Cluster Computing (2019) 22:S8975–S8984
https://doi.org/10.1007/s10586-018-2033-2

Dynamic locally connected layer for person re-identification

Faping Li[1] · Fabing Li[2] · Haizhu Chen[1]

Received: 5 December 2017 / Revised: 15 January 2018 / Accepted: 7 February 2018 / Published online: 27 February 2018

Abstract
Person re-identification is a challenging task due to its large variations on pedestrian pose, camera view, lighting and background. To solve pedestrian misalignment problem, most of the existing works assume that the pedestrian images are horizontally aligned so that the extracted features can be compared correspondingly. However, such assumption is not necessarily true in reality because the pedestrians may be misaligned vertically. To address the misalignment problem, we propose a dynamic locally connected (DLC) layer based on convolutional neural network (CNN). We use human parsing tool to get parsing results of pedestrian images, then map the results to the last feature map of our CNN. By doing this, proposed model is able to locate the human body parts dynamically within DLC layer, thus leads to a more accurate matching on local features. Furthermore, we adopt pre-training with two-step fine-tuning strategy on the small person re-identification datasets, which again boost the model performance. According to the experiments, proposed model achieves competitive results among the state-of-the-art models on four popular person re-identification datasets.

Keywords Person re-identification · Dynamic locally connected layer · Convolutional neural network

1 Introduction

Person re-identification is a task that identifies the same person across long-term multiple cameras [1]. With the development of public monitoring system, it has been in an increasing demand to search person from camera images given query images. The applications such as video surveillance [2] and forensic search [3] are receiving more attention. However, due to the large variations on lighting, pose, occlusion, low resolution and background, the re-identification task has been challenging for industrial use. Earlier works usually divide the task into two parts: (1) hand-crafted features, such as color histograms [4] and texture (LBP [5], Gabor [6], are first designed to transform images to vector representations. (2) A metric learning algorithm is then applied to learn a discriminative space [7–12]. Unfortunately, because of the great variability of the input images, designing a robust and effective feature by hand is difficult and unsustainable.

In recent years, convolutional neural networks(CNNs) have been popular among various computer vision tasks, including person re-identification [13–18]. CNN models outperform traditional methods consistently, especially when large datasets are available. The success of CNNs is owing to their great ability on feature extraction, which is learned end-to-end by neural networks. As for person re-identification task, the CNN models are usually adopted from conventional computer vision tasks such as image classification and face recognition. By adding task specific components, CNN models achieve state-of-the-art performances on most person re-identification datasets. In the field of face recognition [19], the standard data pre-process pipeline includes face alignment generally, and a good align is of great benefit to the final performance. However, few of the existing works pay attention to the misalignment problem in person re-identification. To deal with pose variations, many works assume that the pedestrian images are horizontally aligned, which is not necessarily

✉ Faping Li
DataCleaning@163.com

Fabing Li
27201383@qq.com

Haizhu Chen
82777460@qq.com

[1] ChongQing College of Electronic Engineering, Chongqing 401331, China

[2] ChongQing YuBei Experimental Primary School, Chongqing 401120, China

guaranteed in practice. Figure 1 gives several pedestrian images that are not well aligned. Consequently, such strict assumption may affect the performance of CNN models because of misalignment problem.

To address the problem mentioned above, we propose a novel CNN layer in this paper, named dynamic locally connected (DLC) layer. The DLC layer is a variant of locally connected (LC) layer that first proposed in [20]. Cheng et al. [21] splitted pedestrian images into five horizontal parts to cope with pose variations, which is equivalent to applying LC layer to CNN. As mentioned before, the limit of such method is that the pedestrian images are not strictly aligned horizontally. Therefore, we propose DLC layer based on LC layer, aiming to alleviate the misalignment problem in the re-identification task. Conventional methods such as LC layer are only able to match features from the same area while DLC layer is able to match features from different areas, which is reasonable in re-identification task. We first get the parsing results of pedestrian images by the semantic parsing tool [22], then map the results to three bounding boxes corresponding to the three parts of human body. When it comes to DLC layer, three LC layers conduct operations within the three bounding boxes respectively. By doing this, our CNN model is able to locate human body parts dynamically, and extract features more accurately for metric learning.

Fig. 1 Misaligned pedestrian images from person re-identification datasets. Each column presents the two images of the same identity. It is easy to see that the misalignment problem is unneglectable, otherwise the model accuracy would be affected considerably

Through experiments, we show that the CNN with DLC layer outperforms the baseline models on all the datasets we use, and our proposed model achieves competitive results among state-of-the-arts.

The rest of the paper is organized as follows. We review the related works in Sect. 2. The proposed CNN model is presented in Sect. 3. Section 4 gives the experimental results. At last, Sect. 5 draws a conclusion of the paper.

2 Related work

To tackle person re-identification problems, many proposed models can be roughly divided into two categories: traditional models and deep learning models.

Traditional models Many traditional methods have been proposed to address person re-identification problems. Most of these methods focus on designing a hand-crafted feature representation or a metric learning method for person re-identification.

Hand-crafted feature representation designing is aimed at generating discriminative descriptors for pedestrian images. Liao et al. [7] 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 robust representation against viewpoint changes. To use colors as feature, Yang et al. [23] proposed a color descriptor named salient color names based color descriptor (SCNCD), utilizing salient color names to guarantee that a higher probability will be assigned to the color name which is nearer to the color. To automatically discover patch clusters, Zhao et al. [24] proposed to learn mid-level filters, which are discriminatively learned for identifying specific visual patterns and distinguishing persons, and have good cross-view invariance. A approach is proposed in the work [25] to match images observed in different camera views with complex cross-view transforms. The image representation proposed in work [26] handles both background and illumination variations properly. Khamis et al. [27] proposed to learn a discriminative projection to a joint appearance-attribute subspace instead of only learning the appearance feature. However, due to the large changes on lighting, pose, occlusion, low resolution and background, representation of visual appearance are highly susceptible to these variations. Therefore, achieving a balance between discriminative power and robustness is difficult.

To re-identification problems, metric learning based methods usually follow a similar procedure: first extracting representations from each image, then learning a metric on training data, minimizing inter-category differences and maximizing intra-category similarities.

Li et al. [28] proposed the locally-adaptive decision functions (LADF), combining the distance metric with a locally adaptive thresholding rule for each pair of sample images. KISSME [29] employed a Mahalanobis metric by computing the difference between the intra-class and inter-class covariance matrix, which do not rely on complex optimization problems requiring computationally expensive iterations. Paisitkriangkrai et al. [8] proposed an approach based on structure learning, and utilize two optimization algorithms to directly optimize cumulative matching characteristic (CMC) curve, which is a common evaluation measure in person re-identification. Zhang et al. [9] matched people in a discriminative null space of the training data, trying to overcome the classic small sample size (SSS) problem. Zhang et al. [10] regarded the re-identification task as an imbalanced classification problem, thus they proposed to learn a classifier specifically for each pedestrian. A similarity function was learned in [11], which includes multiple sub-similarity measurements. Liao et al. [7] proposed a subspace and metric learning method called cross-view quadratic discriminant analysis (XQDA). Liao and Li [12] proposed an asymmetric sample weighting strategy to solve a logistic metric learning problem.

Deep learning models Recently, due to its powerful capability of representation extraction, CNN-based models have shown great potential in several computer vision tasks, such as image classification [30], face verification [31], object detection [32]. And more CNN based methods have been proposed to tackle re-identification problem, and the overall performance are improved by a large margin benefiting from the strong representation ability of CNN.

To improve the CNN embeddings, Li et al. [13] proposed one patch matching layer and one maxout-grouping layer. The patch matching layer is used to discriminate the horizontal stripes in across-view images, while the maxout-grouping layer is used to improve the performance of patch matching. Yi et al. [14] designed a Siamese CNN architecture, where the input image is divided into three areas horizontally, and then the three parts are merged to compute a similarity score between images. Ahmed et al. [15] designed a layer to discriminate cross-input neighborhoods so that the local relationships can be captured. They also propose a patch summary layer to summarize the features learned from the previous layers. Wang et al. [16] proposed to learn single-image and cross-image representation jointly to achieve a more robust feature. Xiao et al. [17] proposed a CNN that learns features from multiple domains with a softmax loss, where the main component is called domain guided dropout. Cheng et al. [21] designed a multi-channel parts-based CNN to learn both the global and the local features with an improved triplet loss function. Varior et al. [18] proposed a gating function for CNN

to capture effective subtle patterns. Varior et al. [33] integrated the long short-term memory (LSTM) modules into a Siamese CNN, which is able to process image parts sequentially and emphasize contextual information for learning a better feature. A moderate positive sample mining approach was proposed by Shi et al. [34] so that the learned feature is not sensitive to pose variations. A network named PersonNet was proposed by Wu et al. [35], which uses deeper networks with smaller filter size. The works [36, 37] prove that the combination of identification loss and verification loss is better than either of them. Sun et al. [38] proposed to add an Eigenlayer before the last fully connected (FC) layer to improve the discriminative performance of the model for person re-identification. Chung et al. [39] proposed a two stream convolutional neural network where each stream is a Siamese network to learn spatial and temporal information separately, and an objective combining the Siamese cost of the spatial and temporal streams with the objective of predicting a persons identity. Yu et al. [40] proposed an unsupervised asymmetric metric learning method to learn an asymmetric metric, i.e., specific projection for each view, based on asymmetric clustering on cross-view person images.

Although many existing algorithms which exploited state-of-the-art feature representation or effective metric learning models were proposed, the overall performance on popular datasets e.g., Market1501 and VIPeR still has a big improvement for real-world or industry applications.

3 Proposed model

In this section, we will introduce the model we use in this paper. Firstly, we introduce the basic CNN structure and loss functions of proposed model. Secondly, we present the proposed DLC layer by comparing it to FC and LC layer.

3.1 The CNN architecture

We use vgg16 as our base network due to its success on ImageNet. The network we use is pre-trained on ImageNet to get a good initialization and reduce overfitting. There are two FC layers in the original vgg16 network. When we adopt it to our task, we modify the second FC layer to DLC layer.

Proposed model consists of two types of loss functions: identification loss and verification loss. From the works [36, 37], the combination of the two losses has been shown to be superior to either one of them on person re-identification task. Thus, we apply the two losses to our task. More specifically, identification loss is actually softmax loss, defined as:

$$\hat{p} = softmax(\theta_I \circ f), \tag{1}$$

$$Iden(f, t, \theta_I) = \sum_{i=1}^{K} -p_i \log(\hat{p_i}), \tag{2}$$

where $\circ$ means convolution, f is input and θ_I is parameter. $\hat{p}$ is the probability predicted by the model, and p_i is the target probability.$p_i = 0$ for all i except $p_t = 1$.

Verification loss is actually a binary classifier that identifies if a pair of images belong to the same identity, which is defined as follows:

$$\hat{q} = softmax(\theta_S \circ (f_1 - f_2)), \tag{3}$$

$$Veri(f_1, f_2, s, \theta_S) = \sum_{i=1}^{2} -q_i log(\hat{q_i}), \tag{4}$$

where f_1 and f_2 are the outputs of the two base networks, and θ_S is the convolution parameter. $\hat{q_i}$ is the predicted probability and q_i is the target probability. Because it is a binary classifier, there are only two classes for softmax.

Figure 2 shows the architecture of our CNN model. The model is a siamese CNN network, where the two subnets share the parameters. There is an identification loss for each subnet while the verification loss is shared by the two. Before the verification loss, we do subtraction on the two outputs from the subnets. When training, we feed equal positive and negative image pairs for network optimization.

3.2 Dynamic locally connected layer

Fully connected layer FC layer is frequently used in neural networks, including CNNs. It connects each node of the previous layer to every node of the current layer, which yields considerable amount of parameters. Therefore, FC

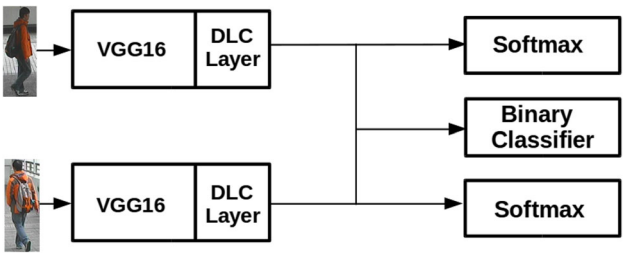


Fig. 2 CNN architecture of proposed model. The model is a siamese CNN, which consists of two base networks sharing the same parameters. In our experiments, we use vgg16 as our base network because of its popularity. The second FC layer of the original vgg16 is replaced by the proposed DLC layer. There is a softmax classifier for each base network, and one binary classifier for the whole siamese network. Before the binary loss, we do subtraction on the two outputs from the base networks. During training, image pairs with positive or negative labels are fed to the siamese network for loss optimization

layer may suffer from overfitting problem when not used properly. Figure 3a illustrates the topology of a FC layer.

Locally connected layer LC layer is first applied to face recognition task by Taigman et al. [20]. The goal of LC layer is to improve the efficiency of filters by having different filters spatially. In terms of person re-identification, LC layer is used to focus on local features so that the model becomes more robust on local pattern matching. Figure 3b shows the topology of a LC layer.

Dynamic locally connected layer Although LC layer considers local features to improve the robustness of model, its performance is limited. Because the real pedestrian images are not necessarily aligned horizontally, applying LC layer still faces the possibility of misalignment. In order to improve the misalignment problem, we propose DLC layer to replace LC layer. To be more specific, we first use a pedestrian parsing tool to get the location of human bodies. We split a pedestrian into three parts, namely head, upper body and lower body. With the help of the parsing tool, we are able to map the three body parts to the last feature map of our CNN. By doing this, the body parts can be located on the feature map dynamically, which yields a more accurate alignment of human body parts. After that, local features of the three body parts can be extracted respectively by LC layer. In other words, DLC layer is an improved LC layer by dynamically tracking three human body parts using pedestrian parsing tool. Figure 3c shows the topology of a LC layer. To give a more detailed illustration on the difference between the three layers, we present the visualization maps on weights connections based on real pedestrian images in Figure 4.

3.3 Training settings

In our paper, we use Caffe [41] to implement our proposed methods and more details about training strategies are discussed in this subsection.

3.3.1 Pre-training

For deep CNN model, a large amount of training data is needed for learning. But the labeled data in scenarios of person re-identification are scarce, we use the labeled data from other scenarios to assist vgg16 network learning, even though the distribution of datasets is quite different. During our experiment, we first utilize large-scale dataset such as ImageNet to learn a pre-trained model, then use that pre-trained model as an model initialization for re-identification dataset. Note that newly added layers are randomly initialized with zero mean and 0.01 standard deviation Gaussian distribution. We find that pre-training is an effective strategy to boost the final performance.

3.3.2 Data augmentation

It is a common method to artificially enlarge the training data using label-preserving transformations to reduce overfitting. We also conduct data augmentation during training

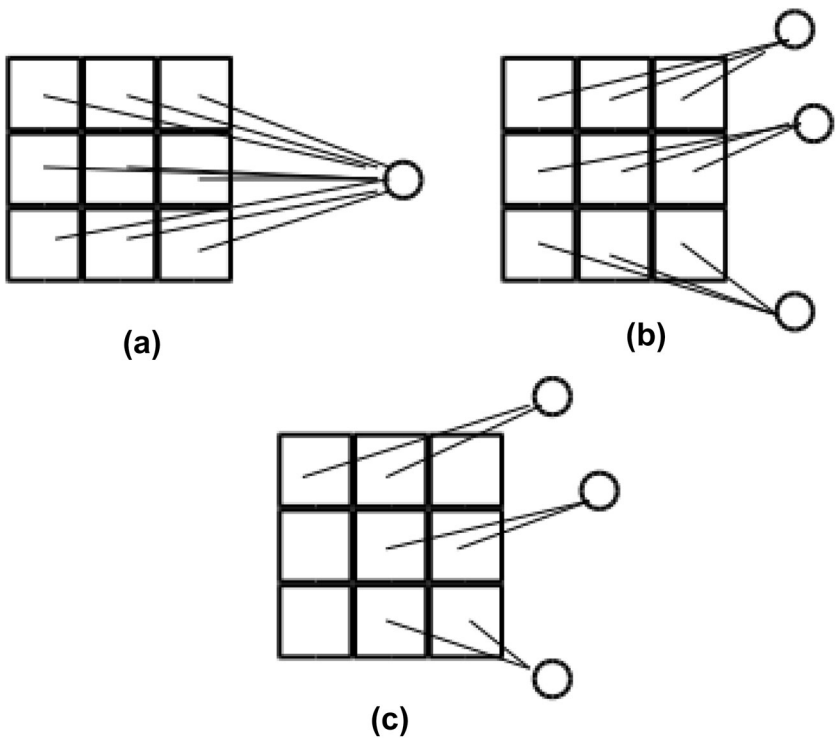


Fig. 3 Topology graphs of FC, LC and DLC layer. **a** FC layer simply connects all the nodes of the previous layer to each node of the next layer. **b** LC layer splits the feature map to several fixed areas in advance(in our case, we split the feature map to three areas vertically), so the connections are only within the same area between two layers. **c** DLC layer can be seen as the dynamic version of LC layer. With the help of human parsing tool, DLC layer is able to track certain parts, then map it onto feature map, which forms the local areas dynamically rather than fixed

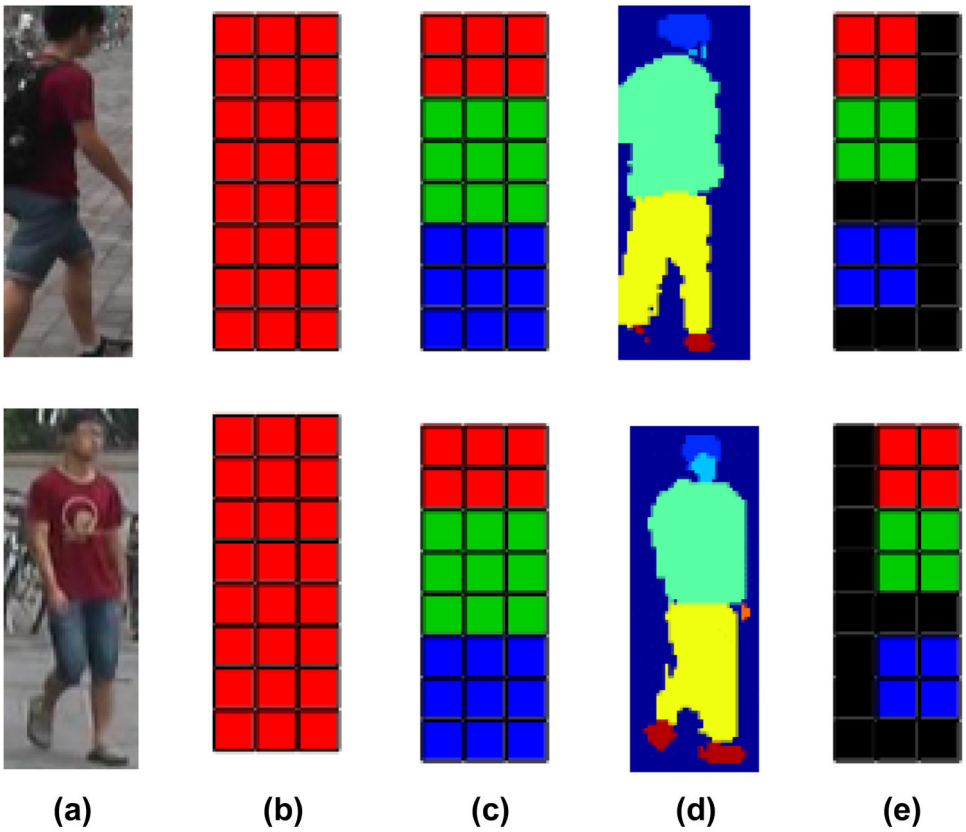


Fig. 4 Comparison of the three layers on the real pedestrian images. **a** Two images of the same identity, with different backgrounds, poses and camera views. The two images are vertically misaligned. **b** Two feature maps of the two input images. When using FC layer, all the nodes on the feature map are connected to the next layer, same for the two maps. **c** When using LC layer, the feature maps are divided into three fixed horizontal areas, indicated by three colors. **d** The parsing results of the two input images, using human parsing tool. **e** The proposed DLC tracks three human body parts according to the parsing results, also getting three areas for local connection. Different from LC layer, DLC layer outputs different area splits if the two images are not well aligned, which is more accurate and robust

to improve the generalization ability of the models. Data augmentation includes two main operations: horizontal flat and random translation. Random translation contains two dimensions, namely horizontal and vertical. The maximum random translation pixels for the two dimensions are 2 and 6 respectively. The mean of the deep person re-identification training data are subtracted to make the optimization easier.

3.3.3 Training strategy

The number of total training iterations is 30k. We set the initial learning rate as 0.001, then decay it by 0.1 after 20K iterations. We set the batch size as 64, and the weight decay as 0.0005. The optimization method is SGD and the momentum is 0.9. Since we use siamese network, we keep the positive and negative ratio of input image pairs as 1:1 to avoid negative sample overwhelming problem.

4 Experiments

We validate the effectiveness of proposed model on four popular person re-identification datasets: Market1501, CUHK03, CUHK01 and VIPeR.

4.1 Datasets

Market1501 Market1501 includes 32668 images of 1501 identities, each identity of them is captured by two cameras at least, and six cameras at most. Among the 1501 identities, 751 of them are training set and the rest 750 are for testing. During testing, only one query image in each camera is chosen for each identity, hence multiple queries are selected for each identity. The pedestrian images of the dataset are captured by Deformable Part Model rather than human labeled, which is close to the scenario in [42] practice. But note that, the selected 3368 queries are hand-drawn, instead of Deformable Part Model as in the gallery. We adopt the provided fixed training and test set, under both the single-query and multi-query evaluation settings.

CUHK03 CUHK03 contains 13163 images of 1360 identities captured from CUHK campus. The dataset are collected by two cameras, where each identity has 4.8 images on average. It has two settings, namely human labeled and model detected. We choose the detected setting because it is close to reality. According to the official protocol, there are 20 splits for training and testing, and the final performance is the average of the 20 splits.

CUHK01 CUHK01 consists of 971 identities collected from CUHK campus. It has two camera views, and each identity of each view contains two images. All the images

are normalized to size 160×160 . Following the standard setting, images from camera A are used as probe and those from camera B as gallery. Similar to CUHK03, its performance is based on 10 random splits, among which 485 identities are for training and 486 of them are for testing.

VIPeR VIPeR contains 632 identities, where each identity has two camera views and one image for each view. All images are scaled to 128×48 pixels. The evaluation is also based on 10 random splits. Half of the dataset are used for training, and another half are for testing.

In our experiments, the images of the four datasets are all resized to 128×48 for implementation convenience.

4.2 Evaluation protocol

Cumulated matching characteristics (CMC) curve is used to evaluate the performance of person re-identification methods for all datasets in this paper. Due to limited space and for simple comparison with published results, we only cover the cumulated matching accuracy at rank-1 in tables rather than plot the complete CMC curves (Fig. 5).

4.3 Comparison with baselines

Table 1 shows the results of our proposed model against its baselines. From the table, we can see that our proposed model outperforms the two baselines on all the datasets in this paper as expected. Also, the model with LC layer has better performance than the FC version because the LC model considers both global feature and local feature. Through this baseline experiment, we show that the proposed DLC layer is effective and robust on pedestrian matching task.

4.4 Comparison with state-of-the-arts

Here, we compare proposed model to the existing state-of-the-art methods on the four re-identification datasets (please see Table 2). The table indicates that proposed model is competitive among those state-of-the-art models. Proposed model with DLC layer achieves the best rank-1 accuracy on Market1501 and CUHK03 data, being 68.2 and 70.1% respectively. On CUHK01 and VIPeR, proposed model also gets comparable results compared to state-of-the-art models. Because CUHK01 and VIPeR are two relatively small datasets, which contain not enough images for our CNN to generalize better. If other big re-identification datasets can be used as pre-training, the accuracy shall increase considerably. To validate the effectiveness of pre-training, we further conduct experiments with data pre-training for relatively small person re-identification datasets.

From the results of Table 2, we see that CUHK01 and VIPeR do not get as good accuracy as CUHK03 and

Fig. 5 The four datasets we use in this paper. We present two identities for each dataset. There are two images from two different camera views for each identity

Table 1 Rank-1 results against baselines

Method	Market1501	CUHK03	CUHK01	VIPeR
Proposed model (FC)	65.0	66.8	65.2	45.0
Proposed model (LC)	67.7	69.1	66.3	47.1
Proposed model (DLC)	**68.2**	**70.1**	**68.0**	**49.9**

Bold values indicate the best result in corresponding column

Table 2 Rank-1 results against state-of-the-arts

Method	Market1501	CUHK03	CUHK01	VIPeR
XQDA [7]	43.79	46.3	–	–
MLAPG [12]	–	51.2	–	–
DNS [9]	61.02	54.7	**69.1**	51.2
LSSCDL [10]	–	51.2	–	42.7
Siamese LSTM [33]	61.6	57.3	–	42.4
EDM [34]	–	52	–	40.9
Joint learning [16]	–	52.2	–	35.8
IDLA [15]	–	45	47.5	34.8
Gated S-CNN [18]	65.88	61.8	–	37.8
CAN [43]	48.24	63.1	–	–
CNN embedding [37]	–	66.1	–	–
SCSP [11]	51.9	–	–	**53.5**
DGD [17]	–	–	66.6	38.6
MCP-CNN [21]	–	–	53.7	47.8
Proposed model (DLC)	**68.2**	**70.1**	68.0	49.9

Bold values indicate the best result in corresponding column

Market1501 because they contains less data for deep learning models. Following the work [36], we first pretrain our base network on the combination of CUHK03 and Market1501, then fine tune it on the two small datasets to boost their performance. To have a more effective transfer learning performance, we use a two-stepped fine-tuning strategy from work [36]. When doing fine-tuning, the target small dataset and the pre-training data are different on class identities, so the weights of the softmax classifier cannot be used any more in the fine-tuning phase. Consequently, we replace the weights of the softmax classifier by a randomly initialized weight. The new weight has N_s nodes, which is the number of class identities of the small dataset. After that, the newly added weights are fine-tuned until the classifier converges while the other parameters are freezed. Then, we do second-step fine-tuning by updating all the parameters altogether. According to work [36], the reason of doing two-step fine-tuning is to prevent the newly added weights from backpropagating wrong gradients to the already pre-trained parameters.

Table 3 shows the the results on dataset CUHK03 and VIPeR with pre-training on CUHK03 and Market1501. Now both of the small datasets achieve the state-of-the-art rank-1 accuracy, which verifies the advantages of pre-training and two-step fine-tuning strategy. It is worth noting that VIPeR gets more rank-1 accuracy improvement because its data is smaller on both identity quantity and image quantity. Therefore, such pre-training strategy will be more significant on small datasets.

Table 3 Rank-1 results against state-of-the-arts, pre-trained on CUHK03 and Market1501

Method	CUHK01	VIPeR
DNS [9]	69.1	51.2
LSSCDL [10]	–	42.7
Siamese LSTM [33]	–	42.4
EDM [34]	–	40.9
Joint Learning [16]	–	35.8
IDLA [15]	47.5	34.8
Gated S-CNN [18]	–	37.8
SCSP [11]	51.9	53.5
DGD [17]	66.6	38.6
MCP-CNN [21]	53.7	47.8
Proposed model (DLC)	**70.4**	**53.7**

Bold values indicate the best result in corresponding column

5 Conclusion

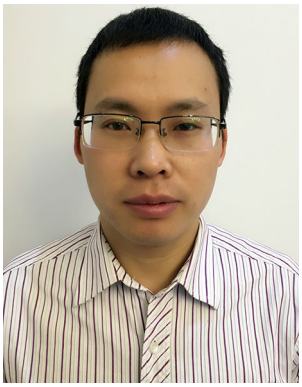
The main contributions of this paper are to propose a novel DLC layer and adopt two-step fine-tuning strategy. The novel CNN layer named dynamic locally connected (DLC) layer to address pedestrian misalignment problem. The proposed layer can locate the human body parts dynamically within DLC layer. In addition, we adopt two-step fine-tuning strategy on the small person re-identification datasets with DLC layer. Compared with existing re-id models, our proposed model achieve competitive results among the state-of-the-art models on four popular person re-identification datasets.

Acknowledgements This work is supported by Scientific Research Project of Chongqing Education Commission (No. KJ1729408) and Teaching Reform Research Project of Chongqing Education Commission (No. 162071).

References

1. Song, B., Kamal, A.T., Soto, C., Ding, C., Farrell, J.A., Roychowdhury, A.K.: Tracking and activity recognition through consensus in distributed camera networks. IEEE Trans. Image Process. **19**(10), 2564–2579 (2010)
2. Gong, S., Cristani, M., Loy, C.C., Hospedales, T.M.: The Reidentification Challenge in Person Re-identification. Springer, London (2014)
3. Vezzani, R., Davide, B., Cucchiara, R.: People reidentification in surveillance and forensics: a survey. ACM Comput. Surv. (2013). https://doi.org/10.1145/2543581.2543596
4. Mignon, A., Jurie, F.: PCCA: a new approach for distance learning from sparse pairwise constraints. In: CVPR (2016)
5. Li, W., Wang, X.: Locally aligned feature transforms across views. In: CVPR (2013)
6. Zheng, W., Gong, S., Xiang, T.: Reidentification by relative distance comparison. IEEE Trans. Pattern Anal. Mach. Intell. **35**(3), 653–668 (2013)
7. Liao, S., Hu, Y., Zhu, X., Li, S.: Person re-identification by local maximal occurrence representation and metric learning. In: CVPR (2015)
8. Paisitkriangkrai, S., Shen, C., Hengel, A.: Learning to rank in person re-identification with metric ensembles. In: CVPR (2015)
9. Zhang, L., Xiang, T., Gong, S.: Learning a discriminative null space for person re-identification. In: CVPR (2016)
10. Zhang, Y., Li, B., Lu, H., Irie, A., Ruan, X.: Sample-specific SVM learning for person re-identification. In: CVPR (2016)
11. Chen, D., Yuan, Z., Chen, B., Zheng, N.: Similarity learning with spatial constraints for person re-identification. In: CVPR (2016)
12. Liao, S., Li, S.: Efficient PSD constrained asymmetric metric learning for person re-identification. In: ICCV (2015)
13. Li, W., Zhao, R., Xiao, T., Wang, X.: DeepReID: deep filter pairing neural network for person re-identification. In: CVPR (2014)
14. Yi, D., Lei, Z., Liao, S., Li, S.: Deep metric learning for person re-identification. In: ICPR (2014)
15. Ahmed, E., Jones, M., Marks, T.: An improved deep learning architecture for person re-identification. In: CVPR (2015)
16. Wang, F., Zuo, W., Lin, L., Zhang, D., Zhang, L.: Joint learning of single-image and cross-image representations for person re-identification. In: CVPR (2016)
17. Xiao, T., Li, H., Ouyang, W., Wang, X.: Learning deep feature representations with domain guided dropout for person re-identification. In: CVPR (2016)
18. Varior, R., Haloi, M., Wang, G.: Gated siamese convolutional neural network architecture for human re-identification. In: ECCV (2016)
19. Sun, Y., Wang, X., Tang, X.: Deep learning face representation from predicting 10,000 classes. In: CVPR (2014)
20. Taigman, Y., Yang, M., Ranzato, M.A., Wolf, L.: DeepFace: closing the gap to human-level performance in face verification. In: CVPR (2014)
21. Cheng, D., Gong, Y., Zhou, S., Wang, J., Zheng, N.: Person re-identification by multi-channel parts-based CNN with improved triplet loss function. In: CVPR (2016)
22. Luo, P., Wang, X., Tang, X.: Pedestrian parsing via deep decompositional network. In: ICCV (2013)
23. Yang, Y., Yang, J., Yan, J., Liao, S., Yi, D., Li, S.: Salient color names for person re-identification. In: ECCV (2014)
24. Zhao, R., Ouyang, W., Wang, X.: Learning mid-level filters for person re-identification. In: CVPR (2014)
25. Li, W., Wang, X.: Locally aligned feature transforms across views. In: CVPR (2013)
26. Ma, B., Su, Y., Jurie, F.: A novel image representation for person re-identification and face verification. In: BMVC (2012)
27. Khamis, S., Kuo, C.-H., Singh, V., Shet, V., Davis, L.: Joint learning for attribute-consistent person re-identification. In: ECCV (2014)
28. Li, Z., Chang, S., Liang, F., Huang, T.S., Cao, J.R.: Learning locally-adaptive decision functions for person verification. In: CVPR, Liangliang and Smith (2013)
29. Kostinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: CVPR (2012)
30. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: NIPS (2012)
31. Sun, Y., Chen, Y., Wang, X., Tang, X.: Deep learning face representation by joint identification-verification. In: NIPS (2014)
32. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: CVPR (2014)
33. Varior, R., Shuai, B., Lu, J., Xu, D., Wang, G.: A siamese long short-term memory architecture for human re-identification. In: ECCV (2016)
34. Shi, H., Yang, Y., Zhu, X., Liao, S., Lei, Z., Zheng, W., Li, S.: Embedding deep metric for person re-identification: a study against large variations. In: ECCV (2016)
35. Wu, L., Shen, C., Hengel, A.V.D.: PersonNet: person re-identification with deep convolutional neural networks (2016). arXiv:1601.07255
36. Geng, M., Wang, Y., Xiang, T., Tian, Y.: Deep transfer learning for person re-identification (2016). arxiv:1611.05244
37. Zheng, Z., Zheng, L., Yang, Y.: A discriminatively learned CNN embedding for person re-identification (2016). arXiv:1611.05666
38. Sun, Y., Zheng, L., Deng, W., Wang, S.: Svdnet for pedestrian retrieval (2017). arXiv:1703.05693
39. Chung, D., Tahboub, K., Delp, E.J.: A two stream siamese convolutional neural network for person re-identification. In: CVPR (2017)
40. Yu, H.X., Wu, A., Zheng, W.S.: Cross-view asymmetric metric learning for unsupervised person re-identification. In: CVPR (2017)

41. Jia, Y., Shelhamer, E., Donahue, J., Karayev, S., Long, J., Girshick, R.B., Guadarrama, S., Darrell, T.: Convolutional architecture for fast feature embedding. In: ACMMM, Caffe (2014)
42. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: a benchmark. In: ICCV (2015)
43. Liu, H., Feng, J., Qi, M., Jiang, J., Yan, S.: End-to-end comparative attention networks for person re-identification. IEEE Trans. Image Process. **26**(7), 3492–3506 (2017)



Faping Li received the M.S. degree in software engineering from Chongqing University, China, in 2009. His current research interest includes software engineering.

Fabing Li received the B.S. degree in computer science in 2003. His current research interest includes software engineering and educational informationization.

Haizhu Chen received the Ph.D. degree in computer science from Chongqing University, China, in 2011. Her current research interests include membranecomputing and optimization algorithm.