Visual Cognition Inspired Vehicle Re-identification via Correlative Sparse Ranking with Multi-view Deep Features

Dengdi Sun[1,2], Lidan Liu[1], Aihua Zheng[1,2(✉)], Bo Jiang[1,2], and Bin Luo[1,2]

[1] School of Computer Science and Technology, Anhui University, Hefei 230601, China
sundengdi@163.com, liulidan0@163.com
{ahzheng214,jiangbo,luobin}@ahu.edu.cn
[2] Key Lab of Industrial Image Processing and Analysis of Anhui Province, Hefei, China

Abstract. Vehicle re-identification has gradually gained attention and widespread applications. However, most of the existing methods learn the discriminative features for identities by single feature channel only. It is worth noting that visual cognition of human eyes is a multi-channel system. Therefore, integrating the multi-view information is a nature way to boost computer vision tasks in challenging scenarios. In this paper, we propose to mine multi-view deep features via correlative sparse ranking for vehicle re-identification. Specifically, first, we employ ResNet-50 and GoogleNet as two baseline networks to generate the attributes (vehicle color and type) aggregated features. Then we explore the feature correlation via enforcing the correlation term into the multi-view sparse coding framework. The original rankings are obtained by the reconstruction coefficients between probe and gallery. Finally, we utilize a re-ranking technique to further boost the performance. Experimental results on public benchmark VeRi-776 dataset demonstrate that our approach outperforms state-of-art approaches.

Keywords: Vehicle re-identification · Correlative sparse ranking
Multi-view · Deep feature

1 Introduction

With the great progress of computer vision [13,17], vehicle re-identification (Re-ID) has recently drawn much more attention due to its potential applications such as intelligent transportation, urban computing and intelligent monitoring. The aim of the vehicle Re-ID is to identify the same vehicle across non-overlapping cameras, where the license plate of the vehicle is scarcely possible to identify due to motion blur, challenging camera view etc. In addition to person Re-ID, vehicle Re-ID has particular challenges: different identities, especially from the same manufacturer, with similar colors and types possibly.

J. Ren et al. (Eds.): BICS 2018, LNAI 10989, pp. 54–63, 2018.
https://doi.org/10.1007/978-3-030-00563-4_6

Recently, many progresses have been made for vehicle Re-ID. Liu et al. [6] proposed a large surveillance-nature dataset (VehicleID) and explored Coupled Clusters Loss to measure the distance of arbitrary two input vehicle images. Zapletal et al. [15] learnt a linear classifier on color histograms and histograms of oriented gradients by vehicle 3D bounding boxes. Zhang et al. [16] designed a classification-oriented loss and triplet sampling method based on the triplet-wise network. Kanacı et al. [2] proposed to transfer the vehicle model representation for more fine-grained Re-ID tasks via a so-called cross-level vehicle recognition method. Liu et al. [7] proposed a big dataset VeRi-776 for vehicle Re-ID, and extracted the Fusion of Attributes and Color features (FACT). Furthermore, some works tried to integrate the spatio-temporal information into vehicle Re-ID process [8,11,12]. Considering that vehicles have specific attributes such as color and type, Liu [8] designed a progressive searching scheme which employed the appearance attributes of vehicle for a coarse filtering. Li et al. [4] designed a unified vehicle Re-ID framework combing identification, attribute recognition, verification and triplet tasks. However, most of existing methods implement the vehicle Re-ID only from a single feature view. Inspired by the vision system of human eyes which can perceive and decompose the multichannel information, we argue that vehicle can be recognised from multi-view characterizations.

In this paper, we propose to explore the multi-view deep feature correlation for vehicle Re-ID. Specifically, we mine the correlation between two feature space via correlative sparse coding by enforcing the correlation constraint into the multi-view sparse coding framework. It can be regarded as a general framework for multi-view feature fusion for any existing networks. In particular, we employ two attributes aggregated Re-ID networks based on ResNet-50 and GoogleNet as the subnetworks. Furthermore, inspired by the satisfactory performance of the re-ranking techniques in person Re-ID, we further utilize the Expanded Cross Neighborhood (ECN) [10] based re-ranking technique to boost the performance of the proposed method.

2 The Proposed Approach

In this section, first, we shall demonstrate the overall architecture of the proposed method, followed by the implementation details.

2.1 Overview

Given a probe vehicle image, the proposed approach regarding the vehicle Re-ID consists of the following three steps as shown in Fig. 1.

(a) Multi-view deep feature learning: We design two deep learning-based [1] subnetworks, namely ResNet-50 and GoogleNet, to extract the multi-view features by aggregating the attribute information into vehicle ID (ID-Att).
(b) Feature fusion via correlative sparse ranking: We propose to explore the correlation between the multi-view feature spaces via correlative sparse coding,

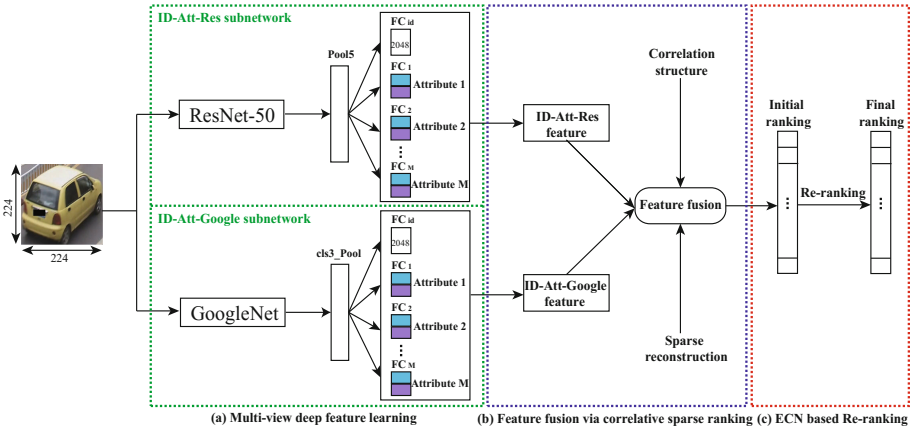

Fig. 1. Overall architecture of the proposed method. (Color figure online)

which enforces the consistency between the sparse coefficients of the multi-view features. The original ranking results are obtained according to the reconstruction coefficients between probe and gallery.

(c) ECN based Re-ranking: The finally ranking results are achieved via Expanded Cross Neighborhood (ECN) [10] based on re-ranking technique.

We shall elaborate the procedure in the following three subsections.

2.2 Multi-view Deep Feature Learning

Inspired by the human visual system, deep learning building hierarchical layers of visual representation to extract the high level features of an image, we design a multi-attribute aggregated deep learning framework to generate the multi-view features for vehicle Re-ID first. As shown in Fig. 1(a), the proposed framework consists of two subnetworks: ResNet-50 and GoogleNet. Specifically, we encode ten color attributes (yellow, orange, green, gray, red, blue, white, golden, brown, black) and nine type attributes (sedan, suv, van, hatchback, mpv, pickup, bus, truck, estate) into the deep framework, since color and type are the most recognizable appearance information.

Without loss of generality, we introduce the ResNet-50 based subnetwork as an example firstly. For the sake of attributes recognition, here we attach $M + 1$ fully-connected (FC) layers at the end, including one ID classification and M attributes, where M is the sum of the number of attributes (colors and types). Specifically, for the FC layer for ID classification, the number of output nodes equals to the number of training vehicle identities, C; while each FC layers for one attribute (color or type) link B output nodes corresponding to the B discriminant results. GoogleNet subnetwork is constructed in the same manner.

Furthermore, we append the dropout [3] layers before all FC layers in two subnetworks. The output of each hidden neuron is set to zero with rate $P = 0.9$

to avoid overfitting. Finally, for each query image, we extract two 2,048-dim feature vectors X^1 and X^2 from the two dropout layers. For gallery image set H, two feature matrices $\boldsymbol{U^1} = [u_1^1, \cdots, u_N^1]$ and $\boldsymbol{U^2} = [u_1^2, \cdots, u_N^2]$ are generated in the same manner, where u_i^1 and u_i^2 represent the feature vector of the i-th gallery image h_i from two subnetworks respectively.

Loss Computation. Given a probe vehicle image q and a gallery set H with N images of C vehicle identities, $H = \{h_i | i = 1, 2, \ldots, N\}$, Let $D = \{x_i, d_i, l_i\}_{i=1}^N$ be the image-label tuple where d_i and $l_i = \{l_i^1, ..., l_i^M\}$ denote the vehicle identity and the attribute labels of the i-th image x_i respectively, where $M = 10$ (colors) $+9$ (types), and $l_i^m \in \{0, 1\}$ indicates the binary attribute label.

Take ResNet-50 as the example. As shown in Fig. 1(a), we use the softmax classification loss function to optimise vehicle identity discrimination in vehicle ID classification branch. The output of the last FC layer is $z^{ID} = [z_1^{ID}, z_2^{ID}, ..., z_C^{ID}] \in \mathcal{R}^C$, where z_i^{ID} is the predicted result of the i-th ID. Thus, the prediction probability of each vehicle ID c calculated by softmax is: $p(c) = \frac{exp(z_c^{ID})}{\sum_{i=1}^{C} exp(z_i^{ID})}$, $c = 1, \cdots, C$. So the total vehicle ID loss is calculated by cross entropy loss function as:

$$L_{ID} = -\sum_{c=1}^{C} f(c) log(p(c)), \tag{1}$$

where $f(c)$ is a $1 \times C$ vector and represents the ground-truth of vehicle ID. Suppose t is the ground-truth of current sample, $f(c) = 1$ when $c = t$, and $f(c) = 0$ for all $c \neq t$.

The output of the FC layer for attribute j is $z_j^{Att} = \left[z_{j_1}^{Att}, \cdots, z_{j_B}^{Att}\right] \in \mathcal{R}^B$, where $B = 2$ denotes the binary discriminant results ("yes" or "no") for attribute j. Therefore, the prediction probability of the j-th attribute label is $p(j_b) = \frac{exp(z_{j_b}^{Att})}{\sum_{j_i=1}^{B} exp(z_{j_i}^{Att})}$, $j_b = 1, \cdots, B$. Similarly, we can compute the loss function of vehicle attribute j as:

$$L_{Att_j} = -\sum_{j_b=1}^{B} f(j_b) log(p(j_b)), \tag{2}$$

where $f(j_b)$ is $1 \times B$ representing the ground-truth of vehicle attribute.

The final ID-Attribute subnetwork loss is defined as:.

$$L = \beta L_{ID} + \frac{1}{M} \sum_{j=1}^{M} L_{Att_j}, \tag{3}$$

where β is a parameter to balance the contribution of vehicle re-identification loss and attribute loss.

2.3 Feature Fusion via Correlative Sparse Ranking

Above, multi-view deep features of the same vehicle are obtained. In this subsection, based on their closely latent correlations, we formulate the multi-view feature fusion problem via a sparse coding framework due to its robustness to noise, and propose a correlative sparse representation to bridge the multi-view features generated from two subnetworks as shown in Fig. 1(b).

Model Formulation. The main idea of sparse coding is to represent a input vector approximately as weighted linear combination of a small number of basis vectors from the dictionary. These basis vectors thus capture high-level patterns in the input data, while the coefficients consist of the sparse representation of the input data. According to this principle, for each query sample, we calculate sparse representation α^k under the k-th channel, where $X^k \approx \boldsymbol{U}^k\alpha^k$, for $k = 1, \cdots, K$, K is the number of the views and $K = 2$ in this paper. The process above can be converted into a ℓ_1-norm sparsity constraint regularized least squares problem:

$$\min_{\alpha^k} \|X^k - \boldsymbol{U}^k\alpha^k\|_2^2 + \lambda^k\|\alpha^k\|_1, \tag{4}$$

where λ^k controls the trade-off between the ℓ_2-norm reconstruction error and the ℓ_1-norm sparsity constraint of the coefficients under the k-th view.

To explore the correlations of multi-view features, it is natural to punish the diversity between sparse coefficients from arbitrary two corresponding views, that is minimizing the Euclidean distance $\|\alpha^{k_1} - \alpha^{k_2}\|_2^2$ for arbitrary $k_1, k_2 \in 1, \ldots, K$ to find the collaborative representation from multi-views of the same vehicle. Thus, the correlative sparse ranking model can be formulated as:

$$\underbrace{\min_{\alpha^k} \sum_{k=1}^{K} \{\|X^k - \boldsymbol{U}^k\alpha^k\|_2^2 + \lambda_k\|\alpha^k\|_1\}}_{\text{sparse reconstruction item}} + \underbrace{\mu \sum_{k_1 \neq k_2} \|\alpha^{k_1} - \alpha^{k_2}\|_2^2}_{\text{correlation item}}, \quad \text{s.t.} \forall \alpha^k \succeq 0, \tag{5}$$

where μ is the trade-off parameter to balance the sparse reconstruction error and the pairwise correlation constraints. At last, the final sparse representation vector for one query to all gallery images is expressed as: $\alpha = \sum_{k=1}^{K} \alpha^k$.

Model Optimization. Due to the non-negativeness of α^k, Eq. (5) can be written as follows:

$$\min_{\alpha^k} \sum_{k=1}^{K} \{\|X^k - \boldsymbol{U}^k\alpha^k\|_2^2 + \lambda_k\alpha^k\mathbf{1}\} + \mu \sum_{k_1 \neq k_2} \|\alpha^{k_2} - \alpha^{k_2}\|_2^2, \quad \text{s.t.} \forall \alpha^k \succeq 0, \tag{6}$$

where $\mathbf{1}$ denotes the vector with all elements as 1. To solve Eq. (6), we convert it to an unconstrained form as:

$$\min_{\alpha^k} \sum_{k=1}^{K} \{\|X^k - \boldsymbol{U}^k\alpha^k\|_2^2 + \lambda_k\alpha^k\mathbf{1}\} + \mu \sum_{k_1 \neq k_2} \|\alpha^{k_1} - \alpha^{k_2}\|_2^2 + \psi(\alpha), \tag{7}$$

Algorithm 1. Optimization Procedure to Eq. (7)

Input: The feature vector X^k of query image q, the gallery feature matrix $\boldsymbol{U}^k$, $k = 1, \cdots, K$, the parameters λ, μ;
Set $\xi = 2.7 \times 10^5$, $\rho_0 = \rho_1 = 1$, $\varepsilon = 10^{-4}$, $maxIter = 200$, $r = 1$.
Output: α^k

1: **while** not converged **do**
2: Update Ω_{r+1}^k by $\Omega_{r+1}^k = \alpha_r^k + \frac{\rho_{r-1}-1}{\rho_r}(\alpha_r^k - \alpha_{r-1}^k)$, where ρ_r is a positive sequence;
3: Update α_{r+1}^k by Eq.(10);
4: Update $\rho_{r+1} = \frac{1+\sqrt{1+4\rho_r^2}}{2}$;
5: Update r by $r = r + 1$;
6: Check the convergence condition: the maximum element change of α^k between two consecutive iterations is less than ε or maximum number of iterations reaches $maxIter$.
7: **return** α^k

where $\psi(\alpha_i^k)$ equals 1 if $\alpha_i^k \geq 0$, and 0 otherwise. α_i^k denotes the representation coefficient of gallery image h_i to the query sample from the k-th subnetwork. In this paper, we utilize the accelerated proximal gradient (APG) approach to optimize efficiently. We denote:

$$\begin{aligned} F &= \min_{\alpha^k} \|X^k - \boldsymbol{U}^k \alpha^k\|_2^2 + \lambda_k \alpha^k \mathbf{1} + \mu \sum_{k_1 \neq k_2} \|\alpha^{k_1} - \alpha^{k_2}\|_2^2, \\ J &= \psi(\alpha). \end{aligned} \tag{8}$$

Obviously, F is a differentiable convex function and J is a nonsmooth convex function. Therefore, according to the APG method, we obtain:

$$\alpha_{r+1}^k = \min_{\alpha^k} \frac{\xi}{2} \|\alpha^k - \Omega_{r+1}^k + \frac{\nabla F(\Omega_{r+1}^k)}{\xi}\|_2^2 + J(\alpha^k), \tag{9}$$

where ξ is the Lipschitz constant, r indicates the current iteration, and α_{r+1}^k denotes the sparse representation coefficients of the query image at the $(r+1)$-th iteration based on the k-th subnetworks. $\Omega_{r+1}^k = \alpha_r^k + \frac{\rho_{r-1}-1}{\rho_r}(\alpha_r^k - \alpha_{r-1}^k)$, where ρ_r is a positive sequence with $\rho_0 = \rho_1 = 1$. Equation (9) can be solved by:

$$\alpha_{r+1}^k = \max(0, \Omega_{r+1}^k - \frac{\nabla F(\Omega_{r+1}^k)}{\xi}). \tag{10}$$

Algorithm 1 summarizes the whole optimization procedure.

After obtaining the collaborative sparse representations α for each query image, we aggregate them as a representation coefficients matrix $\mathbf{A} \in \mathcal{R}^{Q \times N}$. The entry $\alpha_{q,i}$ in $\mathbf{A}$ denotes the representation coefficient of a gallery image h_i to the query image q. Then, the original distance between two vehicle images q and h_i can be calculated by $d(q, h_i) = 1/\alpha_{q,i}$. Therefore, the initial ranking to query sample q is $L(q, H) = \{h_1^0, h_2^0, \cdots, h_N^0\}$, where $d(q, h_i^0) < d(q, h_{i+1}^0)$.

2.4 ECN Based Re-ranking

The initial ranking directly compares the distance between the two images, and ignores the correlations among similar images. In order to enhance retrieval performance, here we calculate the distance by averaging the expanded neighbors of probe and gallery image pairs, that is the Expanded Cross Neighborhood (ECN) [10] distance, as shown in Fig. 1(c).

Formally, given the initial ranking $L(q, H) = \{h_1^0, h_2^0, \cdots, h_N^0\}$ for arbitrary query image q, the expanded neighbors of q are defined as the multi-set $N(q, R)$ including two parts: $N(q, l)$ and $N(l, p)$ which represent the top l samples of the query q and the top p neighbors of each of the elements in set $N(q, l)$ respectively:

$$\begin{aligned} N(q,l) &= \{h_i^0 | i = 1, 2, \cdots, l\}; \\ N(l,p) &= \{N(h_1^0, p), \cdots, N(h_l^0, p)\}. \end{aligned} \tag{11}$$

Then we expand the neighbor of each gallery image as a multi-set $N(h_i, R)$ in the same manner, where R is the total number of neighbors in the set $N(h_i, R)$, and $R = l + l \times p$. Finally, the ECN [10] distance of an image pair (q, h_i) is:

$$ECN(q, h_i) = \frac{1}{2R} \sum_{j=1}^{R} [d(qN_j, h_i) + d(h_i N_j, q)], \tag{12}$$

where qN_j is the j-th neighbor in the query expanded neighbor set $N(q, R)$ and $h_i N_j$ is the j-th neighbor in the i-th gallery image expanded neighbor set $N(h_i, R)$. In practise, $l = 4$, $p = 3$.

3 Experiment

3.1 Experiment Settings

Dataset. We evaluate our method on recent public benchmark dataset VeRi-776 dataset [7] for vehicle re-identification. The dataset contains 51035 images of 776 vehicles captured by 20 cameras in real-world traffic surveillance environment. Specifically, there are 37778 images of 576 vehicles for training, 11579 images of 200 vehicles for testing and 1678 for query. Each vehicle is captured by 2-18 cameras along a circular road. Furthermore, each vehicle image is annotated with corresponding attributes e.g. type and color.

Parameters. During the deep feature extraction, we resize all training images into 256×256 pixels and extract randomly 224×224 patches to data augmentation. We train our models using stochastic gradient descent (SGD) with a batch size of 16, momentum of 0.9, and weight decay of $\lambda = 0.0001$. The learning rate is set to 0.1 at the beginning and is changed to 0.01 in the last few epochs. $\beta = 10$ in Eq. (3) and $\lambda_1 = 0.1$, $\lambda_2 = 0.1$ in Eq. (5).

Evaluation Metric. Following the evaluation protocol of re-identification work [8,11], the mean average precision (mAP), Rank-1 and Rank-5 accuracies are utilized to evaluate the performance of re-identification in camera network.

3.2 Evaluation Results

Comparison with Vehicle Re-ID Methods. We evaluate the performance of the proposed method comparing with the state-of-the-art methods on VeRi-776 dataset and report the results in Table 1. Generally speaking, our method outperforms most of existing state-of-the-art methods. Although the mAP of Siamese+Path-LSTM [11] is comparative to ours, it is worth noting that it has utilized additional spatio-temporal path information. Even though, the Rank-1 and Rank-5 accuracies of ours are significantly higher (without any path information). The specifications of the compared methods in Table 1 are described as follows:

Table 1. The mAP, Rank-1 and Rank-5 comparison on VeRi-776 dataset (in %).

Method	mAP	Rank-1	Rank-5
(1) LOMO [5]	9.64	25.33	46.48
(2) BOW-CN [18]	12.20	33.91	53.69
(3) GoogLeNet [14]	17.89	52.32	72.17
(4) FACT [7]	18.49	50.95	73.48
(5) FACT+Plate-SNN+STR [8]	27.70	61.44	78.78
(6) NuFACT [9]	48.47	76.76	**91.42**
(7) Siamese-Visual [11]	29.48	41.12	60.31
(8) Siamese+Path-LSTM [11]	**58.27**	**83.49**	90.04
Ours	**58.21**	**90.52**	**93.38**

(1) **LOMO** [5]. Local Maximal Occurrence Representation (LOMO) is a local feature descriptor coping with illumination variations and viewpoint changes.
(2) **BOW-CN** [18]. Bag-of-Word based hand-crafted features for vehicle Re-ID.
(3) **GoogLeNet** [14]. Pre-trained on ImageNet [3] and then fine-tuned on the CompCars dataset for semantic feature representation of vehicles.
(4) **FACT** [7]. Fused Appearance features including color, texture and shape.
(5) **FACT+Plate-SNN+STR** [8]. **FACT** [7] with additional plate verification and spatio-temporal relations (STR) based on Siamese Neural Network (SNN).
(6) **NuFACT** [9]. The null space based **FACT** [7] to integrate the multi-level appearance features of vehicles and high-level attribute features.
(7) **Siamese-Visual** [11]. Siamese-CNN with only visual information.
(8) **Siamese+Path-LSTM** [11]. Siamese-CNN together with Path LSTM with visual-spatio-temporal path information.

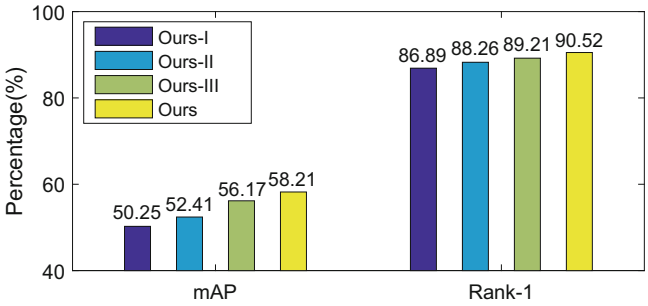

Fig. 2. Component analysis of our method.

Component Analysis. We further evaluate the components of the proposed method with its variants as shown in Fig. 2, where, Ours: the proposed method. Ours-I: only GoogleNet subnetwork without correlative sparse ranking or ECN re-ranking. Ours-II: only ResNet-50 subnetwork without correlative sparse ranking or ECN re-ranking. Ours-III: correlative sparse ranking on two subnetworks but without ECN re-ranking. From Fig. 2 we can see: (1) ResNet-50 subnetwork outperforms the GoogleNet. (2) By correlatively learning on two subnetworks, it can improve the performance of the re-identification. (3) The re-ranking technique can further boost the performance on both mAP and Rank-1.

4 Conclusions

In this paper, inspired by multi-channel visual cognition of human eyes, we discover the correlation between multi-view deep features for vehicle Re-ID. In deep feature extraction, two CNN based attributes aggregated re-identification networks are trained to generate the multi-view deep features. Then we propose the correlative sparse ranking method to jointly learn the coupled sparse coefficients for multi-view features. Furthermore, we re-rank the initial ranking via ECN distance to boost the recognition accuracy. Experimental results on benchmark VeRi-776 demonstrate the promising performance of our method. In the future, we shall further integrate the path and plate information for vehicle Re-ID.

Acknowledgment. This work was partially supported by the National Natural Science Foundation of China (61502006, 61602001 and 61671018), the Key Natural Science Project of Anhui Provincial Education Department (KJ2018A0023) and and Open Project of Anhui University (ADXXBZ201511).

References

1. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

2. Kanaci, A., Zhu, X., Gong, S.: Vehicle re-identification by fine-grained cross-level deep learning. In: Proceedings 5th Activity Monitoring by Multiple Distributed Sensing Workshop, British Machine Vision Conference, London, September 2017
3. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
4. Li, Y., Li, Y., Yan, H., Liu, J.: Deep joint discriminative learning for vehicle re-identification and retrieval. In: IEEE SigPort (2017)
5. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2197–2206 (2015)
6. Liu, H., Tian, Y., Yang, Y., Pang, L., Huang, T.: Deep relative distance learning: tell the difference between similar vehicles. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2167–2175 (2016)
7. Liu, X., Liu, W., Ma, H., Fu, H.: Large-scale vehicle re-identification in urban surveillance videos. In: 2016 IEEE International Conference on Multimedia and Expo (ICME), pp. 1–6 (2016)
8. Liu, X., Liu, W., Mei, T., Ma, H.: A deep learning-based approach to progressive vehicle re-identification for urban surveillance. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 869–884. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_53
9. Liu, X., Liu, W., Mei, T., Ma, H.: Provid: progressive and multimodal vehicle reidentification for large-scale urban surveillance. IEEE Trans. Multimed. **20**, 645–658 (2018)
10. Sarfraz, M.S., Schumann, A., Eberle, A., Stiefelhagen, R.: A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. arXiv preprint arXiv:1711.10378 (2017)
11. Shen, Y., Xiao, T., Li, H., Yi, S., Wang, X.: Learning deep neural networks for vehicle re-id with visual-spatio-temporal path proposals. CoRR. vol. abs/1708.03918 (2017)
12. Wang, Z., et al.: Orientation invariant feature embedding and spatial temporal regularization for vehicle re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 379–387 (2017)
13. Yan, Y., et al.: Cognitive fusion of thermal and visible imagery for effective detection and tracking of pedestrians in videos. Cogn. Comput. **10**, 94–104 (2018)
14. Yang, L., Luo, P., Change Loy, C., Tang, X.: A large-scale car dataset for fine-grained categorization and verification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 3973–3981 (2015)
15. Zapletal, D., Herout, A.: Vehicle re-identification for automatic video traffic surveillance. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 25–31 (2016)
16. Zhang, Y., Liu, D., Zha, Z.J.: Improving triplet-wise training of convolutional neural network for vehicle re-identification. In: 2017 IEEE International Conference on Multimedia and Expo (ICME), pp. 1386–1391 (2017)
17. Zhao, C., Li, X., Ren, J., Marshall, S.: Improved sparse representation using adaptive spatial support for effective target detection in hyperspectral imagery. Int. J. Rem. Sens. **34**, 8669–8684 (2013)
18. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: a benchmark. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1116–1124 (2015)