




A Siamese Pedestrian Alignment Network for Person Re-identification

Yi Zheng^{1,2}, Yong Zhou^{1,2} , Jiaqi Zhao^{1,2} , Meng Jian³, Rui Yao^{1,2} ,
Bing Liu^{1,2}, and Xuning Liu^{1,2}

¹ School of Computer Science and Technology, China University of Mining and Technology, Xuzhou 221116, China

yzhou@cumt.edu.cn

² Engineering Research Center of Mine Digitization of the Ministry of Education of the People's Republic of China, Xuzhou 221116, China

³ Faculty of Information Technology, Beijing University of Technology, Beijing 100124, China

Abstract. Deep learning methods show strong ability in extracting high-level features for images in the field of person re-identification. The produced features help inherently distinguish pedestrian identities in images. However, on deep learning models over-fitting and discriminative ability of the learnt features are still challenges for person re-identification. To alleviate model over-fitting and further enhance the discriminative ability of the learnt features, we propose siamese pedestrian alignment networks (SPAN) for person re-identification. SPAN employs two streams of PAN (pedestrian alignment networks) to increase the size of network inputs over limited training samples and effectively alleviate network over-fitting in learning. In addition, a verification loss is constructed between the two PANs to adjust the relative distance of two input pedestrians of the same or different identities in the learned feature space. Experimental verification is conducted on six large person re-identification datasets and the experimental results demonstrate the effectiveness of the proposed SPAN for person re-identification.

Keywords: Person re-identification · Deep learning · Neural network · Verification loss · Feature learning

1 Introduction

The goal of person re-identification is to find people with the same identity in pictures taken by different cameras. Usually, it is treated as the problem of image retrieval [33], i.e., searching for a specific person (query) in a large image pool (gallery) [32]. Recently, several convolutional neural networks (CNN) based deep learning methods have been developed to deal with person re-identification problem. Since 2014 [12], CNN-based deep learning methods have greatly improved

The first author of this paper is a full-time postgraduate student.

© Springer Nature Switzerland AG 2019

Z. Lin et al. (Eds.): PRCV 2019, LNCS 11857, pp. 409–420, 2019.

https://doi.org/10.1007/978-3-030-31654-9_35

the performance of person re-identification [33] due to its powerful discriminative learning on large-scale datasets.

Currently, deep learning based methods are widely used in computer vision and pattern recognition tasks. In these methods, we usually use feature vectors to represent samples. To describe the relationship between feature vectors and take advantage of this relationship, the distance between feature vectors becomes an important indicator to measure the similarity of feature vectors. The purpose of metric learning is to obtain a metric matrix that can effectively reflect the similarity between samples. In the feature space based on the metric matrix, the distribution of intra-class samples is more compact, while the distribution of inter-class samples is more sparse. Xing *et al.* [25] first proposed the metric learning method in 2003. They improved the results of the clustering algorithm by learning the distance metric between pairs of samples. Weinberger *et al.* [24] combined a meta sample with a similar sample and a dissimilar sample to learn a distance metric that enables the distance between similar samples to be greater than the distance between dissimilar samples. Davis *et al.* [4] proposed the Information Theoretical Metric Learning (ITML). They randomly chose sample pairs and constrains the distances by labels, and proposed a regularization method for metric matrices. Law *et al.* [11] proposed a distance metric learning framework which exploits distance constraints over up to four different examples. This comparison between pairs of samples is often used in the retrieval process.

The pedestrian alignment network (PAN) was proposed in [36] to deal with two types of detector errors in person re-identification caused by excessive background and part missing, which were introduced by automatic pedestrian detectors. PAN treats pedestrian alignment and re-identification tasks as two related issues. It finds an optimal affine transformation from a given pedestrian image without extra annotations. With affine transformation, the pedestrian in the image is aligned to avoid excessive background in person re-identification. The discriminative ability of learnt features could be highly improved with the alignment since the learned feature maps usually exhibit strong activations on human body rather than background [36]. However, the verification loss as a special characteristic of person re-identification is not taken into consideration.

By making use of the pairwise relationships in metric learning, we propose siamese pedestrian alignment network (SPAN) for person re-identification. As an extension of PAN, SPAN takes two streams of input and output constructing a verification network. The main contributions of this paper are summarized as follows:

- The proposed siamese PAN constructs a verification model by combining two streams of PANs for person re-identification, which alleviates the overfitting of PAN greatly by taking pairwise samples to increase the input size in training the siamese network.
- SPAN evolves a verification loss embedded between PAN streams, which introduced the metric learning mechanism to make the images belonging to the same identity get close and push away the image with different identities in the learnt feature space.

- The features learnt from SPAN was cascaded with features extracted from unaligned pedestrian for person retrieval. The ensemble feature help improve the person re-identification performance dramatically.

The rest of this paper is organized as follows. Related works are reviewed and discussed in Sect. 2. Section 3 introduces the proposed method in detail and experimental results and comparisons are discussed in Sect. 4. The conclusions and further discussion are given in Sect. 5.

2 Related Work

2.1 Siamese-Based Methods

Currently, most deep learning based person re-identification methods are developed on the siamese model [7], which is fed with image pairs. Dong *et al.* [27] proposed deep metric learning method to investigate the part model for person re-identification. An input image was partitioned into three overlapping parts and then fed into three siamese CNNs for feature learning. A patch matching layer was designed to extract the result of patch displacement matrices in different horizontal stripes in [12]. The cross input neighborhood difference feature was proposed by Ahmed *et al.* [1], in which the features from one input image were compared to features in neighboring locations of the other image. Zeng *et al.* [29] proposed a person re-identification method based on Mahalanobis distance feature. The long short-term memory (LSTM) modules were incorporated into a siamese network for person re-identification in [21]. The discriminative feature learning ability was further enhanced by adopting LSTMs to process image parts sequentially. Siamese model only takes weak label of whether a pair of images is similar or not, while ignoring the given pedestrian labels. To make use of the pedestrian labels, Zheng *et al.* [34] proposed a discriminative siamese network combining both identification loss and verification loss in discriminative feature learning. It simultaneously learn discriminative and similarity metrics, which effectively improves pedestrian retrieval accuracy dramatically.

2.2 Pedestrian Alignment Network

Pedestrian alignment network (PAN) is able to simultaneously re-localize the person in a given image and categorize the person to one of pre-defined identities. The architecture of PAN is illustrated in Fig. 1. As shown in Fig. 1, PAN consists of a base branch and an alignment branch which includes grid network and ResNet blocks.

Both branches of PAN are classification networks to infer the identity of the input pedestrian images. The base branch performs to extract appearance features of the given image and provide corresponding information for the spatial localization of the person in the image. The alignment branch takes a similar neural network with base branch and further extracts features from the aligned feature maps produced by the grid network. The two branches work together by sequentially combining aligned features for person re-identification.



Fig. 1. The architecture of pedestrian alignment network. It consists of base branch and alignment branch.

3 Siamese Pedestrian Aligned Network

It is still far from satisfactory for PAN in pedestrian feature learning. As we observed, PAN is easily trapped into over-fitting with insufficient training data. In addition, pairwise verification is not taken into consideration, which reveals important clues for person re-identification. With these motivation, we scale up input size of PAN in SPAN by employing pairwise input images for the streams of PANs and embed a verification loss between the two streams to further leverage pairwise relationships in feature learning. The details of the proposed siamese PAN (SPAN) are illustrated in this Section.

3.1 Overview of SPAN

Since there are a large number of parameters in many deep learning models, a large number of labeled samples are needed for model training. However, it is hard to provide training samples of high quality for person re-identification. The siamese model [34] provides an effective way to make full use of the limited training samples by employing pairwise input and metric learning mechanism.

The architecture of the proposed SPAN is shown in Fig. 2. The SPAN consists of siamese base branch and siamese alignment branch. Both of them have two streams and the parameters are shared with each other. The input of the siamese base branch is a pair of pedestrian images with the same or different identities. The siamese base branch is constructed to learn high-level image features by minimizing both identification loss and verification loss. The input of the siamese alignment branch is a pair of pedestrian feature maps obtained from the based branch. The siamese alignment branch works to learn the location of pedestrians and extract discriminative features for further pedestrian retrieval. The ResNet-50 [5] is adopted as the base component of the two branches due to its excellent performance in feature learning, details of ResNet-50 can be seen in [36]. In addition, the verification loss, which bridges the two PAN streams for similarly metric learning, is embedded in both siamese based branch (SBB) and siamese alignment branch (SAB).

3.2 Siamese Base Branch

Besides the initial base branch of PAN, a square layer is further embedded to measure the difference between the high-level features f_a^1 and f_a^2 in the siamese

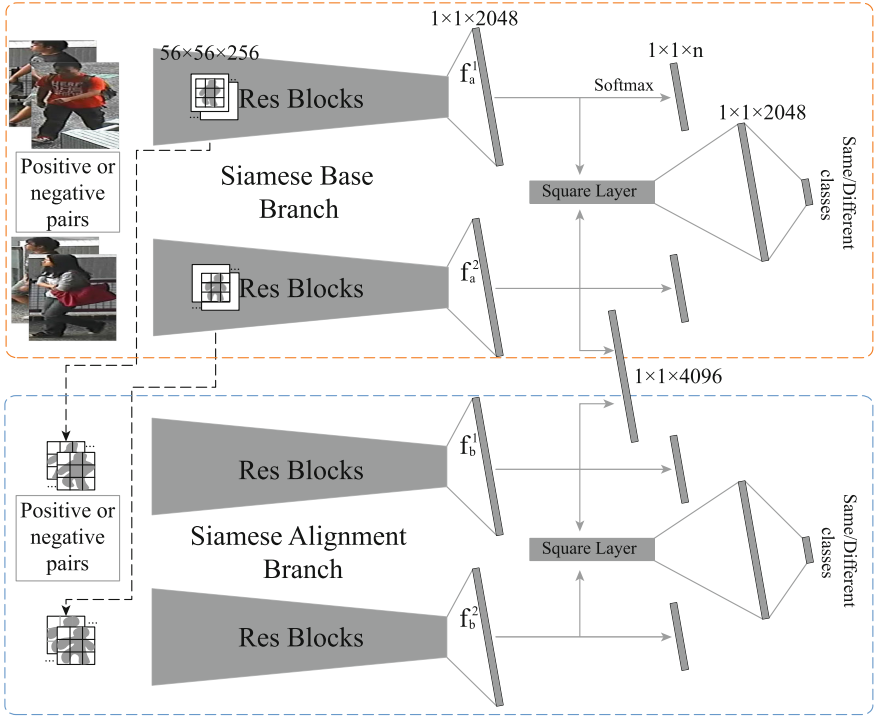


Fig. 2. The architecture of the SPAN model, which consists of the siamese base branch and siamese alignment networks. The output of the two branch is a $1 \times 1 \times 2048$ vector. In the retrieval phase, two 2048-dim vectors, i.e., f_b^1 and f_a^1 are concatenated to describe the input image.

base branch. A convolutional layer is built following the square layer as a verification module, which is constructed to evaluate whether the input pair of images belongs to the same or different identities. The high-level features learned by the additional convolutional layer act as a metric learning module, which forces the distance of pedestrians of the same identity as close as possible, and the distance between different pedestrians as far as possible in the feature space.

Identification: In SBB, the identification loss plays a key role in discriminative feature learning and pedestrian identification. Each stream of SBB contains five Res blocks [5] and they are pre-trained on ImageNet [10] to fine tune the Res blocks. By replacing the fully-connected layer with a convolutional layer and softmax unit [34], we could infer the multi-class probabilities as Eq. 1.

$$\hat{p}^i = \text{softmax}(\phi_b^I \circ f_b^i), i = 1, 2 \quad (1)$$

where, \circ denotes the convolutional operation, f_b^i is a 2048-dim feature vector extracted by one stream of SBB, and ϕ_b^I denotes the parameters of the convolutional layer which connected to the output feature f_b^i . The cross-entropy loss of classification model is formulated as Eq. 2.

$$loss_{C-SBB}(f_b^i, \phi_b^I) = - \sum_{i=1}^2 \sum_{k=1}^K p_k^i \log(\hat{p}_k^i) \quad (2)$$

where, K denotes the number of identity classes in the dataset, and the target class is represented by t . Only if $k = t$ then $p_k^i = 1$, otherwise $p_k^i = 0$, \hat{p}_k^i is the probability of every class predicted by softmax function.

Verification: Depending on the relationship of the input image pair, we further add a verification module in the siamese base branch to adjust relative distances of pedestrian pair in the learnt feature space by a verification loss. Firstly, a 2048-dim tensor f_b^s is obtained by multiplying the difference between f_b^1 and f_b^2 element-wisely as Eq. 3.

$$f_b^s = (f_b^1 - f_b^2) \times (f_b^1 - f_b^2) \quad (3)$$

After the added convolutional layer and softmax output function, f_b^s is mapped to a 2-dim vector $(\hat{q}_b^1, \hat{q}_b^2)$. Here $\hat{q}_b^1 + \hat{q}_b^2 = 1$, which represents the predicted probability whether the two input images belong to the same identity or not. The formulation is represented as Eq. 4.

$$\hat{q}_b = softmax(\phi_b^S \circ f_b^s), \quad (4)$$

where, ϕ_b^S denotes the parameters of the convolutional layer. The verification loss of SBB is denoted by Eq. 5.

$$loss_{V-SBB}(f_b^s, s, \phi_b^S) = - \sum_{i=1}^2 q_b^i \log(\hat{q}_b^i), \quad (5)$$

where, p represents whether the inputted image pairs belong to the same class or not. If the input images belong to the same class, $q_b^1 = 1$ and $q_b^2 = 0$, otherwise $q_b^1 = 0$ and $q_b^2 = 1$.

3.3 Siamese Alignment Branch

Similar to the siamese base branch, each stream of siamese alignment branch is a categorization module for pedestrian identification, including a variation ResNet and a grid network which aligns the position of pedestrians in feature maps [36]. The identification loss of the SAB is constructed as Eqs. 6 and 7.

$$\hat{p}^i = softmax(\phi_a^I \circ f_a^i), i = 1, 2 \quad (6)$$

$$loss_{C-SAB}(f_a^i, t, \phi_a^I) = - \sum_{i=1}^2 \sum_{k=1}^K p_k^i \log(\hat{p}_k^i) \quad (7)$$

where, f_a^i denotes the feature vector extracted by each stream of the siamese alignment branch. ϕ_a^I represents the parameters of the corresponding deep model.

A verification module is embedded in the SAB network which introduces the metric learning mechanization to optimize the pairwise relationships of images. The verification loss in alignment branch is represented by Eq. 9.

$$\hat{q}_a = softmax(\phi_a^S \circ f_a^s) \quad (8)$$

$$loss_{V-SAB}(f_a^s, s, \phi_a^S) = - \sum_{i=1}^2 q_a^i \log(\hat{q}_a^i) \quad (9)$$

Here in Eqs. 8 and 9, f_a^s denotes the 2048-dim vector calculated by f_a^1 and f_a^2 , ϕ_a^S represents the parameters of the added convolutional layer. Similarly, when giving images belong to the same class $q_a^1 = 1$ and $q_a^2 = 0$, otherwise $q_a^1 = 0, q_a^2 = 1$. The vector $(\hat{q}_a^1, \hat{q}_a^2)$ mapped by f_a^s represents the probability if the input images belong to the same class.

In summary, the target of the proposed siamese pedestrian alignment network is to perform aligned pedestrian feature learning by minimizing two cross-entropy losses of the siamese base branch and siamese alignment branch, as Eq. 10.

$$min(\lambda_1 loss_{C-SBB} + \lambda_2 loss_{V-SBB} + \lambda_3 loss_{C-SAB} + \lambda_4 loss_{V-SAB}), \quad (10)$$

here $\lambda_i, i = 1, 2, 3, 4$ represent the weight of each loss in the entire network respectively.

3.4 Feature Concatenation

In the testing phase, for an input pedestrian image x SPAN extracts two features of f_b and f_a by feeding the image to one stream of the SBB and SAB respectively. As the features of SBB and SAB tend to compensate with each other, we concatenate them together to jointly represent features of the image for pedestrian retrieval. After combining features of SBB and SAB, an l^2 -normalization layer is performed to unify feature space as Eq. 11.

$$f = [|f_b|^T, |f_a|^T]^T *, \quad (11)$$

where, $*$ represents the operation of l^2 -normalization.

4 Experimental Results

In this section, six large-scale datasets, i.e., Market-1501 [32], CUHK03 [37], DukeMTMC-reID [16], PKU-Reid [15], MSMT17 [23], Airport [9] are selected to evaluate the performance of the proposed method.

During the training process, our network needs to read at least two images belong to the same identity at the same time. If not, the network could not find vectors to push them together in the feature space. In Airport dataset, some pedestrian identities have only one image, in Market-1501 and MSMT17 dataset there are some identities just have two or three images. Therefore, we have expended these identities with fewer images in the datasets through random cropping. Also, random cropping can cause more serious part missing problems.

4.1 Implementation Details

All of the experiments were based on the MatConvnet [22] and were implemented on a graphics workstation with Intel core i7-8700k CPU, 32 GB memory and a Nvidia GTX 1080Ti. The maximum number of training epochs was set as 85 for the SBB, and the learning rate was set to 10^{-3} for the first 70 epochs then it turned to 10^{-4} for 10 epochs, and 10^{-5} was the learning rate for the last 5 epochs. The mini-batch stochastic gradient descent (SGD) was adapted to update the parameters of the whole network. For SBB, there are three objectives include identification loss and verification loss in our network. So we computed all the gradients produced by every objective respectively and gave them weight before we added them together to update the network. The identification loss $loss_{C-SBB}$ contain two objectives and the weight $\lambda_1 = 0.5$, the weight of verification loss $loss_{V-SBB}$ is $\lambda_2 = 1$.

The settings in SAB were similar to the SBB, except some differences such as the maximum number of training epochs, learning rates, and the objectives and their weights. We set the number of training epochs to 110 and the learning rate was 10^{-3} , 10^{-4} , 10^{-5} for 1st to 70th epochs, 71st to 90th epochs and 91st to 110th respectively. In SAB, there are five objectives, four of them belong to identification loss. We got our best result under the setting of $\lambda_3 = 0.5$ $\lambda_4 = 1$

The mean average precision (mAP) and the rank-1 accuracy are selected to evaluate the performance on these datasets.

4.2 Results

We compare the performance of the proposed model with PAN [36] and a siamese CNN (SCNN) network [34] which is based on ResNet-50. The results of three methods on six datasets are shown in Table 1.

According to [36], the PAN is used to reduce the problems of excessive background and part missing. These problems occur more frequently in pedestrian detectors than hand-drawn bounding boxes. So, the results obtained by CUHK03 (labeled), in which the detected boxes are cropped manually, the PAN does not make enough contributions to align the pedestrian, even may cause the part missing problem because of overcropping.

On Market-1501 and DukeMTMC-reID dataset, we compare SPAN with the state-of-the-art methods. The results are shown in Tables 2 and 3. With re-ranking step the mAP of SPAN reached to **76.54%** on Market-1501, which is better than most results in both published paper and arXiv paper. On DukeMTMC-reID dataset, we reached to a competitive result **mAP = 64.68%** after re-ranking. On CUHK03, the **mAP** had improved to **44.73%** for detected setting, and **48.62%** for labeled setting.

At present, the machine we are using for the experiments had met the out of memory error when performing the re-ranking [37] step on MSMT17 dataset because of the number of identities in the query is over 10,000, that is why we did not get result after re-ranking step on MSMT17 dataset.

Table 1. Comparison of different methods on six datasets. Rank-1 accuracy (%) and mAP(%) are shown in the table.

Dataset	Accuracy	SCNN	PAN	SPAN
Market1501	R-1	81.41	80.10	85.58
	mAP	62.55	58.93	68.13
DukeMTMC-reID	R-1	71.09	67.95	73.42
	mAP	51.14	48.69	54.32
CUHK03 (detected)	R-1	46.28	28.28	47.42
	mAP	42.01	26.65	43.58
CUHK03 (labeled)	R-1	48.78	29.35	48.21
	mAP	45.48	28.07	46.15
MSMT17	R-1	51.78	50.73	52.81
	mAP	26.96	24.93	28.94
Airport	R-1	15.15	8.92	15.56
	mAP	12.68	6.84	13.77
PKU-Reid	R-1	37.71	58.77	63.51
	mAP	33.97	42.75	48.41

Table 2. Comparison of the SPAN with start of the art methods of mAP (%) on Market-1501.**RR**:Re-Ranking [37]

Methods	mAP
PDF [19]	63.41
DSR [6]	64.25
GAN [35]	66.07
SSM [2]+RR	68.80
CamGAN [38]+RR	71.55
DaF [28]	72.42
JLML [13]+RR	72.9
HA-CNN [14]	75.7
SPAN	68.13
SPAN+RR	76.54

Table 3. Comparison of the SPAN with start of the art methods of mAP (%) on Market-1501.**RR**:Re-Ranking [37]

Methods	mAP
SVDNet [20]	56.8
dMpRL [8]	58.56
AACN [26]	59.25
CamGAN [38]	57.61
DPFL [3]	60.6
PSE [18]	62.0
ATWL [17]	63.40
HA-CNN [14]	63.8
SPAN	54.32
SPAN+RR	64.68

However, on PKU-Reid dataset we further improve our result **mAP = 70.27%** and **Rank-1 accuracy = 73.68%** after re-ranking step.

Otherwise, the mAP and Rank-1 accuracy of airport dataset were not improved when using re-ranking.

5 Conclusion

In this paper, we proposed siamese pedestrian alignment network (SPAN) for person re-identification. SPAN is devoted to alleviating the problem of over-fitting and enhancing discriminative ability of the learnt features of PAN model. A verification loss was embedded between the two PAN streams to adjust the relative distance of the two pedestrians with the same or different identifies. The proposed model is able to improve the discriminative learning and similarity measurement at the same time by addressing the misalignment problem and the specific problem of person re-identification. The effectiveness of the proposed method was verified on six large-scale person re-ID datasets. Experimental results shown that the proposed method outperforms PAN and siamese CNN greatly.

Re-ranking [37] is a very effective way to improve the results of re-identification. In [36], the result of PAN had improved from 63% to 76% on Market-1501 dataset. So we can also make SPAN get a better result by adding re-ranking steps.

However, the computation complexity of the proposed model did not take into consideration. In the future, we will focus on optimizing the structure of deep neural networks to reduce the computation complexity and improve its efficiency for person re-identification by using multi-objective learning methods [30, 31].

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China (No. 61572505, No. 61772530, No. U1610124, and No. 61806206), in part by the State’s Key Project of Research and Development Plan of China (No. 2016YFC0600900), in part by the Six Talent Peaks Project in Jiangsu Province under Grant 2015-DZXX-010, under Grant 2018-XYDXX-044, in part by the Natural Science Foundation of Jiangsu Province (No. BK20171192, No. BK20180639), and in part by the China Postdoctoral Science Foundation (No. 2018M642359).

References

1. Ahmed, E., Jones, M., Marks, T.K.: An improved deep learning architecture for person re-identification. In: *Computer Vision and Pattern Recognition*, pp. 3908–3916 (2015)
2. Bai, S., Bai, X., Tian, Q.: Scalable person re-identification on supervised smoothed manifold. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 2530–2539 (2017)
3. Chen, Y., Zhu, X., Gong, S.: Person re-identification by deep learning multi-scale representations. In: *Proceedings of the IEEE International Conference on Computer Vision*, pp. 2590–2600 (2017)
4. Davis, J.V., Kulis, B., Jain, P., Sra, S., Dhillon, I.S.: Information-theoretic metric learning. In: *Proceedings of the 24th International Conference on Machine Learning*, pp. 209–216. ACM (2007)
5. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770–778 (2016)

6. He, L., Liang, J., Li, H., Sun, Z.: Deep spatial feature reconstruction for partial person re-identification: alignment-free approach. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 7073–7082 (2018)
7. Hu, J., Lu, J., Tan, Y.P.: Discriminative deep metric learning for face verification in the wild. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 2014
8. Huang, Y., Xu, J., Wu, Q., Zheng, Z., Zhang, Z., Zhang, J.: Multi-pseudo regularized label for generated data in person re-identification. *IEEE Trans. Image Process.* **28**(3), 1391–1403 (2019)
9. Karanam, S., Gou, M., Wu, Z., Rates-Borras, A., Camps, O., Radke, R.J.: A systematic evaluation and benchmark for person re-identification: features, metrics, and datasets. arXiv preprint [arXiv:1605.09653](https://arxiv.org/abs/1605.09653) (2016)
10. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Advances in Neural Information Processing Systems, pp. 1097–1105 (2012)
11. Law, M.T., Thome, N., Cord, M.: Learning a distance metric from relative comparisons between quadruplets of images. *Int. J. Comput. Vision* **121**(1), 65–94 (2017)
12. Li, W., Zhao, R., Xiao, T., Wang, X.: DeepReID: deep filter pairing neural network for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 152–159 (2014)
13. Li, W., Zhu, X., Gong, S.: Person re-identification by deep joint learning of multi-loss classification. arXiv preprint [arXiv:1705.04724](https://arxiv.org/abs/1705.04724) (2017)
14. Li, W., Zhu, X., Gong, S.: Harmonious attention network for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2285–2294 (2018)
15. Ma, L., Liu, H., Hu, L., Wang, C., Sun, Q.: Orientation driven bag of appearances for person re-identification. arXiv preprint [arXiv:1605.02464](https://arxiv.org/abs/1605.02464) (2016)
16. Ristani, E., Solera, F., Zou, R., Cucchiara, R., Tomasi, C.: Performance measures and a data set for multi-target, multi-camera tracking. In: Hua, G., Jégou, H. (eds.) ECCV 2016. LNCS, vol. 9914, pp. 17–35. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-48881-3_2
17. Ristani, E., Tomasi, C.: Features for multi-target multi-camera tracking and re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 6036–6046 (2018)
18. Saquib Sarfraz, M., Schumann, A., Eberle, A., Stiefelwagen, R.: A pose-sensitive embedding for person re-identification with expanded cross neighborhood re-ranking. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 420–429 (2018)
19. Su, C., Li, J., Zhang, S., Xing, J., Gao, W., Tian, Q.: Pose-driven deep convolutional model for person re-identification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3960–3969 (2017)
20. Sun, Y., Zheng, L., Deng, W., Wang, S.: SVDNet for pedestrian retrieval. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 3800–3808 (2017)
21. Varior, R.R., Shuai, B., Lu, J., Xu, D., Wang, G.: A siamese long short-term memory architecture for human re-identification. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9911, pp. 135–153. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46478-7_9

22. Vedaldi, A., Lenc, K.: MatconvNet: convolutional neural networks for matlab. In: Proceedings of the 23rd ACM International Conference on Multimedia, pp. 689–692. ACM (2015)
23. Wei, L., Zhang, S., Gao, W., Tian, Q.: Person transfer GAN to bridge domain gap for person re-identification. In: IEEE Conference on Computer Vision and Pattern Recognition (2018)
24. Weinberger, K.Q., Blitzer, J., Saul, L.K.: Distance metric learning for large margin nearest neighbor classification. In: Advances in Neural Information Processing Systems, pp. 1473–1480 (2006)
25. Xing, E.P., Jordan, M.I., Russell, S.J., Ng, A.Y.: Distance metric learning with application to clustering with side-information. In: Advances in Neural Information Processing Systems, pp. 521–528 (2003)
26. Xu, J., Zhao, R., Zhu, F., Wang, H., Ouyang, W.: Attention-aware compositional network for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2119–2128 (2018)
27. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Deep metric learning for person re-identification. In: 2014 22nd International Conference on Pattern Recognition (ICPR), pp. 34–39. IEEE (2014)
28. Yu, R., Zhou, Z., Bai, S., Bai, X.: Divide and fuse: a re-ranking approach for person re-identification. arXiv preprint [arXiv:1708.04169](https://arxiv.org/abs/1708.04169) (2017)
29. Zeng, M., Wu, Z., Tian, C., Zhang, L., Zhao, X.: Person re-identification based on a novel mahalanobis distance feature dominated kiss metric learning. *Electron. Lett.* **52**(14), 1223–1225 (2016)
30. Zhao, J., et al.: Multiobjective optimization of classifiers by means of 3D convex-hull-based evolutionary algorithms. *Inf. Sci.* **367–368**, 80–104 (2016)
31. Zhao, J., et al.: 3D fast convex-hull-based evolutionary multiobjective optimization algorithm. *Appl. Soft Comput.* **67**, 322–336 (2018)
32. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: a benchmark. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1116–1124 (2015)
33. Zheng, L., Yang, Y., Hauptmann, A.G.: Person re-identification: past, present and future. arXiv preprint [arXiv:1610.02984](https://arxiv.org/abs/1610.02984) (2016)
34. Zheng, Z., Zheng, L., Yang, Y.: A discriminatively learned CNN embedding for person re-identification. *ACM Trans. Multimedia Comput. Commun. Appl. (TOMM)* **14**(1), 13:1–13:20 (2017)
35. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by GAN improve the person re-identification baseline in vitro. arXiv preprint [arXiv:1701.07717](https://arxiv.org/abs/1701.07717) 3 (2017)
36. Zheng, Z., Zheng, L., Yang, Y.: Pedestrian alignment network for large-scale person re-identification. *IEEE Trans. Circuits Syst. Video Technol.* (2018)
37. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 3652–3661. IEEE (2017)
38. Zhong, Z., Zheng, L., Zheng, Z., Li, S., Yang, Y.: Camera style adaptation for person re-identification. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5157–5166 (2018)