identification plays an important role in video surveillance for public security [1]. Taking a query vehicle image as an input, vehicle re-identification aims to search in the surveillance data and match the same vehicle captured by different cameras. In practical surveillance scenarios, vehicle re-identification is a challenging vehicle matching problem in computer vision, since vehicle images are usually of low resolution and partially occluded and contain variations of illumination, viewpoint, as shown in Fig. 1.

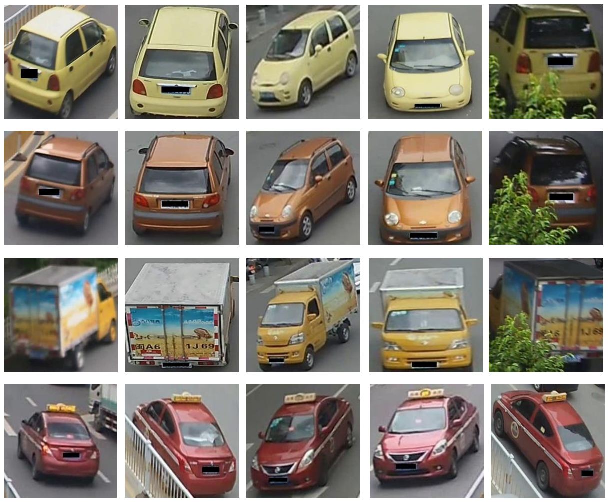


Fig. 1. Classical vehicle samples of the VeRi [7] dataset. Each row denotes the same vehicle captured under different camera-views.

Comparing to person re-identification [2–5], vehicle re-identification is still a frontier research [6]. From the perspective of released vehicle datasets, VeRi [6, 7] and VehicleID [8] are two newest large scale vehicle datasets captured under surveillance scenarios. From the point of vehicle re-identification methods, many person re-identification methods are directly utilized to realize vehicle re-identification. For example, the LOMO [4] and BOW-CN [5] features originally used in person re-identification are applied as baseline methods on the VeRi [6, 7] dataset. Moreover, some well known deep feature learning architectures, such as AlexNet [9], VGGNet [10] and GoogLeNet [11], are used as feature extractors for vehicle re-identification. For example, FACT [6] uses AlexNet [9] to extract features of vehicles, while NuFACT [7] takes GoogLeNet [11] as a feature extractor. DRDL [8] utilizes VGGNet [10] to extract features of vehicles. On the VeRi [6, 7] dataset, comparing the baseline results that reported in [6, 7], it can be found that the deep learning based baseline methods (e.g., FACT, NuFACT and GoogLeNet) obviously outperforms non deep learning based methods (e.g., LOMO and BOW-CN).

Based on large improvements of recent deep learning networks (i.e., VGGNet [10] and DenseNet [12]), we propose a shortly and densely connected convolutional neural network (SDC-CNN) for vehicle re-identification in this paper. The short and dense unit (SDU) is the basic unit of the proposed SDC-CNN. Each SDU contains a short list of densely connected convolutional layers and each convolutional layer is of the same appropriate channels, which ensures the number of connections and the input channel of each convolutional layer are limited in each SDU. Compared to VGGNet, SDC-CNN applies the dense connection mechanism to improve the feature learning ability. While compared to DenseNet, SDC-CNN does not require 1×1 bottleneck layers applied in DenseNet, which ensures SDC-CNN is much simpler than DenseNet.

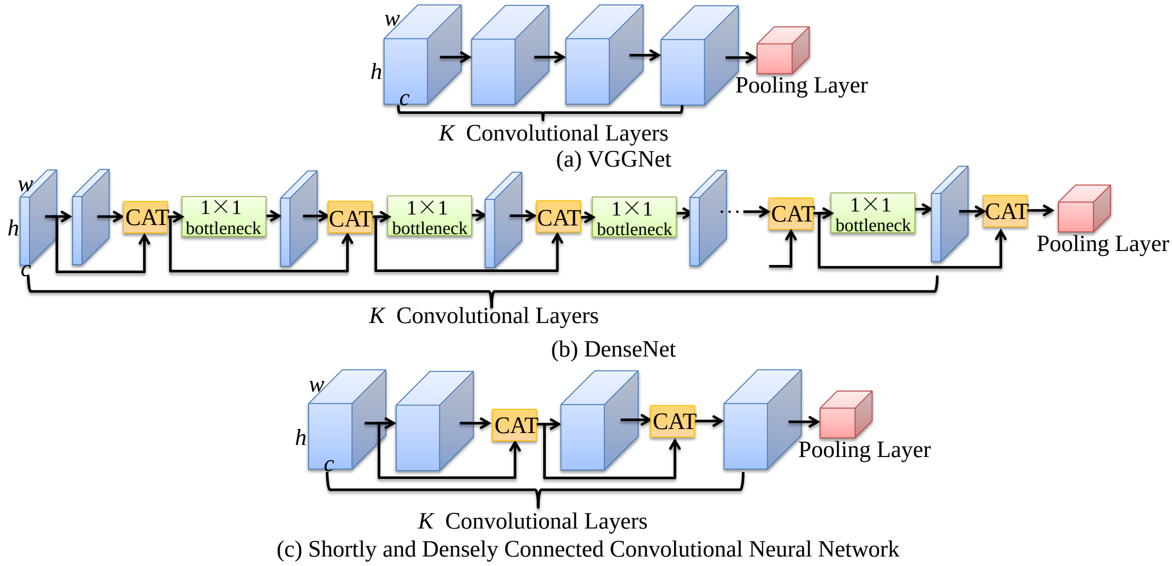


Fig. 2. The diagrams of VGGNet (a), DenseNet (b) and the proposed shortly and densely connected convolutional neural network (c).

Therefore, the proposed SDC-CNN is able to combine the advantages of VGGNet and DenseNet to improve the feature learning ability.

The major contribution of this paper is to design a shortly and densely connected convolutional neural network (SDC-CNN) for vehicle re-identification, which sharply promotes the vehicle re-identification accuracy. The rest of this paper is organized as follows. Section II introduces the proposed shortly and densely connected convolutional neural network based vehicle re-identification method. Section III presents experimental results to validate the superiority of the proposed method. Section IV makes a conclusion for this paper.

TABLE I
THE PARAMETER CONFIGURATION OF THE PROPOSED SDC-CNN.

Name	Channels	Scope of Leaky ReLU	Sub-window $(h \times w)$	Stride
Conv0	64	0.15	3×3	1
SDU1	64	0.15	3×3	1
MP1	64	-	3×3	2
SDU2	128	0.15	3×3	1
MP2	128	-	3×3	2
SDU3	192	0.15	3×3	1
MP3	192	-	3×3	2
SDU4	256	0.15	3×3	1
MP4	256	-	3×3	2
SDU5	320	0	3×3	1
MP5	320	-	3×3	2
AP	320	-	1×4	1
SLN	320	-	4×1	1

II. SHORTLY AND DENSELY CONNECTED CONVOLUTIONAL NEURAL NETWORK BASED VEHICLE RE-IDENTIFICATION

A. Shortly and Densely Connected Convolutional Neural Network

As shown in Fig. 2(a) and Fig. 2(b), both VGGNet [10] and DenseNet [12] emplace multiple same channel convolutional layers before a pooling layer. However, there are two main differences between VGGNet and DenseNet. First, VGGNet [10] uses a short list of convolutional layers and each convolutional is of the same appropriate channels (e.g., c=64), while DenseNet [12] applies a long list of convolutional layers and each convolutional layer is of the same thin channels (e.g., c=8). Second, VGGNet containing K convolutional layers has K connections—one between each layer and its subsequent layer, while DenseNet has $K(K+1)/2$ connections. For each convolutional layer of DenseNet, the feature maps of all preceding layers are used as inputs, and its own feature maps are used as inputs into all subsequent layers, which is the so-called dense connection mechanism in [12]. Based on this dense connection mechanism, the deeper a convolutional layer is, the more feature maps it will be used as the input. For example, the inputs of $k-$th and $(k+1)-$th convolutional layers hold $(k-1)\times c$ and $k \times c$ channel feature maps, respectively. Therefore, in the DenseNet, a 1×1 bottleneck layer is further applied after a concatenation layer to compress feature maps, as shown in Fig. 2(b).

Based on the comparison of VGGNet [10] and DenseNet [12], a short and dense connection mechanism is proposed to construct the shortly and densely connected convolutional neural network (SDC-CNN), as shown in Fig. 2(c). SDC-CNN contains a short list of densely connected convolutional layers and each convolutional layer is of the same appropriate channels. Moreover, 1×1 bottleneck layers are discarded in the SDC-CNN, since the number of convolutional layers in the SDC-CNN is limited. Compared to VGGNet, SDC-CNN applies the dense connection mechanism to improve the feature learning ability. While compared to DenseNet, SDC-CNN is much simpler than DenseNet, by discarding 1×1 bottleneck layers.

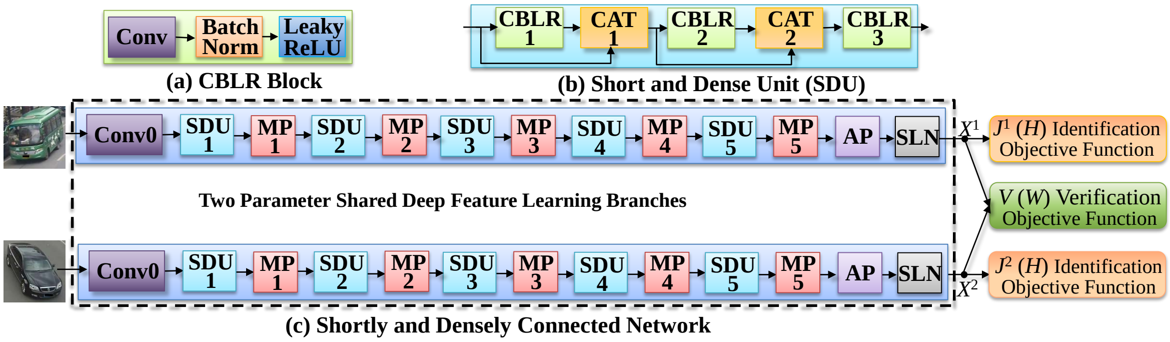


Fig. 3. The diagram of the proposed shortly and densely connected convolutional neural network (SDC-CNN) based vehicle re-identification method. MP, AP and SLN represent max pooling, average pooling and spatial local normalization layers, respectively.

B. SDC-CNN Architecture

As shown in Fig. 3, a shortly and densely connected convolutional neural network (SDC-CNN) based vehicle re-identification method is proposed. First, convolutional, batch normalization [13] and Leaky ReLU [14] layers are sequently packaged to construct a CBLR block, as shown in Fig. 3 (a). Second, three CBLR blocks are densely connected with two concatenation (i.e., CAT1 and CAT2) layers to build a short and dense unit (SDU), as shown in Fig. 3 (b). Third, one convolutional layer (i.e., Conv0), five SDUs (i.e., SDU1-SDU5), five max pooling layers (i.e., MP1-MP5), one average pooling layer (AP) and one spatial local normalization layer (SLN) [15] are packaged in turn to build the deep feature learning branch. Moreover, the SDC-CNN applies a siamese architecture [16], which includes two parameter shared deep feature learning branches, as shown in Fig. 3 (c).

The parameter configuration of the proposed SDC-CNN is listed in Table I. 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, the others are 0.15. The sub-window for Conv0 and SDU represents a filter size, for pooling layers (i.e. MP1-MP5 and AP) means a pooling window size, and for the spatial local normalization (i.e. SLN) layer denotes a local normalization window size. As shown in Table I, Conv0 and five SDUs apply 3×3 sized filters. Five max pooling layers use 3×3 sized pooling windows, while the average pooling layer uses a 1×4 sized pooling window. The spatial local normalization layer utilizes a 4×1 sized normalization window. Moreover, only those strides working on four MP layers are set as 2 pixels, the others are set as 1 pixel.

C. Objective Function Design

Similar to [17, 18], both the identification and verification objective functions are jointly applied to supervise the training process of the proposed network architecture. The same with [17, 18], the softmax function is utilized to construct the identification objective function, as follows:

$$J(H) = \frac{1}{K}[\sum_{k=1}^{K}\sum_{c=1}^{C}\ell(y^{(k)}=c)\log\frac{e^{{H_c}^T X^{(k)}}}{\sum_{p=1}^{C} e^{{H_p}^T X^{(k)}}}] + \frac{1}{2}\alpha\left\|H\right\|_2^2, \tag{1}$$

where $H = [H_1, H_2, ..., H_C] \in \Re^{d \times C}$ is the projection matrix used to predicate a vehicle's class label; $X^{(k)}$ is the deep learning feature of k-th training sample and $y^{(k)} \in \{1, 2, 3, ..., C\}$ is the corresponding class label. $\alpha \geq 0$ is a parameter used to balance the weigth of the L_2 regularization item. K and C represents the numbers of the training samples and classes. $\ell(\cdot)$ is an indicator function.

Moreover, the hybrid similarity function proposed in our previous person re-identification work [3] is applied to calculate the similarity of an input vehicle pair, which is formulated as follows:

$$s_k = W_d^T|X_k^1 - X_k^2| + W_m^T(X_k^1 .* X_k^2), \tag{2}$$

where $W = \begin{bmatrix} W_d \\ W_m \end{bmatrix} \in \Re^{d+d}$ is the projection vector used to calculate the hybrid similarity s_k between the deep learning feature pair $X_k^1 \in \Re^d$ and $X_k^2 \in \Re^d$; .* denotes the the element-wise multiplication operation between X_k^1 and X_k^2.

Based the Eq. (2), the log-logistic function [19] is further used to built the verification objective function for vehicle re-identification, as formulated in Eq. (3).

$$V(W) = \frac{1}{N}[\sum\nolimits_{n=1}^{N} \log(1 + e^{-t_n s_n})] + \frac{1}{2}\beta\left\|W\right\|_2^2, \tag{3}$$

where $t_n \in \{-1, 1\}$, if $t_n = 1$ means that X_n^1 and X_n^2 in a deep learning feature pair is with the same class label, otherwise, the deep learning feature pair is with different class labels; N is the number of training pair samples; $\beta \geq 0$ is a parameter applied to balance the weight of the L_2 regularization item.

Jointing the softmax based identification objective function (Eq. (1)) and the hybrid similarity function based verification objective function (Eq. (3)), the final objective function of this paper is obtained, as follows:

$$O(W, H) = V(W) + \lambda[J^1(H) + J^2(H)], \tag{4}$$

where $J^1(H)$ and $J^2(H)$ are two the same softmax based identification objective functions assigned to the two parameter shared feature learning branches, respectively, as shown in Fig. 3; λ is a constant applied to balance the contribution of the identification objective functions. Note that the mini-batch stochastic gradient descent [9] algorithm is applied to optimize the proposed method in this paper.

III. Experiment and Analysis

To validate the superiority of the proposed shortly and densely connected convolutional neural network (SDC-CNN) based vehicle re-identification method, it is evaluated and compared with multiple state-of-the-art methods on two challenging datasets, VeRi [7] and VehicleID [8].

A. Dataset and Evaluation Protocol

Both VeRi [7] and VehicleID [8] are two large scale datasets released by the Institute of Digital Media, Peking University. The cumulative match characteristic (CMC) curve [2, 3] and mean average precision (MAP) are applied to evaluate the vehicle re-identification performance on the VeRi and VehicleID datasets.

VeRi is captured by 20 cameras in unconstrained traffic scenarios and each vehicle is captured by 2-18 cameras in different viewpoints, illuminations, occlusions and resolutions. The VeRi dataset is divided into a training subset containing 37,781 images of 576 vehicles and a testing subset with 11,579 images of 200 vehicles. For the evaluation, one can select one image of each vehicle captured from each camera as the query and obtain a query set containing 1,678 images. Furthermore, only the cross-camera vehicle re-identification is evaluated, which means that if a probe image and a gallery image are captured under the same camera viewpoint, the corresponding matching result will be excluded in the final performance evaluation.

VehicleID is captured during daytime by multiple real-world surveillance cameras distributed in a small city in China. There are 221,763 images of 26,267 vehicles in the entire dataset. Each vehicle is captured only from a front or back viewpoint. The training subset consists of 110,178 images of 13,134 vehicles. In addition, VehicleID provides three testing subsets, test800, test1600 and test2400, for evaluating the performance in different scales. Test800 includes 800 gallery and 6,532 probe images. Test1600 contains 1600 gallery and 11,395 probe images. Test2400 consists of 2400 gallery and 17,638 probe images.

B. Training Configuration

All images in the VeRi and VehicleID datasets are scaled to 128 $\times$ 128 pixels, and each image is further augmented by horizontal mirror and randomly rotating operations. The randomly rotating operation is applied to randomly rotate an image in ranges $[-3°, 0°]$ and $[0°, 3°]$.The weights in each layer are initialized with a normal distribution $N(0, 0.01)$, and the biases are initialized to 0. The L_2 regularization weights α in Eq. (1) and β in Eq. (3) are set to 0.005. The balance constant λ in Eq. (4) is set to 0.5. The momentums are assigned as 0.9. Each mini-batch is composed of 64 positive and 64 negative image pairs, and each image pair is randomly chosen from the whole dataset. The learning rates start with 0.01 and the minimum learning rates are 0.001. The learning rates are gradually decreased along with the training progress, that is, if the objective function of Eq.(4) is convergent at a phase, the learning rates will be decreased to 10% of the original values.

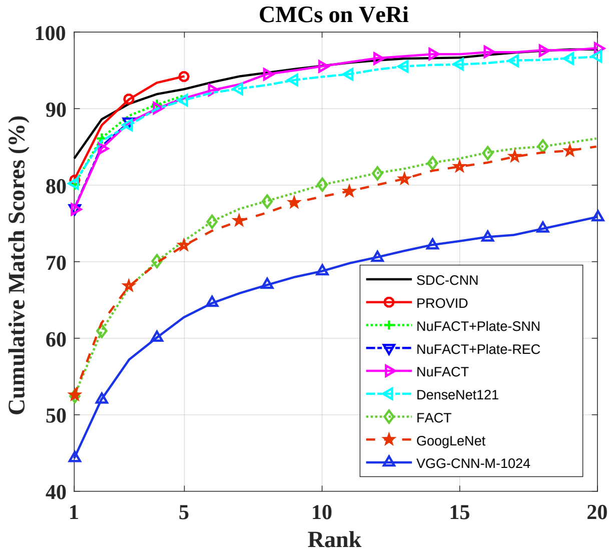


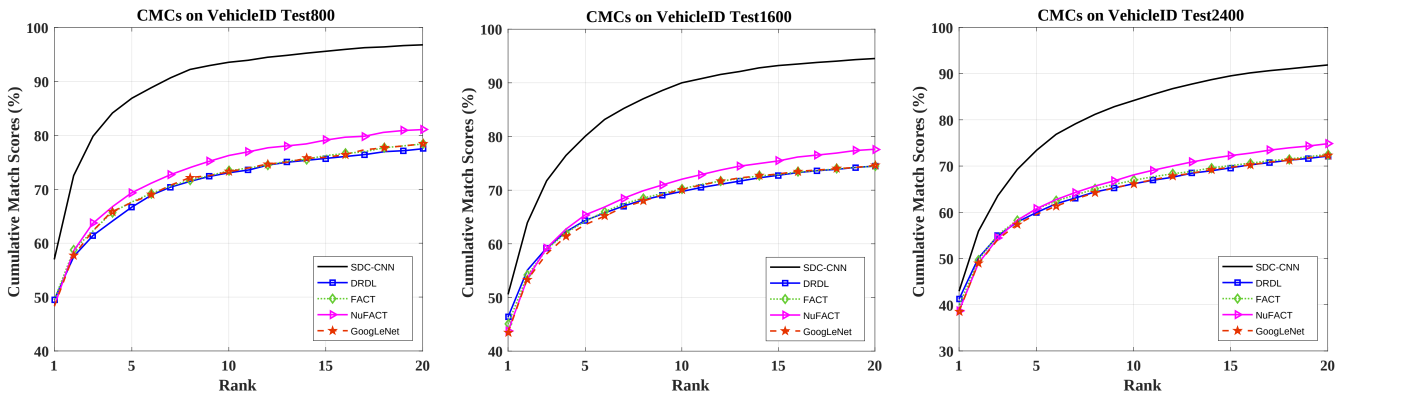
Fig. 4. The CMC curve comparison of the proposed SDC-CNN method and multiple state-of-the-art methods on VeRi.

TABLE II
THE PERFORMANCE (%) COMPARISON OF THE PROPOSED SDC-CNN AND MULTIPLE STATE-OF-THE-ART METHODS ON VERI.

Methods	MAP	Rank=1	Rank=5
SDC-CNN	**53.45**	**83.49**	92.55
PROVID [7]	53.42	81.56	**95.11**
NuFACT + Plate-SNN [7]	50.87	81.11	92.79
NuFACT + Plate-REC [7]	48.55	76.88	91.42
NuFACT [7]	48.47	76.76	91.42
DenseNet121 [12]	45.06	80.27	91.12
FACT [6]	18.75	52.21	72.88
GoogLeNet [20]	17.89	52.32	72.17
VGG-CNN-M-1024 [8]	12.76	44.10	62.63
BOW-CN [5]	12.20	33.91	53.69
LOMO [4]	9.64	25.33	46.48
BOW-SFIT [21]	1.51	1.91	4.53

C. Comparison with State-of-the-art Methods

1) Experiments on VeRi: The performance comparison of the proposed SDC-CNN and multiple state-of-art methods are shown in Fig. 4 and Table II. Firstly, from Table II, it can be found that the proposed SDC-CNN acquires the highest MAP and rank-1 identification rate. Secondly, compared with three vehicle license plate aided methods, PROVID [7], NuFACT + Plate-SNN [7] and NuFACT + Plate-REC [7], the proposed SDC-CNN method still defeat Plate-REC [7], although it is beaten by PROVID [7] and NuFACT + Plate-SNN [7] by a lower rank-5 identification rate. As shown in Table II, without the help of plates, both NuFACT [7] and FCAT [6] are outperformed by the proposed SDC-CNN method. Thirdly, compared with those very deep models, DenseNet121 [12] and

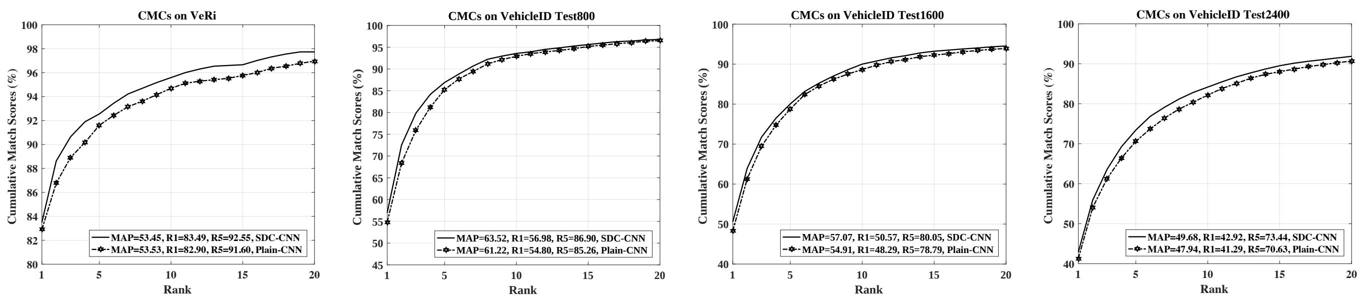


(a) (b) (c)

Fig. 5. The CMC curve comparisons of the proposed SDC-CNN method and multiple state-of-the-art methods on VehicleID Test800 (a), Test1600 (b) and Test2400 (c), respectively.

TABLE III
THE PERFORMANCE (%) COMPARISON OF THE PROPOSED SDC-CNN AND MULTIPLE STATE-OF-THE-ART METHODS ON VEHICLEID.

Method	Test800			Test1600			Test2400			Average		
	MAP	Rank=1	Rank=5	MAP	Rank=1	Rank=5	MAP	Rank=1	Rank=5	MAP	Rank=1	Rank=5
SDC-CNN	63.52	**56.98**	**86.90**	57.07	**50.57**	**80.05**	49.68	**42.92**	**73.44**	56.76	**50.16**	**80.13**
DRDL [8]	N/A	48.91	66.71	N/A	46.36	64.38	N/A	40.97	60.02	N/A	45.41	63.70
FACT [6]	N/A	49.53	67.96	N/A	44.63	64.19	N/A	39.91	60.49	N/A	44.69	64.21
NuFACT [7]	N/A	48.90	69.51	N/A	43.64	65.34	N/A	38.63	60.72	N/A	43.72	65.19
GoogLeNet [20]	N/A	47.90	67.43	N/A	43.45	63.53	N/A	38.24	59.51	N/A	43.20	60.04
LOMO [4]	N/A	19.74	32.14	N/A	18.95	29.46	N/A	15.26	25.63	N/A	17.98	3.76
BOW-CN [5]	N/A	13.14	22.69	N/A	12.94	21.09	N/A	10.20	17.89	N/A	12.09	20.56
BOW-SIFT [21]	N/A	2.81	4.23	N/A	3.11	5.22	N/A	2.11	3.76	N/A	2.68	3.76



(a) (b) (c) (d)

Fig. 6. The performance (%) comparisons of the proposed SDC-CNN method and multiple state-of-the-art methods on VeRi (a), VehicleID Test800 (b), Test1600 (c) and Test2400 (d), respectively. R1 and R5 represents rank-1 and rank-5 identification rates, respectively.

GoogLeNet [20], the proposed SDC-CNN method also obtains better results.

2) Experiments on VehicleID: The performance comparison results are shown in Fig. 5 and Table III. As shown in Fig. 5, it can be clearly found that the proposed SDC-CNN method obviously defeat those state-of-the-art methods on the VehicleID dataset. From Table III, one can find that deep learning based methods (i.e., DRDL [8], FACT [6], NuFACT [7] and GoogLeNet [20]) obviously defeat traditional methods (i.e., LOMO [4], BOW-CN [5] and BOW-SIFT [21]) on this large scale dataset. Moreover, compared with those deep learning based methods, DRDL [8], FACT [6], NuFACT [7] and GoogLeNet [20], the proposed SDC-CNN method consistently outperforms them by higher rank-1 and rank-5 identification rates on Test800, Test1600 and Test2400 datasets, and the best average result is consequently obtained.

D. Role of Short and Dense Connection Mechanism

Base on comparison results on the VeRi and VehicleID datasets, the accuracy superiority of the proposed SDC-CNN method has been validated. In this section, we further evaluate the role of the short and dense connection mechanism in the proposed method. To facilitate the description, we abbreviate the plain case that abandoning the short and dense connection mechanism while keeping the same basic network architecture as Plain-CNN.

As shown in Fig. 6, on the VeRi and VehicleID datasets, the proposed SDC-CNN outperforms Plain-CNN. For example, on the VeRi dataset, the rank-1 and rank-5 identification rates of SDC-CNN are 0.59% and 0.95% higher than those of Plain-CNN, although the MAP of SDC-CNN is a bit lower than that of Plain-CNN. Moreover, as shown in Figs. 6(b), 6(c) and 6(d), both for MAP, rank-1 and rank-5 identification rates, the proposed SDC-CNN consistently defeats Plain-CNN on three testing subset (i.e., Test800, Test1600, Test2400) of the VehicleID dataset. These results clearly show that the short and dense connection mechanism is efficient for improving the vehicle re-identification accuracy.

IV. Conclusion

This paper presented a shortly and densely convolutional neural network (SDC-CNN) for vehicle re-identification. The proposed SDC-CNN is mainly composed of short and dense units (SDUs), necessary pooling and normalization layers, where a short and dense connection mechanism is designed to make that each SDU contain a short list of densely connected convolutional layers and each convolutional layer is of the same appropriate channels. Consequently, the feature learning ability is improved and the architecture of SDC-CNN is simplified. Extensive experiments show that the proposed SDC-CNN is able to achieve better performance than multiple state-of-the-art vehicle re-identification methods on the VeRi and VehicleID datasets.

Acknowledgment

This work was supported in part by the National Natural Science Foundation of China under the Grants 61602191, 61672521, 61375037, 61473291, 61572501, 61572536, 61502491, 61372107 and 61401167, in part by the Natural Science Foundation of Fujian Province under the Grants 2018J01090 and 2016J01308, in part by the Scientific and Technology Founds of Xiamen under the Grant 3502Z20173045, in part by the Promotion Program for Young and Middle-aged Teacher in Science and Technology Research of Huaqiao University under the Grants ZQN-PY418 and ZQN-YX403, and in part by the Scientific Research Funds of Huaqiao University under the Grant 16BS108.

References

[1] C. C. Loy, T. Xiang, and S. Gong, "Multi-camera activity correlation analysis," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2009, pp. 1988–1995.

[2] D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *Workshop on Performance Evaluation of Tracking and surveillance*, 2007.

[3] J. Zhu, H. Zeng, S. Liao, Z. Lei, C. Cai, and L. Zheng, "Deep hybrid similarity learning for person re-identification," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. PP, no. 99, pp. 1–1, doi:10.1109/TCSVT.2017.2734740, 2017.

[4] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, "Person re-identification by local maximal occurrence representation and metric learning," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 2197–2206.

[5] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, "Scalable person re-identification: A benchmark," in *Proceedings of the IEEE International Conference on Computer Vision*, 2015, pp. 1116–1124.

[6] X. Liu, W. Liu, H. Ma, and H. Fu, "Large-scale vehicle re-identification in urban surveillance videos," in *Proceedings of the IEEE International Conference on Multimedia and Expo*, 2016, pp. 1–6.

[7] X. Liu, W. Liu, T. Mei, and H. Ma, "Provid: Progressive and multi-modal vehicle re-identification for large-scale urban surveillance," *IEEE Transactions on Multimedia*, vol. 20, no. 3, pp. 645–658, 2018.

[8] H. Liu, Y. Tian, Y. Wang, L. Pang, and T. Huang, "Deep relative distance learning: Tell the difference between similar vehicles," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 2167–2175.

[9] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "Imagenet classification with deep convolutional neural networks," in *Proceedings of Annual Conference on Neural Information Processing Systems*, 2012, pp. 1097–1105.

[10] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[11] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, and A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 1–9.

[12] G. Huang, Z. Liu, L. van der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 2261–2269.

[13] S. Ioffe and C. Szegedy, "Batch normalization: Accelerating deep network training by reducing internal covariate shift," *arXiv preprint arXiv:1502.03167*, 2015.

[14] B. Xu, N. Wang, T. Chen, and M. Li, "Empirical evaluation of rectified activations in convolutional network," *arXiv preprint arXiv:1505.00853*, 2015.

[15] A. Vedaldi and K. Lenc, "Matconvnet: Convolutional neural networks for matlab," in *Proceedings of the 23rd ACM international conference on Multimedia*. ACM, 2015, pp. 689–692.

[16] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Deep metric learning for person re-identification," in *Proceedings of the IEEE International Conference on Pattern Recognition*, 2014.

[17] Y. Sun, Y. Chen, X. Wang, and X. Tang, "Deep learning face representation by joint identification-verification," in *Proceedings of Annual Conference on Neural Information Processing Systems*, 2014, pp. 1988–1996.

[18] Z. Zheng, L. Zheng, and Y. Yang, "A discriminatively learned cnn embedding for person re-identification," 2016.

[19] R. C. Gupta, O. Akman, and S. Lvin, "A study of log-logistic model in survival analysis," *Biometrical Journal*, vol. 41, no. 4, pp. 431–443, 1999.

[20] L. Yang, P. Luo, C. L. Chen, and X. Tang, "A large-scale car dataset for fine-grained categorization and verification," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3973–3981.

[21] L. Zheng, S. Wang, W. Zhou, and Q. Tian, "Bayes merging of multiple vocabularies for scalable image retrieval," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1963–1970.