Local Heterogeneous Features for Person Re-Identification in Harsh Environments

HAIJIA ZHANG¹, TONGZHEN SI¹, ZHONG ZHANG¹, (Senior Member, IEEE), RONGHUA ZHANG², HAO MA¹, AND SHUANG LIU¹, (Senior Member, IEEE)

¹Tianjin Key Laboratory of Wireless Mobile Communications and Power Transmission, Tianjin Normal University, Tianjin 300387, China
²College of Information Science and Technology, Shihezi University, Shihezi 832003, China

Corresponding author: Zhong Zhang (zhong.zhang8848@gmail.com)

This work was supported in part by National Natural Science Foundation of China under Grant No. 61711530240, in part by Natural Science Foundation of Tianjin under Grant No. 19JCZDJC31500, in part by the Open Projects Program of National Laboratory of Pattern Recognition under Grant No. 202000002, and in part by the Tianjin Higher Education Creative Team Funds Program.

ABSTRACT

Local features could learn semantic information for pedestrian images and they are very important for person re-identification (Re-ID) in harsh environments. However, most approaches only optimize one kind of local feature, which results in incomplete local features. In this paper, we propose Local Heterogeneous Features (LHF) to extract discriminative local features from three aspects. To this end, we utilize three kinds of losses to learn three kinds of local features, i.e., local discriminative features, local relative features, local compact features. As for local discriminative features, we split the attention maps into three horizontal sub-regions and perform the classification operation. Then, we divide the attention maps into two horizontal sub-regions, and we synchronously apply the triplet loss and center loss to learn local relative features and local compact features. Finally, we utilize local discriminative features to represent pedestrian. We evaluate LHF on public person Re-ID datasets and prove LHF is meaningful for local feature learning.

INDEX TERMS

Person re-identification, local heterogeneous features, harsh environments.

I. INTRODUCTION

Person re-identification (Re-ID) in harsh environments has an important practical application value, and it aims to distinguish the same identity under different camera sensors [1]–[5]. Because of the complexity of real scene, the pedestrian images possess large variances in illuminations, clothes, occlusions, backgrounds, and so on. Hence, person Re-ID still maintains many challenges. In recent years, deep learning approaches [6]–[8] become the mainstream approaches that bring some effective strategies to learn powerful feature representations. Some approaches [9]–[11] based on deep learning obtain a large breakthrough, which pushes person Re-ID to a high level.

The intuitive approach is to learn robust features from the global pedestrian images. The global features could search for the most salient region to represent pedestrian images [12]–[15]. However, the complexity of pedestrian images restricts the development of person Re-ID, because many non-salient details may be ignored. The local features contain subtle information which is important for person Re-ID. As shown in Fig. 1, the images in the first row have different identities, but very similar global appearances. It is hard to distinguish these pedestrian images using global features. The pedestrian images in the second row have the same identity, but large variations. The local features can extract effective information from head, legs or feet so as to recognize pedestrian images accurately.

FIGURE 1. (a) the pedestrian images with different identities, (b) the pedestrian images with the same identity.
In order to excavate discriminative features, some researchers utilize deep learning to learn local features [16]–[19], which further promotes person Re-ID performance. However, these methods split original pedestrian images or feature maps, and then utilize the cross-entropy function to optimize local features. Thus, the existing methods can not fully learn local information of pedestrian images.

To overcome the shortcoming, in this paper we propose Local Heterogeneous Features (LHF) to learn discriminative local features from different views for person Re-ID in harsh environments. The proposed LHF contains three kinds of local features, i.e., local discriminative features, local relative features, local compact features. Specifically, we simultaneously employ three kinds of loss functions to learn local features. To learn local discriminative features, we split attention maps into three horizontal regions and perform the pooling operation to extract features. Afterwards, Batch Normalization (BN) is utilize to implement a normalization operation. Then, the cross-entropy loss is utilized to perform a classification task. At the same time, we divide the attention maps into two horizontal regions. Then, we utilize the triplet loss and the center loss to learn local relative features and local compact features, respectively. By optimizing local features from three different views, we can fully learn local information from pedestrian images.

Different from previous part-based methods, the paper has the following three contributions:

- The local features are all learned from attention maps which contain abundant semantic information.
- We learn three kinds of local features from different views to discover local cues from pedestrian images.
- Extensive experiments on Re-ID datasets, i.e., Market1501 [20], CUHK03 [21] and DukeMTMC-reID [22], prove the proposed LHF is effective, which exceeds other competitive methods.

II. RELATED WORK

Person Re-ID shows great application value and potential. Traditional methods [23]–[27] usually extract low-level features, i.e., colour features, edge features, texture features, etc, to represent pedestrian images. Recently, it undergoes a large reformation from hand-crafted features to automatic deep features. The Convolutional Neural Networks (CNNs) [28]–[31] reveal a huge advantage on feature representations, and therefore they dominate the person Re-ID field.

Many researchers propose effective approaches using CNNs to learn global features so as to search for the most salient character. For examples, Zheng et al. [32] combine the classification model and the verification model to jointly optimize global features. Huang et al. [33] utilize adversarial occluded samples to enlarge training images and then retrain the deep model to obtain discriminative global features. Si et al. [34] design the Discriminative-Aware Integration (DAI) that employ different deep models to optimize global features, and then integrates these global features to represent images. Luo et al. [14] apply multiple tricks on the deep model to extract global features that promote the performance of person Re-ID.

The global features often lose some subtle information so that restrict the development of person Re-ID. So, some methods propose meaningful strategies to learn local features, which further improves person Re-ID. For instances, Zhang and Si [35] split original pedestrian images into three sub-regions and then employ the verification model to recognize the positive or negative sub-region pairs. Sun et al. [17] split the attention maps into six horizontal regions and apply six cross-entropy functions to optimize local features independently. Huang et al. [36] leverage the pose estimation algorithm to predict seventeen key points and then partition the attention maps into eight sub-regions so as to learn local features. Fu et al. [37] propose Horizontal Pyramid Matching (HPM) to learn different scale local features and use these local features to obtain pedestrian features.

These methods mentioned above learn local features using the cross-entropy loss. Different from them, we exploit three kinds of loss functions from multiple views to optimize local features. Then, we utilize local discriminative features to represent pedestrian images, which achieves a new level.

III. APPROACH

In the section, we introduce the structure of LHF. Then, we describe three kinds of local features, i.e., local discriminative features, local relative features and local compact features. Finally, we introduce how to optimize the overall deep model.

A. THE STRUCTURE OF LHF

The proposed LHF employs ResNet-50 [6] as backbone to extract image features due to the succinct structure. Different from the traditional setting, we modify the last stride from 2 to 1 and abandon the rear structure. This is because changing the stride from 2 to 1 can remove the down-sampling and increase the size of feature maps from $24 \times 8$ to $12 \times 4$. The feature maps maintain more semantic information. The overall backbone is shown in Table 1. We can find that the backbone contains a convolution layer, a pooling layer and four blocks. Each of block includes different number of convolution layers. As shown in Block 1 in Table 1, $1 \times 1$ denotes the size of filter and 64 indicates the number of filters. 1 is the stride of filter movement in the Stride column. $(1,1)$ represents padding one column and one row with pixel 0. $\times 3$ denotes the corresponding Block has three such structures. We crop each pedestrian image into $384 \times 128$, and then these pedestrian images are fed into the backbone. We obtain feature maps with the size of $24 \times 8 \times 2048$ after the backbone.

In order to learn local heterogeneous features, we propose an efficient model structure as shown in Fig. 2. We design three branches to learn local discriminative features, local relative features and local compact features, respectively. Local discriminative features could make full use of the
identity label and obtains the discriminative information. The local relative features could decrease the intra-class variation and increase the inter-class variation, and the local compact features could further decrease the intra-class variation.

**B. LOCAL DISCRIMINATIVE FEATURES**

As shown in Fig. 2, we apply the first branch to optimize local discriminative features. Specifically, we first partition the feature maps extracted from CNN into three horizontal sub-regions. Then, we perform the average pooling operation to extract local discriminative features $f_d^i$ ($i = 1, 2, 3$). Afterwards, local discriminative features are fed into the classifier. Here, each classifier contains a fully connected (FC) layer and a softmax, where the neuron number of the FC layer is equal to the identity number. We obtain the prediction probability that the $i$-th sub-region of each pedestrian image belongs to the $k$-th identity. It is formulated as:

$$p_i(k) = \frac{e^{a_i^k}}{\sum_{j=1}^{K} e^{a_j^k}}, \quad k \in \{1, 2, \cdots, K\}$$  \hspace{1cm} (1)
where \( K \) denotes the total identities in the training set and \( a^i_\theta \) represents the \( k \)-th neural value in the classifier.

Each local feature has an identity label which is the same as the label of corresponding original pedestrian image and each local feature corresponding to a cross-entropy loss function. Then, we employ the cross-entropy loss to compute the loss value of each pedestrian image:

\[
L_d = - \sum_{i=1}^{3} \sum_{k=1}^{K} q_i(k) \log(p(k))
\]

where \( q_i(k) \) is the true identity distribution and is written as:

\[
q_i(k) = \begin{cases} 
0 & k \neq t \\
1 & k = t 
\end{cases}
\]

where \( t \) is the true identity.

So, the total loss of each batch for local discriminative feature learning is written as:

\[
L_D = - \sum_{b=1}^{B} L_D^b
\]

where \( L_D^b \) is the loss of the \( b \)-th pedestrian image in one batch and \( B \) is the batch size.

### C. LOCAL RELATIVE FEATURES

We utilize the second branch to learn local relative features. From Fig. 2, we split the feature maps into two horizontal sub-regions and then perform Batch Normalization (BN) operation. Afterwards, we implement the average pooling process to extract local relative features \( f^i_b(i = 1, 2) \). We apply the triplet loss to optimize \( f^i_b \):

\[
L_R = \sum_{b=1}^{B} \sum_{i=1}^{2} [a + \max D((f^i_b)^h, f^i_m) - \min D((f^i_b)^h, f^i_m)]^+ 
\]

where \( a \) indicates the threshold, \((f^i_b)^h\) denotes the \( i \)-th sub-region feature of \( b \)-th pedestrian image in one batch, \( f^i_m \) represents the \( i \)-th sub-region feature of the \( m \)-th pedestrian image with the same identity of \((f^i_b)^h\), \( f^i_m \) is the \( i \)-th sub-region feature of the \( m \)-th pedestrian image with the different identity from \((f^i_b)^h\). Here, \( D[\cdot, \cdot] \) is to compute the Euclidean distance between two different feature vectors, \( \max D \) outputs the maximum value, \( \min D \) outputs the minimum value and \( [x]^+ = \max(x, 0) \).

### D. LOCAL COMPACT FEATURES

As shown in the third branch in Fig. 2, we apply center loss to optimize the local compact features \( f^i_c(i = 1, 2) \). Similarly, we divide the feature maps into two horizontal sub-regions. Since the three losses have different functions, we also add a BN layer for better convergence, and perform BN operation. Then, we implement the average pooling procedure to obtain \( f^i_c \). The loss is formulated as:

\[
L_C = \sum_{b=1}^{B} \sum_{i=1}^{2} \|((f^i_c)^h - t^i_b)^2
\]

where \((f^i_c)^h\) denotes the \( i \)-th sub-region feature of \( b \)-th pedestrian image in one batch and \( t^i_b \) indicates the \( i \)-th sub-region feature center of the \( b \)-th pedestrian image.

### E. MODEL OPTIMIZATION

We combine three loss functions to optimize local features and the final loss value is formulated as:

\[
L = L_D + \eta L_R + \rho L_C
\]

where \( \eta \) and \( \rho \) are the weight coefficients to control the importance of local relative features and local compact features.

The cross-entropy loss can make full use of the identity label and obtains the discriminative information. The triplet loss could decrease the intra-class variation and increase the inter-class variation, and the center loss could further decrease the intra-class variation. We add the three loss values as the final loss value. Then, we utilize the final loss value to compute the gradient and perform the back-propagating process. Afterwards, we utilize Stochastic Gradient Descent (SGD) to update parameters of the deep model. Local discriminative features are extracted from feature maps, which is optimized by three kinds of loss functions. So, the feature maps contain rich semantic information. We only utilize local discriminative features to represent pedestrian images, which has low dimension and complexity but better performance. So, we only utilize local discriminative features as final pedestrian features. In the test process, we concatenate three local discriminative features \( f_d = [f^1_d, f^2_d, f^3_d] \) to represent pedestrian features. Then, the feature \( f_d \) is utilized to evaluate the proposed LHF.

### IV. EXPERIMENTS

In this section, we introduce person Re-ID datasets, and then present the experimental settings. Afterwards, we evaluate three local features, i.e., local discriminative features, local relative features and local compact features. Finally, we compare LHF with other competitive approaches.

### A. DATABASES

**Market1501** [20]. The dataset possesses 1,505 identities. It totally contains 32,668 images and has two parts. One part includes 12,936 pedestrian images, which is called the training samples. The other part has 19,732 pedestrian images, which is the test set. These images are detected by DPM [38].

**DukeMTMC-reID** [22]. The dataset contains 1,812 identities and is captured by eight different cameras. These identities has 36,411 pedestrian images. Specifically, the training data includes 16,522 images, the test data has 17,661 images, and the query data possesses 2,228 images.

**CUHK03** [21]. The dataset possesses 14,097 pedestrian images of 1,467 identities. We employ the same image partitioning strategy as [33], [39]. The training samples have 7,365 images and the gallery samples have 5,332 images. The other 1,400 pedestrian images are treated as the query samples. We evaluate the proposed LHF using both two types of image samples.
TABLE 2. The performance of different local features. Rank 1 (%) and mAP (%) are shown.

| Methods | Market1501 rank1 mAP | DukeMTMC-reID rank1 mAP | CUHK03 (Detected) rank1 mAP | CUHK03 (Labeled) rank1 mAP |
|---------|-----------------------|--------------------------|-----------------------------|---------------------------|
| $f_d$   | 90.9 77.1             | 81.2 64.4                | 60.4 53.5                   | 64.1 55.8                 |
| $f_d + f_r$ | 93.2 84.7            | 85.6 75.7                | 64.9 63.5                   | 68.2 66.9                 |
| $f_d + f_r + f_c$ | 94.9 86.1          | 87.1 77.6                | 66.4 64.7                   | 70.1 67.3                 |

FIGURE 3. Some pedestrian images on three person Re-ID datasets.

The three datasets are captured at different environments and have different settings. Some pedestrian samples are list in Fig 3. In order to comprehensively evaluate LHF, we employ the mAP and Rank 1 as the criterion.

B. EXPERIMENT SETTINGS

In the experiments, the batch size is set to 64 which includes 16 identities and each identity possesses four different images. We randomly select 16 identities, and randomly choose four pedestrian images for each identity. To optimize the deep model, each image is cropped into $384 \times 128$, and then perform a normalization operation. $a$ in Eq. 5 is 0.3. The parameters $\eta$ and $\rho$ in Eq. 7 are 1 and 0.0005, respectively. In addition, the epoch number denotes the total number of training iterations. Based on the observation, the deep model achieves converge when epoch is 120, and therefore we set the epoch to 120. As for learning rate, we utilize different values at multiple steps and it is formulated as:

$$l(e) = \begin{cases} 
3.5 \times 10^{-5} \times \frac{e}{10}, & e \leq 10 \\
3.5 \times 10^{-4}, & 10 < e \leq 75 \\
3.5 \times 10^{-5}, & 30 < e \leq 70 \\
3.5 \times 10^{-6}, & 70 < e \leq 100 \\
3.5 \times 10^{-7}, & 100 < e \leq 120 
\end{cases}$$ (8)

where $e$ represents the current epoch.

C. PERFORMANCE OF LOCAL FEATURES

We evaluate three kinds of local features and the performances are shown in Table 2.

$f_d$ denotes local discriminative features and it achieves a good performance. This proves that $f_d$ is efficient. Then,
TABLE 4. Comparison LHF with other competitive methods on DukeMTMC-reID. Rank1 (%) and mAP (%) are shown.

| Methods     | rank1 | mAP  |
|-------------|-------|------|
| BoW+KISSME [20] | 25.1  | 12.2 |
| LOMO+XQDA [24] | 30.8  | 17.0 |
| GAN [22]    | 67.7  | 47.1 |
| OIM [54]    | 68.1  | -   |
| FAPL-d [44] | 71.9  | 52.3 |
| TriNet+Era [56] | 73.0  | 56.6 |
| APR [55]    | 73.9  | 55.6 |
| SVNet [9]   | 76.7  | 56.8 |
| MultiScale [48] | 79.2  | 60.6 |
| AOS [33]    | 79.2  | 62.1 |
| AWTL [49]   | 79.8  | 63.4 |
| HA-CNN [57] | 80.5  | 63.8 |
| DAI [34]    | 81.3  | 65.9 |
| DuATM [51]  | 81.8  | 64.6 |
| JALo [50]   | 82.5  | 66.9 |
| CTL [60]    | 85.4  | 71.9 |
| CAMA [61]   | 85.8  | 72.9 |
| SMGN [53]   | 86.1  | 57.3 |
| ASB [14]    | 86.4  | 76.1 |
| Ours        | 87.1  | 77.6 |

we add local relative features \( f_r \) to \( f_d \), which called \( f_d + f_r \). This method greatly raises the accuracy of person Re-ID. This verifies \( f_r \) can further distinguish different identities. Finally, we add center loss to learn local compact features \( f_c \). As a result, \( f_d + f_r + f_c \) obtains the best result, which demonstrates that \( f_c \) is meaningful for person Re-ID. The three kinds of local features are complementary and could learn the local information from different aspects.

D. COMPARISON WITH OTHER APPROACHES

In this subsection, we compare LHF with other competitive approaches on person Re-ID datasets.

On Market1501, we compare the proposed LHF with other approaches and the results are listed in Table 3. LHF obtains rank1 accuracy 94.9% and mAP 86.1%, which surpasses other competitive approaches. This denotes LHF could effectively learn local features. The three kinds of local feature jointly promote person Re-ID. In addition, ASB method employs different kinds of loss functions to optimize global features. Our method is better than ABS, which verifies local features are meaningful.

As listed in Table 4, we evaluate LHF on DukeMTMC-reID, and the results exceed other approaches. LHF obtains 87.1% and 77.6% at rank 1 accuracy and mAP, respectively. LHF achieves the best performance, which verifies the three kinds of local features are effective. Similarly, ASB utilizes different loss functions to optimize the global features. LHF utilizes the same functions to learn local features that is higher than ABS, which proves local information is meaningful.

E. PARAMETERS ANALYSIS

We perform experiments to verify the performance of different \( \alpha \) on Market1501 and the results are shown in Table 6. The results indicate that the method obtains the best performance when \( \alpha = 0.3 \).

\( \rho \) and \( \eta \) are the weight coefficients of the center loss and the triplet loss. We conduct the experiments to evaluate the two parameters and the results are shown in Table 7. When \( \rho = 0.0005 \) and \( \eta = 1 \), LHF achieves the best result.

V. CONCLUSION

We have proposed LHF for person Re-ID in harsh environments. The proposed LHF could learn three kinds of local features from multiple angles. Specifically, we employ the cross-entropy loss to optimize local discriminative features. At the same time, we apply the triplet loss to optimize local relative features and utilize the center loss to learn...
local compact features. Finally, we employ local discriminative features to represent pedestrian. Experiments on three datasets show that LHF effectively promotes the performance of person Re-ID.

The proposed LHF could achieve better results on three large-scale person Re-ID datasets. However, occlusion is an important challenge in person Re-ID. The method does not consider how to learn discriminative pedestrian features under occlusion. In the future, we will utilize the local features to overcome the occlusion problem.

REFERENCES

[1] J. Berclaz, F. Fleuret, and P. Fua, “Multi-camera tracking and atypical motion detection with behavioral maps,” in Proc. Eur. Conf. Comput. Vis., 2008, pp. 112–125.

[2] M. Soudel, “People detection, tracking and re-identification through a video camera network,” in Institut National de Recherche en Informatique et en Automatique. Paris, France: Smartert, 2013.

[3] S. Gong, M. Cristani, C. Loy, and T. M. Hospedales, “The re-identification challenge” in Person re-Identification. London, U.K.: Springer, 2014, pp. 1–20.

[4] Z. Wang, R. Hu, C. Chen, Y. Yu, J. Jiang, C. Liang, and S. Satoh, “Person reidentification via discrepancy matrix and matrix metric,” IEEE Trans. Cybern., vol. 48, no. 10, pp. 3006–3020, Oct. 2018.

[5] Z. Zhang, M. Huang, S. Liu, B. Xiao, and T. Durrani, “Fuzzy multi-layer clustering and fuzzy label regularization for unsupervised person re-identification,” IEEE Trans. Fuzzy Syst., early access, May 2, 2019, doi: 10.1109/TFUZZ.2019.2914626.

[6] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in Proc. Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2016, pp. 770–778.

[7] K. Simonyan and A. Zisserman, “Very deep convolutional networks for large-scale image recognition,” in Proc. Int. Conf. Learn. Represent., 2015, pp. 135–153.

[8] M. Ye, J. Li, A. J. Ma, L. Zheng, and P. C. Yuen, “Dynamic graph matching for unsupervised video-based person re-identification,” IEEE Trans. Image Process., vol. 28, no. 6, pp. 2976–2990, Jun. 2019.

[9] Y. Sun, L. Zheng, W. Deng, and S. Wang, “SVDNet for pedestrian retrieval,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 3820–3828.

[10] L. Zhao, X. Li, Y. Zhuang, and J. Wang, “Deeply-learned part-aligned representations for person re-identification,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 3219–3228.

[11] J. Song, Y. Yang, Y.-Z. Song, T. Xiang, and T. M. Hospedales, “Generalizable person re-identification by domain-invariant mapping network,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 719–728.

[12] A. Hermans, L. Beyer, and B. Leibe, “In defense of the triplet loss for person re-identification,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 3129–3137.

[13] H. Luo, Y. Gu, X. Liao, S. Lai, and W. Jiang, “Bag of tricks and a native vocabulary learning for high-level person representation,” in Proc. Eur. Conf. Comput. Vis. (ECCV), Dec. 2018, pp. 1114–1128.

[14] H. Zhang et al., “LHFs for Person Re-ID in Harsh Environments,” IEEE Access, vol. 8, 2020, pp. 112–125.

[15] Y. Sun, L. Zheng, Y. Yang, Q. Tian, and S. Wang, “Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline),” in Proc. Eur. Conf. Comput. Vis., 2018, pp. 480–496.

[16] J. Li, S. Zhang, J. Wang, W. Gao, and Q. Tian, “Global-local temporal representations for video person re-identification,” in Proc. IEEE/CVF Int. Conf. Comput. Vis. (ICCV), Oct. 2019, pp. 3958–3967.

[17] L. Zheng, Y. Huang, H. Lu, and Y. Yang, “Pose-invariant embedding for deep person re-identification,” IEEE Trans. Image Process., vol. 28, no. 9, pp. 4500–4509, Sep. 2019.

[18] L. Zheng, L. Shen, L. Tian, S. Wang, J. Wang, and Q. Tian, “Scalable person re-identification: A benchmark,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 2288–2295.

[19] W. Li, R. Zhao, T. Xiao, and X. Wang, “DeepReID: Deep filter pairing neural network for person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2014, pp. 152–159.

[20] Z. Zheng, L. Zheng, and Y. Yang, “Unlabeled samples generated by GAN improve the person re-identification baseline in vitro,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Oct. 2017, pp. 3754–3762.

[21] Z. Zhang, C. Wang, B. Xiao, W. Zhou, S. Liu, and C. Shi, “Cross-view action recognition via a continuous virtual path,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit., Jun. 2013, pp. 2690–2697.

[22] S. Liao, Y. Hu, X. Zhu, and S. Z. Li, “Person re-identification by local maximal occurrence representation and metric learning,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2015, pp. 2197–2206.

[23] Y. Shen, W. Lin, J. Yan, M. Xu, J. Wu, and J. Wang, “Person re-identification with correspondence structure learning,” in Proc. IEEE Int. Conf. Comput. Vis. (ICCV), Dec. 2015, pp. 3200–3208.

[24] Y. Yang, Z. Lei, S. Zhang, H. Shi, and S. Z. Li, “Metric embedded discriminative vocabulary learning for high-level person representation,” in Assoc. for the Advance of Artificial Intelligence, 2016, pp. 3648–3654.

[25] L. Wu, C. Shen, and A. V. D. Hengel, “Deep linear discriminant analysis with deep convolutional neural networks,” in Proc. Adv. Neural Inf. Process. Syst., 2012, pp. 1097–1105.

[26] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, “Inception-v4, inception-resnet and the impact of residual connections on learning,” in Proc. 31st AAAI Conf. Artif. Intell., 2017, pp. 1–12.

[27] M. Pang, Y.-M. Cheung, B. Wang, and R. Liu, “Robust heterogeneous discriminative analysis for face recognition with single sample per person,” Pattern Recognit., vol. 48, no. 10, pp. 3006–3020, Oct. 2018.

[28] Z. Zheng, L. Zheng, and Y. Yang, “A discriminatively learned CNN embedding for person reidentification,” ACM Trans. Multimedia Comput., Commun., Appl., vol. 14, no. 1, pp. 1–20, Jan. 2018.

[29] H. Huang, D. Li, Z. Zhang, X. Chen, and K. Huang, “Adversarially occluded samples for person re-identification,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 3008–3017.

[30] T. Si, Z. Zhang, and S. Liu, “Discrimination-aware integration for person re-identification in camera networks,” IEEE Access, vol. 7, pp. 33107–33114, 2019.

[31] Z. Zhang and T. Si, “Learning deep features from body and parts for person re-identification in camera networks,” EURASIP J. Wireless Commun. Netw., vol. 2018, no. 1, pp. 1–22, Dec. 2018.

[32] H. Huang, W. Yang, X. Chen, X. Zhao, K. Huang, J. Lin, G. Huang, and D. Du, “EANet: Enhancing alignment for cross-domain person re-identification,” 2018, arXiv:1812.11369. [Online]. Available: http://arxiv.org/abs/1812.11369

[33] Y. Fu, Y. Wei, Y. Zhou, H. Shi, G. Huang, X. Wang, Z. Yao, and T. Huang, “Horizontal pyramid matching for person re-identification,” in Proc. AAAI Conf. Artif. Intell., vol. 33, Jul. 2019, pp. 8295–8302.

[34] D. Forsyth, “Object detection with discriminatively trained part-based models,” Computer, vol. 47, no. 2, pp. 6–7, Feb. 2014.

[35] Zhong, L. Zhang, D. Gao, and S. Li, “Re-ranking person re-identification with k-Reciprocal encoding,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 3652–3661.

[36] L. Wu, C. Shen, and A. V. D. Hengel, “Deep linear discriminant analysis on Fisher networks: A hybrid architecture for person re-identification,” Pattern Recognit., vol. 65, pp. 238–250, May 2017.

[37] R. R. Varior, M. Haloi, and G. Wang, “Gated siamese convolutional neural network architecture for human re-identification,” in Proc. Eur. Conf. Comput. Vis., 2016, pp. 791–808.

[38] I. B. Barbosa, M. Cristani, B. Caputo, A. Rognhaugen, and T. Theoharis, “Looking beyond appearances: Synthetic training data for deep CNNs in re-identification,” Comput. Vis. Image Understand., vol. 167, pp. 50–62, Feb. 2018.
[43] D. Li, X. Chen, Z. Zhang, and K. Huang, “Learning deep context-aware features over body and latent parts for person re-identification,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 384–393.

[44] G. Ding, S. Khan, Z. Tang, and F. Porikli, “Feature mask network for person re-identification,” Pattern Recognit. Lett., to be published, doi: 10.1016/j.patrec.2019.02.015.

[45] L. He, J. Liang, H. Li, and Z. Sun, “Deep spatial feature reconstruction for partial person re-identification: Alignment-free approach,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 7073–7082.

[46] C. Song, Y. Huang, W. Ouyang, and L. Wang, “Mask-guided contrastive attention model for person re-identification,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 1179–1188.

[47] H. Yao, S. Zhang, R. Hong, Y. Zhang, C. Xu, and Q. Tian, “Deep representation learning with part loss for person re-identification,” IEEE Trans. Image Process., vol. 28, no. 6, pp. 2860–2871, Jun. 2019.

[48] Y. Chen, X. Zhu, and S. Gong, “Person re-identification by deep learning multi-scale representations,” in Proc. IEEE Int. Conf. Comput. Vis. Workshops (ICCVW), Oct. 2017, pp. 2590–2600.

[49] E. Ristani and C. Tomasi, “Features for multi-target multi-camera tracking and re-identification,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 6036–6046.

[50] W. Xiang, J. Huang, X. Qi, X. Hua, and L. Zhang, “Homocentric hypersphere feature embedding for person re-identification,” in Proc. IEEE Int. Conf. Image Process. (ICIP), Sep. 2019, pp. 1237–1241.

[51] J. Si, H. Zhang, C.-G. Li, J. Kuen, X. Kong, A. C. Kot, and G. Wang, “Dual attention matching network for context-aware feature sequence based person re-identification,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 5363–5372.

[52] X. Zhang, H. Luo, X. Fan, W. Xiang, Y. Sun, Q. Xiao, W. Jiang, C. Zhang, and J. Sun, “AlignedReID: Surpassing human-level performance in person re-identification,” 2017, arXiv:1711.08184. [Online]. Available: http://arxiv.org/abs/1711.08184

[53] D.-X. Li, G.-Y. Fei, and S.-W. Teng, “Learning large margin multiple granularity features with an improved siamese network for person re-identification,” Symmetry, vol. 12, no. 1, p. 92, 2020.

[54] T. Xiao, S. Li, B. Wang, L. Lin, and X. Wang, “Joint detection and identification feature learning for person search,” in Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR), Jul. 2017, pp. 3376–3385.

[55] Y. Lin, L. Zheng, Z. Zheng, Y. Wu, Z. Hu, C. Yan, and Y. Yang, “Improving person re-identification by attribute and identity learning,” Pattern Recognit., vol. 95, pp. 151–161, Nov. 2019.

[56] Z. Zhong, L. Zheng, G. Kang, S. Li, and Y. Yang, “Random erasing data augmentation,” 2017, arXiv:1708.04896. [Online]. Available: http://arxiv.org/abs/1708.04896

[57] W. Li, X. Zhu, and S. Gong, “Harmonious attention network for person re-identification,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit., Jun. 2018, pp. 2285–2294.

[58] L. Zheng, Y. Yang, and A. G. Hauptmann, “Person re-identification: Past, present and future,” 2016, arXiv:1610.02984. [Online]. Available: http://arxiv.org/abs/1610.02984

[59] H. Shi, Y. Yang, X. Zhu, S. Liao, Z. Lei, W. Zheng, and S. Z. Li, “Embedding deep metric for person re-identification: A study against large variations,” in Proc. Eur. Conf. Comput. Vis., 2016, pp. 732–748.

[60] T. Si, Z. Zhang, and S. Liu, “Compact triplet loss for person re-identification in camera sensor networks,” Ad Hoc Netw., vol. 95, Dec. 2019, Art. no. 101984.

[61] W. Yang, H. Huang, Z. Zhang, X. Chen, K. Huang, and S. Zhang, “Towards rich feature discovery with class activation maps augmentation for person re-identification,” in Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. (CVPR), Jun. 2019, pp. 1389–1398.

HAIJIA ZHANG is currently pursuing the master’s degree with Tianjin Normal University, Tianjin, China. His research interests include computer vision, deep learning, and sensor networks.

ZHONG ZHANG (Senior Member, IEEE) received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China. He is an Associate Professor with Tianjin Normal University, Tianjin, China. He has published about 110 articles in international journals and conferences such as the IEEE TRANSACTIONS ON FUZZY SYSTEMS, Pattern Recognition, the IEEE TRANSACTIONS ON CIRCUITS AND SYSTEMS FOR VIDEO TECHNOLOGY, the IEEE TRANSACTIONS ON INFORMATION FORENSICS AND SECURITY, Signal Processing (Elsevier), CVF, ICPR, and ICIP. His research interests include computer vision, remote sensing, and deep learning.

RONGHUA ZHANG is an Associate Professor with Shihezi University, Shihezi, China. She has published about 20 articles in major international journals and conferences. Her research interests include computer vision and digital image processing.

HAO MA is currently pursuing the B.S. degree with the College of Electronics and Communication Engineering, Tianjin Normal University, Tianjin, China. His current research interests include wireless communication, wireless body area networks, co-channel interference, pattern recognition, and sensor networks.

SHUANG LIU (Senior Member, IEEE) received the Ph.D. degree from the Institute of Automation, Chinese Academy of Sciences, Beijing, China. She is an Associate Professor with Tianjin Normal University, Tianjin, China. She has published over 50 articles in major international journals and conferences. Her research interests include computer vision, remote sensing, and deep learning.

TONGZHEN SI is currently pursuing the master’s degree with Tianjin Normal University, Tianjin, China. His research interests include computer vision, person reidentification, and deep learning.