



HAL
open science

Person Re-identification with a Body Orientation-Specific Convolutional Neural Network

Yiqiang Chen, Stefan Duffner, Andrei Stoian, Jean-Yves Dufour, Atilla
Baskurt

► **To cite this version:**

Yiqiang Chen, Stefan Duffner, Andrei Stoian, Jean-Yves Dufour, Atilla Baskurt. Person Re-identification with a Body Orientation-Specific Convolutional Neural Network. *Advanced Concepts for Intelligent Vision systems*, Sep 2018, Poitiers, France. hal-01895374

HAL Id: hal-01895374

<https://hal.science/hal-01895374v1>

Submitted on 15 Oct 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Person Re-identification with a Body Orientation-Specific Convolutional Neural Network

Yiqiang Chen¹, Stefan Duffner¹, Andrei Stoian², Jean-Yves Dufour², and Atilla Baskurt¹

¹ Univ Lyon, INSA-Lyon, CNRS, LIRIS, F-69621, Villeurbanne, France

² Thales Services, ThereSIS, Palaiseau, France

Abstract. Person re-identification consists in matching images of a particular person captured in a network of cameras with non-overlapping fields of view. The challenges in this task arise from the large variations of human appearance. In particular, the same person could show very different appearances from different points of view. To address this challenge, in this paper we propose an Orientation-Specific Convolutional Neural Network (OSCNN) framework which jointly performs body orientation regression and extracts orientation-specific deep representations for person re-identification. A robust joint embedding is obtained by combining feature representations under different body orientations. We experimentally show on two public benchmarks that taking into account body orientations improves the person re-identification performance. Moreover, our approach outperforms most of the previous state-of-the-art re-identification methods on these benchmarks.

Keywords: Person Re-identification, Convolutional Neural Network, Mixture of Experts

1 Introduction

Person re-identification is the problem of identifying people across images that have been captured by different surveillance cameras with non-overlapping views. The task is increasingly receiving attention because of its important applications in video surveillance such as cross-camera tracking, multi-camera behavior analysis and forensic search.

However, this problem is challenging due to the large variations of lightings, poses, viewpoints and backgrounds. The main difficulty is that the pedestrian appearance can be very different with different body orientations under different viewpoints, *i.e.* images of the same person can look quite different and images of different persons can look very similar (see Fig. 1). Moreover, low image resolution and partial occlusion in images make the problem even harder.

Most existing approaches consider that pedestrian images come from a single domain. The viewpoint-invariant feature representations are either designed



Fig. 1. Some image examples from a person re-identification dataset. Pedestrian appearance can be very different due to different body orientations.

“manually” or learned automatically by a deep neural network. Though, re-identification can be considered as a multi-domain problem, *i.e.* pedestrians with the same body orientation have similar silhouettes and those with different body orientations have dissimilar appearance. Some metric learning approaches, for example, learn to transfer the feature space from one camera to another. But this requires a model for all the combination of cameras. Some other metric learning methods learn to transfer the different view-specific feature spaces to a common subspace where features are discriminative. This addresses the lighting and background variations, but it cannot be generalised to new camera views, and pedestrian images still have variations from different body orientations even if they come from the same camera.

To tackle this issue, we use a multi-task deep Convolutional Neural Network (CNN) to perform body orientation regression in a gating branch, and in another branch separate orientation-specific layers are learned as local experts. The combined orientation-specific CNN feature representations are used for the person re-identification task. Our main contributions are:

- a mixture-of-expert deep CNN to model the multi-domain pedestrian images for person re-identification. We show that learning and combining different feature embeddings of different orientations improves the re-identification performance,
- a novel multi-task CNN framework with combined person orientation estimation and re-identification, where the estimated body orientation is used to steer the orientation specific mixture of experts for re-identification,
- an experimental evaluation showing that our approach outperforms most state-of-the-art methods on the CUHK01 and Market-1501 datasets.

2 Related Work

Existing person re-identification approaches generally build a robust feature representation or learn a distance metric. The features used for re-identification are mainly variants of color histograms, Local Binary Patterns (LBP) or Gabor features. Some approaches use features that are specifically designed to be robust to common appearance variations. For example, Gray *et al.* [6] extract RGB, YUV and HSV channels and LBP texture histograms in horizontal stripes as feature vector. Liao *et al.* [13] propose the LOMO features. Color and SILTP histograms

are extracted in sliding windows and only the maximal occurrence is kept along each horizontal strip. The main metric learning methods include Mahalanobis metrics like KISSME [7], Local Fisher discriminant Analysis (LFDA) [17] and Cross-view Quadratic Discriminant Analysis (XQDA) [13].

With the recent success of deep learning for computer vision applications, many deep convolution neural network approaches have been proposed for person re-identification. For example, Li *et al.* [11] adopted a filter pairing neural network (FPNN) to model the displacement of body parts for person re-identification. Amed *et al.* [1] introduced an improved Siamese architecture using the difference of feature maps to measure the similarity. Cheng *et al.* [4] proposed a variant of the triplet loss function and a CNN network processing parts and the entire body. Varior *et al.* [22] proposed a Siamese CNN integrating a gate layer to capture effective subtle patterns in the feature map. The deep networks proposed in [8] and [27] learn a body part alignment and localisation and extract body part regions in an unified framework.

Most existing methods for person re-identification focus on developing a robust representation to handle the variations of view. Some methods take into account the view as extra information. For example, Ma *et al.* [15] divide the data according to the additional camera position information and learn a specific distance metric for each camera pair. Lisanti *et al.* [14] proposed to apply Kernel Canonical Correlation Analysis which finds a common subspace between the feature space from disjoint cameras. Yi *et al.* [25] proposed to apply a Siamese CNN to person re-identification. Similar to [14], the weights of two subnetworks are not shared to learn a camera view projection to a common feature space. In these approaches, camera information is used but the body orientation which is only partly due to different camera views is not modelled. That is, in the same camera view, pedestrians can exhibit different orientations and thus largely different appearances in the resulting images.

In order to solve this issue, Bak *et al.* [2] perform an orientation-driven feature weighting and the body orientation is calculated according to the walking trajectory. Some other approaches [23, 18] deal with the orientation variations of pedestrian images by using Mixture of Experts. The expert neural networks map the input to the output, while a gating network produces a probability distribution over all experts' final predictions. Verma *et al.* [23] applied an orientation-based mixture of experts to the pedestrian detection problem. Sarfraz *et al.* [18] proposed to learn the orientation sensitive units in a deep neural network to perform attribute recognition. Garcia *et al.* [5] used orientations estimated by a Kalman filter and then trained two SVM classifiers for pedestrian images matching with respectively similar orientations and dissimilar orientations. And the approach of Li *et al.* [9] learns a mixture of experts, where samples were softly distributed into different experts via a gating function according to the viewpoint similarity.

Sharing the idea of mixture of experts, we propose to build a multi-domain representation in different orientations with deep convolutional neural networks. Intuitively, an orientation-specific model should have a better generalization abil-

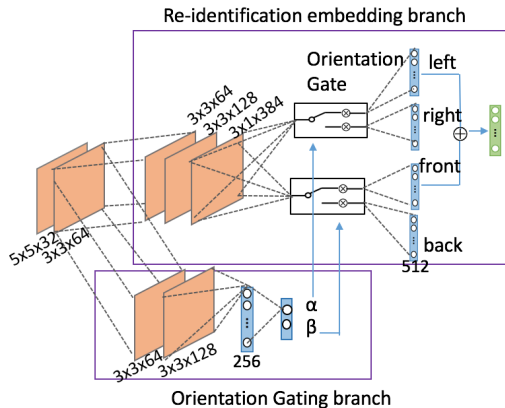


Fig. 2. Overview of the OSCNN architecture.

ity than a camera view-specific model, since we cannot incorporate all possible surveillance camera views. Further, instead of using discrete orientations for the gating activation function, in our method, we use a regressor to estimate an accurate and continuous body orientation. This allows to continuously weight different expert models for re-identification and also avoids combining contradictory orientations.

3 Proposed method

The overall procedure of our re-identification approach OSCNN is shown in Fig. 2. The network contains an orientation gating branch and a re-identification branch consisting of 4 feature embeddings regarding the 4 main orientations: left, right, frontal and back. The final output feature representation is a linear combination of the four expert outputs and is steered by an orientation gate unit which is a function of the estimated orientation.

3.1 OSCNN architecture

The proposed neural network architecture consists of two convolution layers shared between an orientation gating branch and a re-identification feature embedding branch. In the re-identification branch, there are 3 further convolution layers followed by 4 separate, parallel fully-connected layers of 512 dimensions, each one corresponding to a local expert. Thus, our network learns different projections to a common feature space, as shown Fig. 3 .

In the orientation regression branch, 2 convolution layers and 2 fully connected layers are connected to the common convolutional layers. The estimated orientation output by the orientation gating branch is represented by a two-dimensional Cartesian vector $[\alpha, \beta]$ constructed by projecting the orientation

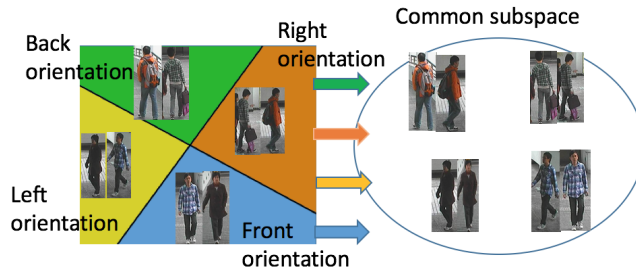


Fig. 3. Pedestrian images from different orientations could be considered as different domains. Our method learns different orientation-specific projections into a common feature space.

angle on the left-right axis (x) and on the front-back (y) axis and then normalizing it to a unit vector. Based on this vector, the orientation gate selects and weights either the left or the right component and either the front or the back component of the re-identification branch. Thus, we use four different local experts corresponding to left, right, front and back orientations and any orientation can be represented by the combination of these orientations. Let $f_{\{left, right, front, back\}}$ be the output feature vectors of the 4 different orientation branches. The final re-identification output vector is the sum of the left-right component and the front-back component:

$$f_{output} = \max(\alpha, 0)f_{left} + \max(-\alpha, 0)f_{right} + \max(\beta, 0)f_{front} + \max(-\beta, 0)f_{back} \quad (1)$$

Different from the classic mixture of experts approach, our orientation gate is set before the local experts, and we perform a regression instead of a classification. The advantage of our orientation gate is that it avoids combining contradictory orientations like front and back. Computationally, only two among four orientations are used and combined according to the sign of α and β . This further allows saving computation.

3.2 Training

There are two stages to train the model as shown in Fig. 4. In the first stage, the orientation regressor and a general re-identification feature embedding are both trained in parallel with two separate objective functions. In the second stage, the network is specialized to different orientations. These two steps are detailed in the following.

Multi-task network training We start training the network with pedestrian identity labels and orientation labels respectively. **Identification:** for identification learning, we temporarily add an N -dimensional fully-connected layer to the re-identification branch, N being the number of the identities in the training

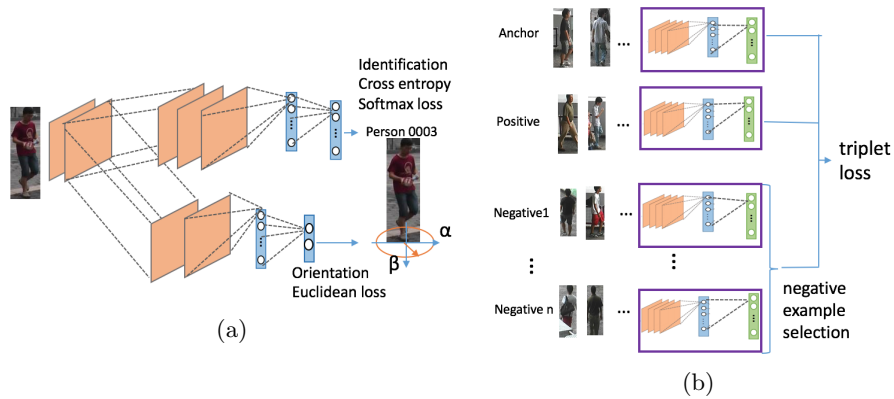


Fig. 4. The two training steps of our method. (a) In the first step, we train the model with identity and orientation labels. (b) Then, we fine-tune the model to train the orientations-specific layers with hard triplets.

set. The estimated probability of the i^{th} identity is calculated with the softmax function: $p(i) = \frac{\exp(z_i)}{\sum_{j=1}^N \exp(z_j)}$, where $z = [z_1, z_2, \dots, z_N]$ is the output of this last fully connected layer. Then, we train the CNN by minimizing the cross-entropy loss:

$$L_{id} = - \sum_{i=1}^N \log(p(i)) l_{id}(i), \quad (2)$$

where l_{id} is the ground truth one-hot coded identity vector for a given example.

Orientation regression: for the body orientation, we use the Euclidean loss to train the orientation regression of α and β . For a given training example, we have:

$$L_{orien} = \frac{(\alpha - \hat{\alpha})^2 + (\beta - \hat{\beta})^2}{2} \quad (3)$$

where $\hat{\alpha}, \hat{\beta}$ are predicted orientation labels of the example. Due to the difficulty in estimating the precise body angle, even for humans, orientation is annotated with 8 discrete labels. For training we convert the orientation class to the vector $[\alpha, \beta]$. To get a more robust orientation learning, we add a uniform random noise of 10 degrees to the orientation labels.

For datasets that have both identity and orientation labels, we train the network with a combined loss $L_{multi-task} = L_{id} + \lambda L_{orien}$. Then, orientation and identification are learned jointly. Otherwise, the two branches are trained separately.

Orientation-specific fine-tuning with triplets In the second training stage, we fine-tune the network parameters using similarity metric learning in order

to specialize the 4 different local experts. For the re-identification branch, we remove the last fully-connected layer and duplicate four times the the first fully-connected layer. Two orientation gates are integrated to select and weight different orientation projections. Since the different choices and weightings are performed according to the orientation of the person in the input image, the four orientation-specific layers are updated in different ways, whereas the other layers keep their pre-trained weights.

For the similarity metric learning, we propose to use an improved triplet loss with hard example selection. Unlike with classic triplet loss, a $(n+2)$ -tuple of images instead of a triplet is projected into the feature space. The tuple includes one anchor image \mathbf{a} , one positive image of the same person p and k negative images of different persons n^j . Training enforces that the projection of the positive example is placed closer to the anchor than the projection of the closest negative example among the k negative examples. This constraint is defined as following:

$$\min_{j=1..k} (\|f(a) - f(n^j)\|_2^2) - \|f(a) - f(p)\|_2^2 > m \quad (4)$$

The negative example that is closest to the anchor is considered the hardest example. The network is thus updated efficiently by pushing the hardest example further away from the anchor. In classic triplet loss, a part of the triplets does not violate the triplet constraint and thus is useless for learning. The selection among k negative examples reduces the number of unused training data and can make the training more efficient. To further enhance the loss function, as [4], we add a term including the distance between the anchor example and the positive example. The loss function for N training examples is defined as follows:

$$E_{triplet} = -\frac{1}{N} \sum_{i=1}^N [\max(\|f(a_i) - f(p_i)\|_2^2 - \min_{j=1..k} (\|f(a_i^j) - f(n_i^j)\|_2^2) + m, 0) + \gamma \|f(a_i) - f(p_i)\|_2] \quad (5)$$

3.3 Implementation details

The first convolutional layer has a kernel size of 5×5 and the following have a kernel size of 3×3 . All following max-pooling layers have a kernel size of 2×2 except the last one in the re-identification branch which has a kernel size of 3×1 without zero-padding increasing the number of channels and reducing the number of parameters by reducing their size to a single column. Batch normalization and a Leaky ReLU activation function with a slope of 0.2 are applied after the max-pooling layers and fully connected layers. The first fully-connected layers of the re-identification branch and the orientation gating branch output a vector of respectively 512 and 256 dimensions. Dropout is applied to the fully-connected layers to reduce the risk of overfitting. The optimization is performed by Stochastic Gradient Descent with a learning rate of 0.005, a momentum of 0.9 and a batch size of 50. The constant k is set to 5 and γ is set to 0.002 as [4].

Predicted labels	True labels							
	right	back right	back	back left	left	front left	front	front right
right	393	70	3	0	1	0	8	91
back right	94	376	58	2	0	0	2	9
back	4	133	357	14	1	0	0	3
back left	1	1	44	179	24	6	2	1
left	0	0	4	100	355	59	2	0
front left	0	1	5	10	90	490	77	1
front	2	2	7	1	3	208	484	24
front right	41	2	5	0	0	3	59	209

Fig. 5. Orientation confusion matrix on Market-1203.

4 Experiments

4.1 Datasets

The **Market-1501 Dataset** [28] is one of the largest publicly available datasets for human re-identification with 32668 annotated bounding boxes of 1501 subjects. All images are resized to 128×48 . The dataset is split into 751 identities for training and 750 identities for testing as in [28].

The **Market-1203 Dataset** [16] is a subset of Market-1501 containing 8570 images from 1203 identities under two camera views. 8 body orientations are annotated. We use 601 identities for training and 602 identities for the test. The test on Market-1203 is performed in the way as Market-1501, that means, we pick one image for each identity and each camera view as query (if there’s only one image, no image will be picked) and the rest as gallery images. The gallery images from the same identity and the same camera view as the query will be considered as “junk images” which have zero impact on search accuracy. The rank 1 accuracy (R1) and the mean average precision (mAP) are used for performance evaluation.

The **CUHK01 Dataset** [10] contains 971 subjects, each of which has 4 images under 2 camera views. We manually annotated each image with 8 body orientations. According to the protocol in [1], the data set is divided into a training set of 871 subjects and a test set of 100 and the extra data from the CUHK03 dataset [11] is also used in training. The CUHK03 dataset is a large person re-identification dataset with 13164 images of 1360 identities. We evaluate in two ways with only CUHK01 data and with CUHK01 plus CUHK03 data in training. The images are all resized to 160×60 . The Cumulative Match Curve (CMC) is employed as evaluation measure.

For all datasets, to reduce over-fitting, we perform data augmentation by randomly flipping the images and by cropping central regions with random perturbation. For the tests on Market-1501 and on CUHK01 with extra data from CUHK03, since only a part of the images has orientation annotations, the re-identification branch and the orientation gating branch are trained separately. For the test on Market-1203 and the one using only the CUHK01 dataset, we

Methods	R1	mAP
Baseline	62.0	64.6
OSCNN	63.8	66.4

Table 1. Experimental evaluation on the Market-1203 dataset.

Methods	R1	mAP
Baseline	77.3	53.9
OSCNN	78.9	55.2
OSCNN+re-rank [30]	83.9	73.5
LOMO+XQDA [13]	43.8	22.2
PersonNet [24]	37.2	18.6
Gated SCNN [4]	65.9	39.6
Divide fues re-rank [26]	82.3	72.4
LSRO [29]	78.1	56.2
DeepContext [8]	80.3	57.5
K-reciprocal re-rank [30]	77.1	63.6
SVDnet [21]	82.3	62.1
JLML [12]	85.1	65.5

Table 2. Experimental evaluation on the Market-1501 dataset.

perform a joint multi-task training with the combined loss from Section.3.2 and $\lambda = 0.01$ determined by a cross-validation.

4.2 Experimental results

Orientation regression evaluation. We first evaluate the performance of orientation regression. We tested the model after the first training stage on Market-1203 dataset. The confusion matrix is shown in Fig. 5. We calculated also the accuracy rate proposed in [16], *i.e.* result is considered correct if the predicted and true orientation classes are equal or adjacent. Since person appearances obtained in adjacent orientations are very similar, the exact orientation is less important. Thus, this accuracy evaluation criterion is more suitable for the person re-identification problem. On the Market-1203 test set, we can get an accuracy rate of 97.7%.

Orientation gate evaluation To evaluate the effectiveness of our OSCNN, we set up a baseline method. The baseline performs identity learning with soft-max loss, then fine-tuning on hard triplets without the orientation gate. The results on Market-1203, Market-1501 and CUHK01 are respectively shown in Tables 1, 2 and 3. Compared to the baseline, integrating the orientation-based local experts in the CNN framework could achieve a 1.8% point improvement for rank1 on CUHK01, 1.6% and 1.3% points for rank 1 and mAP on Market-1501 and 1.8% and 1.8% points for rank 1 and mAP on Market-1203 . This demonstrates the effectiveness of the orientation gate and the specific projections into a common feature subspace.

Methods	R1	R5	R10	R20
Basline(CUHK01)	76.6	93.8	97.0	98.8
OSCNN(CUHK01)	78.2	94.1	97.3	99.1
OSCNN(CUHK01+03)	83.5	96.4	99.0	99.5
LOMO+XQDA [13]	63.2	83.9	90.1	94.2
ImporvedDL [1]	65.0	88.7	93.1	97.2
PersonNet [24]	71.1	90.1	95	98.1
Deep Embedding [19]	69.4	-	-	-
Norm X-Corr [20]	81.2	-	97.3	98.6
Multi-task [3]	78.5	96.5	97.5	-

Table 3. Experimental evaluation on the CUHK01 dataset.

Comparison with state-of-the-art We compared our OSCNN to the state-of-the-art approaches on Market-1501 and CUHK01. Following the test protocol in [1, 24, 3], we added also the CUHK03 images to the training for the test on the CUHK01 and we compared to the methods only using these two datasets for training. As Table. 3 shows, our method is superior to most results in the state-of-the-art. Even without much extra CUHK03 training data, our method shows a competitive performance.

On the Marke-1501 dataset, our OSCNN achieves the same level results as some state-of-the-art methods. Although the result is under the best score of the state-of-the-art, the advantage of our approach is that the model doesn't need a pre-training step with a much larger pre-training dataset composed of ImageNet as [29, 8, 30, 21]. And our model has less complexity (1.15×10^8 FLOPs of our model compared to 1.45×10^9 FLOPs of JLML and to 3.8×10^9 FLOPs of SVDNet). Recently some state-of-the-art approaches show the re-ranking [30, 26] which uses information from nearest neighbors in the gallery can significantly improve the performance. As Table. 2 shows, our approach can largely benefit from this technique and achieves a state-of-the-art result on Market-1501.

5 Conclusion

In this paper, we presented a person re-identification approach based on an orientation specific CNN architecture and learning framework. Four orientation-based local experts are trained to project pedestrian images of specific orientations into a common feature subspace. An orientation gating branch learns to predict the body orientation and an orientation gate unit uses the estimated orientation to select and weight the local experts to compute the final feature embedding. We experimentally showed that the orientation gating improves the performance of person re-identification, and our approach outperforms most of the previous state-of-the-art re-identification methods on two public benchmarks.

Acknowledgement

This work was supported by the Group Image Mining (GIM) which joins researchers of LIRIS Lab. and THALES Group in Computer Vision and Data Mining. We thank NVIDIA Corporation for their generous GPU donation to carry out this research.

References

1. Ahmed, E., Jones, M., Marks, T.K.: An improved deep learning architecture for person re-identification. In: CVPR. pp. 3908–3916 (2015)
2. Bak, S., Zaidenberg, S., Boulay, B., Bremond, F.: Improving person re-identification by viewpoint cues. In: IEEE International Conference on Advanced Video and Signal Based Surveillance (AVSS). pp. 175–180. IEEE (2014)
3. Chen, W., Chen, X., Zhang, J., Huang, K.: A multi-task deep network for person re-identification. In: AAAI. pp. 3988–3994 (2017)
4. 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. pp. 1335–1344 (2016)
5. García, J., Martinel, N., Foresti, G.L., Gardel, A., Micheloni, C.: Person orientation and feature distances boost re-identification. In: International Conference on Pattern Recognition (ICPR). pp. 4618–4623. IEEE (2014)
6. Gray, D., Tao, H.: Viewpoint invariant pedestrian recognition with an ensemble of localized features. In: ECCV. pp. 262–275 (2008)
7. Koestinger, M., Hirzer, M., Wohlhart, P., Roth, P.M., Bischof, H.: Large scale metric learning from equivalence constraints. In: CVPR. pp. 2288–2295 (2012)
8. Li, D., Chen, X., Zhang, Z., Huang, K.: Learning deep context-aware features over body and latent parts for person re-identification. In: CVPR. pp. 384–393 (2017)
9. Li, W., Wang, X.: Locally aligned feature transforms across views. In: CVPR. pp. 3594–3601 (2013)
10. Li, W., Zhao, R., Wang, X.: Human reidentification with transferred metric learning. In: ACCV (2012)
11. Li, W., Zhao, R., Xiao, T., Wang, X.: Deepreid:deep filter pairing neural network for person re-identification. In: CVPR. pp. 152–159 (2014)
12. Li, W., Zhu, X., Gong, S.: Person re-identification by deep joint learning of multi-loss classification. In: International Joint Conference on Artificial Intelligence (2017)
13. Liao, S., Hu, Y., Zhu, X., Li, S.Z.: Person re-identification by local maximal occurrence representation and metric learning. In: CVPR (2015)
14. Lisanti, G., Masi, I., Del Bimbo, A.: Matching people across camera views using kernel canonical correlation analysis. In: Proceedings of the International Conference on Distributed Smart Cameras. p. 10. ACM (2014)
15. Ma, L., Yang, X., Tao, D.: Person re-identification over camera networks using multi-task distance metric learning. *IEEE Transactions on Image Processing* 23(8), 3656–3670 (2014)
16. Ma, L., Liu, H., Hu, L., Wang, C., Sun, Q.: Orientation driven bag of appearances for person re-identification. *arXiv preprint arXiv:1605.02464* (2016)
17. Pedagadi, S., Orwell, J., Velastin, S., Boghossian, B.: Local fisher discriminant analysis for pedestrian re-identification. In: CVPR. pp. 3318–3325 (2013)

18. Sarfraz, M.S., Schumann, A., Wang, Y., Stiefelhagen, R.: Deep view-sensitive pedestrian attribute inference in an end-to-end model. In: British Machine Vision Conference (BMVC) (2017)
19. Shi, H., Yang, Y., Zhu, X., Liao, S., Lei, Z., Zheng, W., Li, S.Z.: Embedding deep metric for person re-identification: A study against large variations. In: ECCV. pp. 732–748 (2016)
20. Subramaniam, A., Chatterjee, M., Mittal, A.: Deep neural networks with inexact matching for person re-identification. In: NIPS. pp. 2667–2675 (2016)
21. Sun, Y., Zheng, L., Deng, W., Wang, S.: Svdnet for pedestrian retrieval. In: International Conference on Computer Vision (2017)
22. Variator, R.R., Haloi, M., Wang, G.: Gated siamese convolutional neural network architecture for human re-identification. In: ECCV. pp. 791–808 (2016)
23. Verma, A., Hebbalaguppe, R., Vig, L., Kumar, S., Hassan, E.: Pedestrian detection via mixture of cnn experts and thresholded aggregated channel features. In: ICCV Workshops. pp. 163–171 (2015)
24. Wu, L., Shen, C., Hengel, A.v.d.: Personnet: Person re-identification with deep convolutional neural networks. arXiv preprint arXiv:1601.07255 (2016)
25. Yi, D., Lei, Z., Liao, S., Li, S.Z.: Deep metric learning for person re-identification. In: International Conference on Pattern Recognition. pp. 34–39 (2014)
26. Yu, R., Zhou, Z., Bai, S., Bai, X.: Divide and fuse: A re-ranking approach for person re-identification. In: British Machine Vision Conference (BMVC) (2017)
27. Zhao, H., Tian, M., Sun, S., Shao, J., Yan, J., Yi, S., Wang, X., Tang, X.: Spindle net: Person re-identification with human body region guided feature decomposition and fusion. In: CVPR. pp. 1077–1085 (2017)
28. Zheng, L., Shen, L., Tian, L., Wang, S., Wang, J., Tian, Q.: Scalable person re-identification: A benchmark. In: International Conference on Computer Vision (2015)
29. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: International Conference on Computer Vision (2017)
30. Zhong, Z., Zheng, L., Cao, D., Li, S.: Re-ranking person re-identification with k-reciprocal encoding. In: CVPR (2017)