

RESEARCH

Open Access



Learning local embedding deep features for person re-identification in camera networks

Zhong Zhang^{1,2*} and Meiyang Huang^{1,2}

Abstract

In this paper, we propose a novel feature learning method named local embedding deep features (LEDf) for person re-identification in camera networks. In order to learn the structural information of pedestrian, we first utilize the verification network that does not require explicit identity labels to obtain the local summing maps. We then combine all local summing maps of a pedestrian image to form the holistic summing map which has the same identity label with the original pedestrian image. Finally, we take the holistic summing maps as the input to train the identification network, and then obtain the LEDf from the last fully connected layer. The proposed LEDf fully considers the structural information by learning the local features and meanwhile possesses strong discriminative ability by learning global features. The experimental results on two large-scale datasets (Market-1501 and CUHK03) demonstrate that the proposed LEDf achieves better results than the state-of-the-art methods.

Keywords: Camera networks, Person re-identification, Local summing map, Holistic summing map

1 Introduction

Person re-identification in camera networks [1–4] aims at matching the query person in a large image gallery where the images are captured from the different camera sensors in the network as shown in Fig. 1. The performance of person re-identification in sensor networks is closely related to many other applications, such as person retrieval, behavior analysis, long-term person tracking, and so on [5]. However, person re-identification in sensor networks is a very challenging problem for two reasons. Firstly, the same person observed in different cameras often undergoes high variations in illumination, poses, viewpoints, and occlusions. Secondly, a surveillance camera in the public captures hundreds of pedestrians within 1 day, and some of them have similar appearance. The core of person re-identification is to find a discriminative representation and a good metric to evaluate the similarity between pedestrian images.

Recently, the progress of person re-identification mainly consists of feature learning using the convolutional neural network (CNN) on large-scale datasets [6–11]. There

are two major kinds of CNN models for person re-identification, i.e., verification network and identification network. The verification network [12–14] treats person re-identification as a binary classification task, and its structure is shown in Fig. 2a. The input of verification network is a pair of pedestrian images, and the output indicates whether the input image pair belongs to the same class or not. The verification network maps the images into a discriminative space where two images from the same person are close to each other and two images from different person are far apart. Unlike the verification network, the identification network [15–18] regards person re-identification as a multi-class classification task, which takes full advantage of the identity labels of pedestrian images. The identification network learns the non-linear functions from the training images, and its structure is shown in Fig. 2b. Although the identification network has stronger discriminative ability than that of verification network, most existing methods train the identification network using the whole pedestrian images, which discards the structural information provided by the region of a pedestrian. The structural information, which is prone to be robust for the changes in viewpoints and poses, could be fully discovered by learning the regions of pedestrians. However, it is hard to define the identity

*Correspondence: zhong.zhang8848@gmail.com

¹Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin, China

²College of Electronic and Communication Engineering, Tianjin Normal University, Tianjin, China

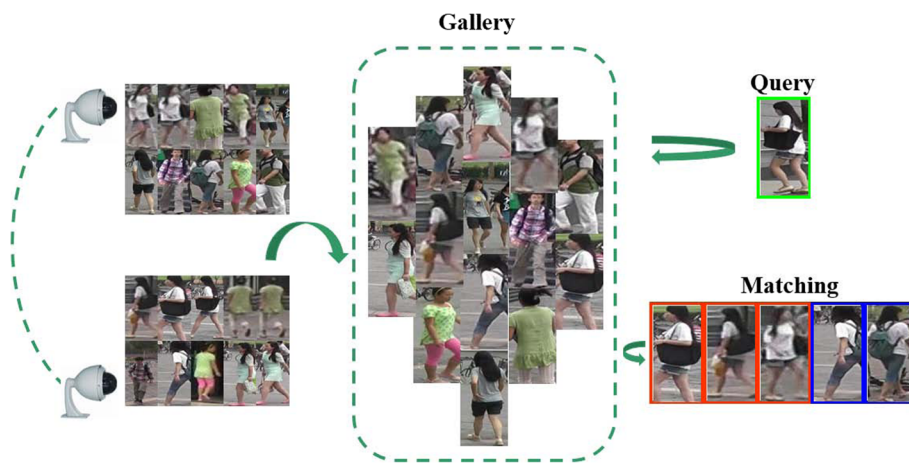


Fig. 1 The definition of person re-identification in sensor networks. The green bounding box indicates a query image which is utilized to search images of the same pedestrian from the gallery. The red bounding boxes and the blue bounding boxes show the correctly and wrongly matched pedestrian images from the gallery, respectively

label for the region of a pedestrian, and the identity label is essential for training the identification network.

In this paper, we propose a novel feature learning method named local embedding deep features (LEDf) for person re-identification in sensor networks which train an identification network in a local way. For learning the structural information, we first divide each pedestrian image into several regions. The corresponding regions of two images from the same pedestrian are considered as local similar pairs as shown in Fig. 3a, and

the corresponding regions of two images from different pedestrians are considered as local dissimilar pairs as shown in Fig. 3b. We expect to force the regions to the same feature space, and therefore, we first utilize the verification network, which does not require explicit identity labels, to learn the local features for similar and dissimilar pairs. Then, we extract local feature maps visualized in Fig. 4a and average them as the local summing map visualized in Fig. 4b for each region in the verification network. Since the discrimination of

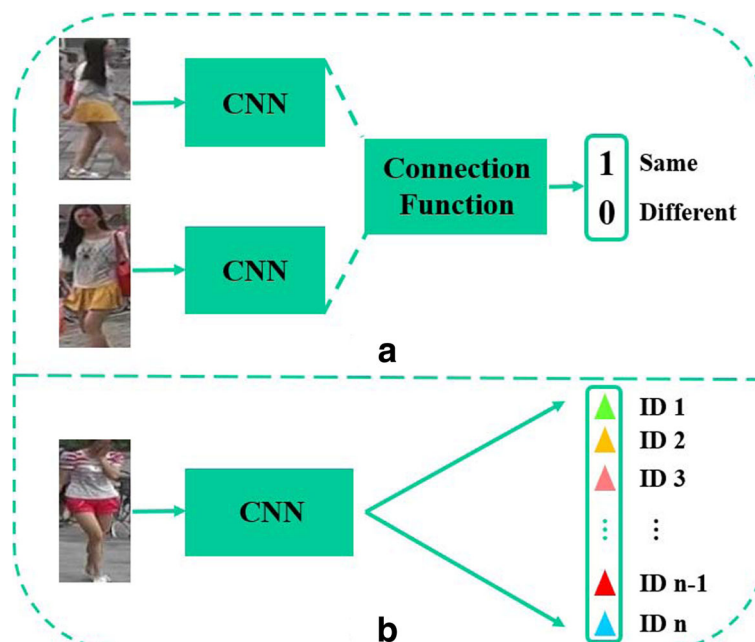


Fig. 2 Two types of CNNs. **a** The verification network. **b** The identification network



Fig. 3 Two kinds of local image pairs. **a** The corresponding regions of two images belonging to the same pedestrian are considered as local similar pairs. **b** The corresponding regions of two images belonging to different pedestrians are considered as local dissimilar pairs

identification network is stronger than that of verification network, we utilize the identification network to learn the global information. However, the local summing maps do not have the identity labels. To overcome this drawback, we connect all local summing maps of a pedestrian image to form the holistic summing map visualized in Fig. 4c. Afterwards, the holistic summing map is assigned the same identity label with the original image, so that we can train the identification network using the holistic summing maps and their identity labels. We extract the final feature representation for each pedestrian image from the fully connected layer of the identification network. The extracted features are denoted as LEDF which fully consider the structural information of pedestrian images by analyzing the local features and meanwhile possess strong discrimination by using the identification network. We validate the proposed

LEDF on two large-scale person re-identification datasets (Market-1501 [19] and CUHK03 [7]), and the experimental results demonstrate the effectiveness of the proposed LEDF.

The rest of this paper is organized as follows. In Section 2, we review and discuss the related works. In Section 3, we introduce the proposed LEDF in detail. Then, we show the experimental results in Section 4. Finally, we conclude this paper in Section 5.

2 Related work

With the development of sensor network theories [20, 21], there are many applications in sensor networks. The person re-identification is one of the most important applications. In this section, we review the related works for person re-identification in two aspects: the traditional methods and the CNN-based methods.

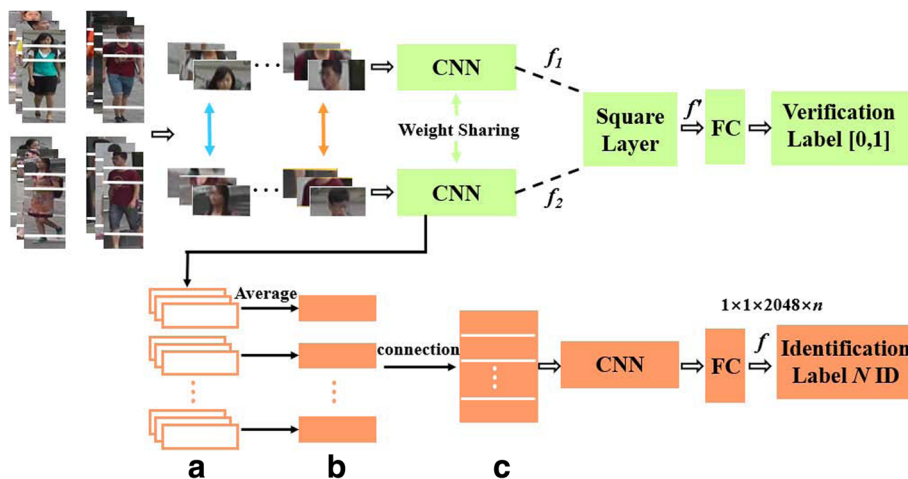


Fig. 4 The network structure of the proposed method. The verification network includes two CNNs, one square layer, one additional fully connected layer and one loss. The identification network consists of one CNN, one fully connected layer and one loss. **a**, **b**, and **c** indicate local feature map, local summing map, and holistic summing map, respectively

The traditional methods for person re-identification mainly include two key steps, i.e., feature representation and metric learning. Many approaches propose effective features for describing the appearances of person under various environments [22–28]. Cheng et al. [22] utilized the Custom Pictorial Structure (CPS) to extract an ensemble of features considering both part-based color information and color displacement for person re-identification. Gheissari et al. [23] employed the spatiotemporal segmentation algorithm to generate salient edges that are robust to the color of clothing. In [24], a novel descriptor based on a simple seven-dimensional feature vector was proposed for person re-identification. Liao et al. [25] proposed an effective feature representation called Local Maximal Occurrence (LOMO), which analyzes the horizontal occurrence of local features and maximizes the occurrence to make a stable representation against viewpoint changes. Besides robust feature representation, metric learning methods play an essential role in person re-identification [29–34]. Koestinger et al. [29] presented KISSME method to learn a distance metric from equivalence constraints based on a statistical inference perspective. Ye et al. [30] proposed the Regularized Classical Linear Discriminant Analysis (RLDA) metric which provides a simple strategy to overcome the singularity problem by applying a regularization term. In [31], a discriminative Mahalanobis metric learning method was proposed to obtain a simplified formulation. Zheng et al. [32] formulated person re-identification as a distance learning problem. They aim to learn the optimal distance which maximizes the matching accuracy regardless of the choice of features.

Recently, the convolutional neural network (CNN) has made great progress in the task of person re-identification, and they are mainly based on the verification network and the identification network. For the verification network, Ahmed et al. [35] added the neighborhood difference layer and the subsequent layer to compare convolutional features in the patch level and summarize the neighborhood differences of each patch. In [36], Ding et al. utilized a series of triplet samples to train the networks which consider the distances between pedestrian images from the same class and the distances between pedestrian images from the different classes, simultaneously. To fully discover the structural information, some methods train the CNN in part-based way. Varior et al. [9] combined the verification network with some gating functions to selectively emphasize common local patterns by comparing the mid-level features across pairs of images. In [37], pedestrian images were cropped into three overlapped parts and three networks were trained jointly. Similarly, Cheng et al. [38] presented to learn both the full-body and local-body features for the pedestrian images. However, the verification network does not fully consider the

identity information of images and cannot predict the identity labels of pedestrians. In order to take full advantages of the identity labels, the identification network which has strong discrimination is employed for person re-identification. Zheng et al. [39] fine tuned the identification network on three large datasets and achieved competitive results. Wu et al. [40] proposed the feature extraction network named Feature Fusion Net (FFN) for pedestrian image representation, which combines CNN embeddings with the hand-crafted features in a fully connected layer. Moreover, Zheng et al. [10] combined the verification network with the identification network to learn more discriminative pedestrian descriptors. However, the abovementioned methods learn the identification network in a holistic way, which ignores the structural information.

3 Method

3.1 Overall network

The architecture of the proposed network contains two parts, i.e., the verification network and the identification network, as shown in Fig. 4.

The verification network includes two CNNs, one square layer, one fully connected layer and one loss function. The square layer combines the outputs of two CNNs, i.e., f_1 and f_2 resulting in a vector f' . Since the verification network regards person re-identification as a binary classification task, we add a fully connected layer to convert f' into a 2-dim vector that determines whether the input local pair belongs to the same class or not. The softmax function is utilized to obtain the predicted probability of the input local pair belonging to the same class.

The identification network consists of one CNN, one fully connected layer and one loss function. We set the final fully connected layer of the CNN with N -dim for fine-tuning the network on person re-identification datasets, where N denotes the number of identity labels. We regard the vector f from the fully connected layer as the final feature representation for pedestrian images.

3.2 Learning the local summing maps

The identification network could enhance the feature discrimination for person re-identification. Most existing methods train the identification network using the whole pedestrian images, which ignores the structural information of pedestrians. The structural information could be provided by local regions. However, the region identity label of a pedestrian image is difficult to define, and meanwhile, the identity label is essential for training the identification network.

To solve this dilemma, we first utilize the verification network to learn the structural information from regions, which does not require explicit identity labels. Concretely, the verification network takes the local similar

and dissimilar pairs as the input. Then we train the verification network using weight sharing strategy which means that the parameters of two CNNs are the same. The square layer is employed to connect the outputs of two CNNs

$$f' = (f_1 - f_2)^2 \quad (1)$$

where f' is the output vector of square layer. Then, f' is fed into the fully connected layer and afterwards passes through the softmax function resulting in a 2-dim vector $(\tilde{p}_1, \tilde{p}_2)$ which indicates the predicted probability of the input local pair belonging to the same class. Here, $\tilde{p}_1 + \tilde{p}_2 = 1$. The verification network takes the person re-identification as a binary classification task and we utilize the cross-entropy as the loss function

$$\text{Loss}_{vr} = - \sum_{i=1}^2 p_i \log(\tilde{p}_i) \quad (2)$$

where

$$\tilde{p}_i = \frac{e^{x_i}}{\sum_{j=1}^2 e^{x_j}} \quad (3)$$

where \tilde{p}_i is the predicted probability, x_i is the output of the fully connected layer, and p_i is the true label. If the input is a local similar pair, then $p_1 = 1$ and $p_2 = 0$; otherwise, $p_1 = 0$ and $p_2 = 1$.

After training the verification network, we extract the local summing map for each region as shown in Fig. 4a and b. The local summing map could obtain the local and complete structural information for pedestrian images. The local summing map of the k -th region is formulated as

$$Z_k = \sum_{m=1}^M z_k^m / M \quad (4)$$

where M is the number of local feature maps, and z_k^m denotes the m -th local feature map of the k -th region.

3.3 Local embedding deep features

To complement with the verification network, we utilize the identification network to take full advantages of the identity labels and learn global features with strong discrimination. However, there is still a problem that we cannot assign the identity label to the local summing map. To deal with the problem, we connect all local summing maps to form the holistic summing map for a pedestrian image

$$Z = [Z_1; Z_2; \dots; Z_k; \dots; Z_K] \quad (5)$$

where K is the number of cropped regions in each pedestrian image, and the holistic summing map Z has the same identity label with the original image. The holistic summing maps contain both the local and global information of pedestrians, and they make possible to train the identification network in a local way. We take the holistic

summing map as the input of the identification network as shown in Fig. 4c. In the identification network, we also employ the cross-entropy loss

$$\text{Loss}_{id} = - \sum_{n=1}^N q_n \log(\tilde{q}_n) \quad (6)$$

where

$$\tilde{q}_n = \frac{e^{y_n}}{\sum_{j=1}^N e^{y_j}} \quad (7)$$

where $\tilde{q}_n \in [0, 1]$ is the predicted probability that the input holistic summing map belongs to the n -th class, y_n is the output of the fully connected layer, and q_n is the true class label. Assuming that t is the true identity label, q_n is defined as

$$q_n = \begin{cases} 0 & n \neq t \\ 1 & n = t \end{cases} \quad (8)$$

In the test stage, we first obtain the holistic summing maps from the verification network, and then feed them into the identification network. We also extract the LEDF from the fully connected layer of the identification network, and the dimension of the proposed LEDF is 2048-dim when the ResNet-50 [41] network is treated as the CNN model. The Euclidean distances between LEDF are sorted from gallery to the query to obtain the rank list.

4 Results and discussion

In this section, we evaluate the proposed LEDF on two large-scale datasets, Market-1501 [19] and CUHK03 [7]. During the experiments, we implement the verification network and the identification network using the Mat-ConvNet [42] package, and utilize the ResNet-50 [41] network as the CNN model for training and testing. For the verification network, we set the training epoch to 75, and set the learning rate to 0.001 and 0.0001 in the first 70 epochs and the final 5 epochs, respectively. The batch size is 24 and the ration between local similar pairs and local dissimilar pairs is 1:1 in the verification network. The identification network has the same epoch and learning rate with the verification network, and the batch size of the identification network is set to 48. The mini-batch stochastic gradient descent (SGD) is utilized to update the parameters and the weight decay is set to 0.0005 for two types of networks. There are four blocks in the ResNet-50 network, and each block contains nine convolution layers. We extract the local feature maps from the ninth convolution layer of the first block in the ResNet-50 network. The number of divided regions for each pedestrian image is the same, and each pedestrian image is divided into three regions. We resize each region to 224×224 as the input of verification network in pairs.

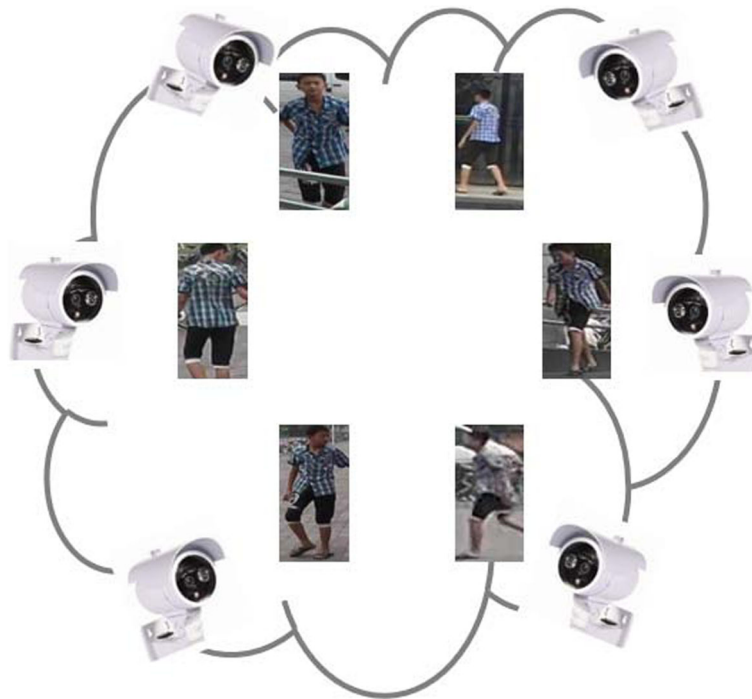


Fig. 5 The same pedestrian under camera networks suffers from large variations in illumination, poses and viewpoints

4.1 Datasets

Market-1501 dataset [19] is a large-scale person re-identification dataset. It contains 32,668 annotated bounding boxes of 1501 identities collected with six cameras. There are 12,936 pedestrian images of 751 identities for training set and 19,732 pedestrian images of 750 identities for test set. The Market-1501 dataset [19] also includes 3368 query images. For each query, we search the same pedestrian images from the test set. All images in this dataset are automatically detected by the Deformable Part Model (DPM) detector [43], which is closer to the realistic setting. Figure 5 shows that the same pedestrian under a camera network suffers from large variations in illumination, poses, and viewpoints.

CUHK03 dataset [7] contains 14,097 images of 1467 identities. Each identity is captured by two cameras on the CUHK campus and has 9.6 images in average. There are two kinds of bounding boxes, i.e., labeled bounding box and detected bounding box. We evaluate the proposed LEDF using the detected bounding box. The detected bounding box possesses some misalignments and part missing, which is closer to the realistic performance. According to the experimental setting in [6, 7], we partition this dataset into a training set of 1367 pedestrians and a test set of 100 pedestrians, and the performance of single-shot setting is reported with 20 random splits.

The rank-1 accuracy computed by the Euclidean distance and mean average precision (mAP) is utilized to

evaluate the performance of the proposed LEDF on Market1501 [19] and CUHK03 [7] datasets.

4.2 Evaluation

4.2.1 Comparison with the state-of-the-art algorithms

We compare the proposed LEDF with other state-of-the-art algorithms on the Market1501 dataset [19]. From Table 1, it can be seen that the proposed LEDF achieves 80.77% rank-1 accuracy and 61.34% mAP, outperforming the other precious works. Especially, the proposed LEDF

Table 1 Comparison with state-of-the-art results on the Market-1501 dataset [19]

Method	Rank-1	mAP
DADM [44]	39.4	19.6
BoW+KISSME [19]	44.42	20.76
Multiregion CNN [45]	45.58	26.11
Fisher Network [46]	48.15	29.94
CAN [47]	48.24	24.43
LS [48]	51.90	26.35
DNS [49]	55.43	29.87
Gate Reid [9]	65.88	39.55
Verif-identif [10]	79.51	59.87
LEDF	80.77	61.34

The rank-1 accuracy and mAP are listed. The italicized part shows the best results

Table 2 Comparison with state-of-the-art results on the CUHK03 dataset [7] using the single-shot setting

Method	Rank-1	Rank-5	Rank-10	mAP
KISSME [29]	11.7	33.3	48.0	–
DeepReID [7]	19.9	49.3	67.4	–
BoW+HS [12]	24.3	–	–	–
LOMO+XQDA [12]	46.3	78.9	88.6	–
SI+CI [50]	52.2	84.3	94.8	–
DNS [49]	54.7	80.1	88.3	–
Verif-identif [10]	83.4	97.1	98.7	86.4
LEDF	<i>84.6</i>	<i>96.8</i>	<i>98.5</i>	<i>86.9</i>

The rank-1, rank-5, and rank-10 accuracy and mAP are listed. The italicized part shows the best results

is 1.26 and 1.47% higher than the state-of-the-art method Verif-identif [10] in rank-1 accuracy and mAP, respectively. The two methods both utilize the verification network and identification network to learn the features, while the proposed LEDF explicitly considers the structural information of pedestrian.

Table 2 shows the results reported on the CUHK03 dataset [7] with the single-shot setting. There are only one correct image in the gallery, and the rank-1, rank-5, and rank-10 accuracy and mAP are listed. As shown in Table 2, the proposed LEDF yields 84.6, 96.8, and 98.5% at rank-1, rank-5, and rank-10 accuracy, respectively, which is comparable with the state-of-the-art methods. In addition, the mAP also achieves a new state of the art.

4.2.2 Evaluation with different inputs in the identification network

As shown in Table 3, we evaluate the performance of identification network with different inputs which contain the original pedestrian images (ResNet-50 (O)), the whole feature maps extracted from the original pedestrian images (ResNet-50 (W)), and the holistic summing maps (LEDF). Experimental results show that ResNet-50 (W) outperforms the ResNet-50 (O) because whole feature maps learned by CNN model contain high-level semantic information. From Table 3, it can be also seen that the proposed

Table 3 Results on Market-1501 [19] and CUHK03 [7] with different inputs in the identification network

Method	Market-1501		CUHK03	
	Rank-1	mAP	Rank-1	mAP
ResNet-50 (O)	73.69	51.48	71.5	75.8
ResNet-50 (W)	75.03	53.52	75.8	79.4
LEDF	<i>80.77</i>	<i>61.34</i>	<i>84.6</i>	<i>86.9</i>

“O” and “W” indicate the original pedestrian images and the whole feature maps extracted from the original pedestrian images, respectively. The italicized part shows the best results

Table 4 The rank-1 accuracy and mAP with varying number of cropped regions on the Market-1501 dataset

<i>K</i>	1	2	3	4	5
Rank-1	75.03	77.65	<i>80.77</i>	76.32	73.86
mAP	53.52	56.13	<i>61.34</i>	54.69	51.52

The italicized part shows the best results

LEDF achieves better results than ResNet-50 (W). It is because that the LEDF method could fully consider the structural information of pedestrian images by analyzing the local features in the holistic summing maps.

4.3 Evaluation on different number of regions

The purpose of dividing each pedestrian image into several regions is to fully consider the structural information. Hence, it is very important to determine the number of regions. In order to choose the optimal number of regions, we conduct experiments on the Market-1501 dataset based on the above-mentioned experimental setting, and the rank-1 accuracy and mAP with varying number of cropped regions are listed. From Table 4, it can be observed that training the identification network in a local way plays an essential part in improving rank accuracy and mAP, and we achieve the best result when $K = 3$.

5 Conclusions

In this paper, we have proposed a novel feature learning method named LEDF for person re-identification in camera networks, which trains the identification network in a local way. There are two characteristics of the proposed LEDF. First, we learn the local features by dividing each pedestrian image into several regions, and take pairs of regions as the input of the verification network to extract local summing maps. Second, all local summing maps of a pedestrian image are connected to form the holistic summing map. The holistic summing maps have the same labels with the original images, and contain the local and global information. We learn the identification network by taking the holistic summing maps as input, and then extract the final LEDF. The experimental results have indicated that the proposed LEDF improves the performance of person re-identification in camera networks.

Abbreviations

CNN: Convolutional neural network; LEDF: Local embedding deep features; mAP: Mean average precision; SGD: Stochastic gradient descent

Acknowledgements

The authors would like to thank the editor and anonymous reviewers for their helpful comments and suggestions in improving the quality of this paper.

Funding

This work was supported by National Natural Science Foundation of China under Grant No. 61711530240, Natural Science Foundation of Tianjin under Grant No. 17JCZDJC30600, the Fund of Tianjin Normal University under Grant

No. 135202RC1703, the Open Projects Program of National Laboratory of Pattern Recognition under Grant No. 201700001, and the China Scholarship Council No. 201708120040.

Availability of data and materials

The databases are available online.

Authors' contributions

ZZ contributed to the main idea and algorithm design of the study. MH conducted the experiments, and ZZ analyzed the results. ZZ and MH drafted and revised the manuscript. Both authors read and approved the final manuscript.

Authors' information

Zhong Zhang received the Ph.D. degree from Institute of Automation, Chinese Academy of Sciences. He is currently an Associate Professor at Tianjin Normal University.

Meiyan Huang is currently pursuing the M.S. degree at Tianjin Normal University.

Competing interests

The authors declare that they have no competing interests.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 15 March 2018 Accepted: 16 April 2018

Published online: 24 April 2018

References

1. T D'Orazio, G Cicirelli, in *IEEE International Conference on Image Processing*. People re-identification and tracking from multiple cameras: A review (IEEE, Orlando, 2012), pp. 1601–1604
2. F Zhao, X Sun, H Chen, R Bie, Outage performance of relay-assisted primary and secondary transmissions in cognitive relay networks. *EURASIP J. Wirel. Commun. Netw.* **2014**(1), 60 (2014)
3. MZ Hasan, H Al-Rizzo, M Günay, Lifetime maximization by partitioning approach in wireless sensor networks. *EURASIP J. Wirel. Commun. Netw.* **2017**(1), 15 (2017)
4. F Zhao, L Wei, H Chen, Optimal time allocation for wireless information and power transfer in wireless powered communication systems. *IEEE Trans. Veh. Technol.* **65**(3), 1830–1835 (2016)
5. WS Zheng, S Gong, T Xiang, in *Person Re-Identification*. Group association: assisting re-identification by visual context (Springer, London, 2014), pp. 183–201
6. L Zheng, Y Yang, AG Hauptmann, Person re-identification: past, present and future. *arXiv preprint arXiv:1610.02984* (2016)
7. W Li, R Zhao, T Xiao, X Wang, in *IEEE Conference on Computer Vision and Pattern Recognition*. Deepreid: deep filter pairing neural network for person re-identification (IEEE, Columbus, 2014), pp. 152–159
8. L Wu, C Shen, A Hengel, Personnet: person re-identification with deep convolutional neural networks. *arXiv preprint arXiv:1601.07255* (2016)
9. RR Viorio, M Haloi, G Wang, in *European Conference on Computer Vision*. Gated siamese convolutional neural network architecture for human re-identification (Springer, Cham, 2016), pp. 791–808
10. Z Zheng, L Zheng, Y Yang, A discriminatively learned cnn embedding for person re-identification. *ACM Trans. Multimed. Comput. Commun. Appl.* **14**(1), 13 (2017)
11. T Xiao, S Li, B Wang, L Lin, X Wang, End-to-end deep learning for person search. *arXiv preprint arXiv:1604.01850* (2016)
12. J Bromley, I Guyon, Y LeCun, E Säckinger, R Shah, in *Advances in Neural Information Processing Systems*. Signature verification using a siamese time delay neural network (NIPS, San Francisco, 1994), pp. 737–744
13. T Sikora, The mpeg-4 video standard verification model. *IEEE Trans. Circ. Syst. Video Technol.* **7**(1), 19–31 (1997)
14. KR Paap, SL Newsome, JE McDonald, RW Schvaneveldt, An activation–verification model for letter and word recognition: The word-superiority effect. *Psychol. Rev.* **89**(5), 573 (1982)
15. T Xiao, H Li, W Ouyang, X Wang, in *IEEE Conference on Computer Vision and Pattern Recognition*. Learning deep feature representations with domain guided dropout for person re-identification (IEEE, Las Vegas, 2016), pp. 1249–1258
16. R Girshick, in *IEEE International Conference on Computer Vision*. Fast r-cnn (IEEE, Santiago, 2015), pp. 1440–1448
17. D Heinke, GW Humphreys, Attention, spatial representation, and visual neglect: simulating emergent attention and spatial memory in the selective attention for identification model (saim). *Psychol. Rev.* **110**(1), 29 (2003)
18. A Abbott, D Collins, A theoretical and empirical analysis of a state of the art talent identification model. *High Ability Stud.* **13**(2), 157–178 (2002)
19. L Zheng, L Shen, L Tian, S Wang, J Wang, Q Tian, in *IEEE International Conference on Computer Vision*. Scalable person re-identification: A benchmark (IEEE, Santiago, 2015), pp. 1116–1124
20. F Zhao, W Wang, H Chen, Q Zhang, Interference alignment and game-theoretic power allocation in mimo heterogeneous sensor networks communications. *Signal Process.* **126**, 173–179 (2016)
21. F Zhao, B Li, H Chen, X Lv, Joint beamforming and power allocation for cognitive mimo systems under imperfect csi based on game theory. *Wirel. Pers. Commun.* **73**(3), 679–694 (2013)
22. DS Cheng, M Cristani, M Stoppa, L Bazzani, V Murino, Custom pictorial structures for re-identification. **2**(5), 6 (2011)
23. N Gheissari, TB Sebastian, R Hartley, in *IEEE Conference on Computer Vision and Pattern Recognition*. Person reidentification using spatiotemporal appearance (IEEE, New York, 2006), pp. 1528–1535
24. B Ma, Y Su, F Jurie, in *European Conference on Computer Vision*. Local descriptors encoded by fisher vectors for person re-identification (Springer, Berlin, 2012), pp. 413–422
25. S Liao, Y Hu, X Zhu, SZ Li, in *IEEE Conference on Computer Vision and Pattern Recognition*. Person re-identification by local maximal occurrence representation and metric learning (IEEE, Boston, 2015), pp. 2197–2206
26. F Zhao, H Nie, H Chen, Group buying spectrum auction algorithm for fractional frequency reuse cognitive cellular systems. *Ad Hoc Netw.* **58**, 239–246 (2017)
27. Y Hu, S Liao, Z Lei, D Yi, S Li, in *IEEE Conference on Computer Vision and Pattern Recognition*. Exploring structural information and fusing multiple features for person re-identification (IEEE, Portland, 2013), pp. 794–799
28. O Hamdoun, F Moutarde, B Stanculescu, B Steux, in *IEEE International Conference on Distributed Smart Cameras*. Person re-identification in multi-camera system by signature based on interest point descriptors collected on short video sequences (IEEE, Stanford, 2008), pp. 1–6
29. M Koestinger, M Hirzer, P Wohlhart, PM Roth, H Bischof, in *IEEE Conference on Computer Vision and Pattern Recognition*. Large scale metric learning from equivalence constraints (IEEE, Providence, 2012), pp. 2288–2295
30. J Ye, T Xiong, Q Li, R Janardan, J Bi, V Cherkassky, C Kambhampettu, in *ACM International Conference on Information and Knowledge Management*. Efficient model selection for regularized linear discriminant analysis (ACM, New York, 2006), pp. 532–539
31. M Hirzer, PM Roth, M Köstinger, H Bischof, in *European Conference on Computer Vision*. Relaxed pairwise learned metric for person re-identification (Springer, Berlin, 2012), pp. 780–793
32. WS Zheng, S Gong, T Xiang, in *IEEE Conference on Computer Vision and Pattern Recognition*. Person re-identification by probabilistic relative distance comparison (IEEE, Providence, 2011), pp. 649–656
33. KQ Weinberger, J Blitzer, LK Saul, in *Advances in Neural Information Processing Systems*. Distance metric learning for large margin nearest neighbor classification (NIPS, Vancouver, 2006), pp. 1473–1480
34. Z Li, S Chang, F Liang, TS Huang, L Cao, JR Smith, in *IEEE Conference on Computer Vision and Pattern Recognition*. Learning locally-adaptive decision functions for person verification (IEEE, Portland, 2013), pp. 3610–3617
35. E Ahmed, M Jones, TK Marks, in *IEEE Conference on Computer Vision and Pattern Recognition*. An improved deep learning architecture for person re-identification (IEEE, Boston, 2015), pp. 3908–3916
36. S Ding, L Lin, G Wang, H Chao, Deep feature learning with relative distance comparison for person re-identification. *Pattern Recog.* **48**(10), 2993–3003 (2015)
37. D Yi, Z Lei, S Liao, SZ Li, in *International Conference on Pattern Recognition*. Deep metric learning for person re-identification (IEEE, Stockholm, 2014), pp. 34–39
38. D Cheng, Y Gong, S Zhou, J Wang, N Zheng, in *IEEE Conference on Computer Vision and Pattern Recognition*. Person re-identification by

- multi-channel parts-based cnn with improved triplet loss function (IEEE, Las Vegas, 2016), pp. 1335–1344
39. L Zheng, Z Bie, Y Sun, J Wang, C Su, S Wang, Q Tian, in *European Conference on Computer Vision*. Mars: A video benchmark for large-scale person re-identification (Springer, Cham, 2016), pp. 868–884
 40. S Wu, YC Chen, X Li, AC Wu, JJ You, WS Zheng, in *IEEE Winter Conference on Applications of Computer Vision*. An enhanced deep feature representation for person re-identification (IEEE, Lake Placid, 2016), pp. 1–8
 41. K He, X Zhang, S Ren, J Sun, in *IEEE Conference on Computer Vision and Pattern Recognition*. Deep residual learning for image recognition (IEEE, Las Vegas, 2016), pp. 770–778
 42. A Vedaldi, K Lenc, in *ACM International Conference on Multimedia*. Matconvnet: Convolutional neural networks for matlab (ACM, New York, 2015), pp. 689–692
 43. P Felzenszwalb, D McAllester, D Ramanan, in *IEEE Conference on Computer Vision and Pattern Recognition*. A discriminatively trained, multiscale, deformable part model (IEEE, Anchorage, 2008), pp. 1–8
 44. C Su, S Zhang, J Xing, W Gao, Q Tian, in *European Conference on Computer Vision*. Deep attributes driven multi-camera person re-identification (Springer, Cham, 2016), pp. 475–491
 45. E Ustinova, Y Ganin, V Lempitsky, in *IEEE International Conference on Advanced Video and Signal Based Surveillance*. Multi-region bilinear convolutional neural networks for person re-identification (IEEE, Lecce, 2017), pp. 1–6
 46. L Wu, C Shen, VDA Hengel, Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification. *Pattern Recognit.* **65**, 238–250 (2017)
 47. H Liu, J Feng, M Qi, J Jiang, S Yan, End-to-end comparative attention networks for person re-identification. *IEEE Trans. Image Process.* **26**(7), 3492–3506 (2017)
 48. D Chen, Z Yuan, B Chen, N Zheng, in *IEEE Conference on Computer Vision and Pattern Recognition*. Similarity learning with spatial constraints for person re-identification (IEEE, Las Vegas, 2016), pp. 1268–1277
 49. L Zhang, T Xiang, S Gong, in *IEEE Conference on Computer Vision and Pattern Recognition*. Learning a discriminative null space for person re-identification (IEEE, Las Vegas, 2016), pp. 1239–1248
 50. F Wang, W Zuo, L Lin, D Zhang, L Zhang, in *IEEE Conference on Computer Vision and Pattern Recognition*. Joint learning of single-image and cross-image representations for person re-identification (IEEE, Las Vegas, 2016), pp. 1288–1296

Submit your manuscript to a SpringerOpen[®] journal and benefit from:

- Convenient online submission
- Rigorous peer review
- Open access: articles freely available online
- High visibility within the field
- Retaining the copyright to your article

Submit your next manuscript at ► [springeropen.com](https://www.springeropen.com)