Journal of Electronic Imaging

JElectronicImaging.org

Deep triplet-group network by exploiting symmetric and asymmetric information for person reidentification

Benzhi Yu
Ning Xu

SPIE. IS&T imaging.org

Benzhi Yu, Ning Xu, "Deep triplet-group network by exploiting symmetric and asymmetric information for person reidentification," *J. Electron. Imaging* **27**(3), 033033 (2019), doi: 10.1117/1.JEI.27.3.033033.

Deep triplet-group network by exploiting symmetric and asymmetric information for person reidentification

Benzhi Yu[a,b,*] **and Ning Xu**[a,b]
[a]Wuhan University of Technology, School of Computer Science and Technology, Wuhan, China
[b]Wuhan University of Technology, Hubei Key Laboratory of Transportation Internet of Things, Wuhan, China

Abstract. Deep metric learning is an effective method for person reidentification. In practice, impostor samples generally possess more discriminative information than other negative samples. Specifically, existing triplet-based deep-learning methods cannot effectively remove impostors, because they cannot consider congeners of impostor and it may produce new impostors when removing existing impostors. To utilize discriminative information in triplets and make impostor and its congeners more clustering, we design oversymmetric and oversymmetric relationships and apply these two constraints to triplet and impostors' congeners to train our deep triplet-group network with original individual images rather than handcrafted features. Extensive experiments with five benchmark datasets demonstrate that our method outperforms the state-of-the-art methods with regards to the rank-*N* matching accuracy. [DOI: 10.1117/1.JEI.27.3.033033]

Keywords: person reidentification; deep metric learning; deep learning; impostor; symmetric and asymmetric information.

Paper 171038 received Jan. 1, 2018; accepted for publication May 11, 2018; published online Jun. 9, 2018.

1 Introduction

Person reidentification (PR-ID) is a very important branch of computer vision and has been widely used in many safety-critical applications, such as video surveillance and forensics. The basic task of PR-ID shown in Fig. 1 is to determine whether or not two images from nonoverlapping cameras show the same person of interest. However, in real-world applications, there are many significant challenges for PR-ID because an image pair of a person is usually captured by different cameras with significantly different backgrounds, levels of illumination, viewpoints, occlusions, and image resolutions. To overcome these issues, many PR-ID methods have been proposed in recent years and can be generally classified into two categories: feature representation[1,2] and metric learning methods.[3,4] For feature representation methods, Schwartz and Davis[1] proposed a high-dimensional feature extraction algorithm. Baltieri et al.[2] proposed a view-independent signature method by mapping the local descriptors extracted from RGB-D sensors on an articulated body model. The pose priors and subject-discriminative features were used to reduce the effects of viewpoint changes.[5] Li et al.[6] proposed a cross-view multilevel dictionary learning model to improve the representation power, which contains dictionary learning at different representation levels, including image level, horizontal part level, and patch level. For metric learning methods, Cheng et al.[3] introduced a new and essential ranking graph Laplacian term, which can minimize the intrapedestrian compactness and maximize the interpedestrian dispersion. Li and Wang[7] presented a method that learns different metrics from the images of a person obtained from different cameras. In addition, Jing et al.[4] combined semicoupled low-rank discriminant dictionary learning to achieve super-resolution PR-ID, and Li et al.[8] also proposed for low-resolution PR-ID, which jointly learns a pair of dictionaries and a mapping to bridge the gap across lower and higher resolution images to incorporate positive and negative pair information and using the projective dictionary to boost PR-ID efficiency.

With the development of deep-learning methods, deep representation learning has recently achieved great success due to its highly effective learning ability. Several deep PR-ID models achieve a great improvement in the accuracy, such as deep metric learning (DML) for practical PR-ID,[9] a multitask deep network (MDN) for PR-ID,[10] and a deep linear discriminant analysis of Fisher networks for PR-ID.[11] However, existing deep-learning-based methods require learning a deep metric network by maximizing the distance among interclass samples and minimizing the distance among intraclass samples simultaneously. These methods do not effectively use the discriminant information among different samples. Therefore, triplet-based PR-ID models have been proposed to improve the efficiency of exploiting discriminant information through three samples, including a multiscale triplet CNN,[12] distance metric learning with asymmetric impostors (LISTEN),[13] and a body-structure-based triplet convolutional neural network.[14]

Although these triplet-based methods can improve the performance of PR-ID, they did not consider constraint from impostors' congeners samples (IC samples). As shown in Fig. 2, some new impostors may be produced when removing existing impostors by existing impostor-based methods. Therefore, how to alleviate effects of these samples is an important problem on PR-ID.

1.1 *Motivation*

Research in Refs. 12–14 has demonstrated that triplet-based methods can develop more discriminant information than

*Address all correspondence to: Benzhi Yu, E-mail: yubzh_whut@163.com

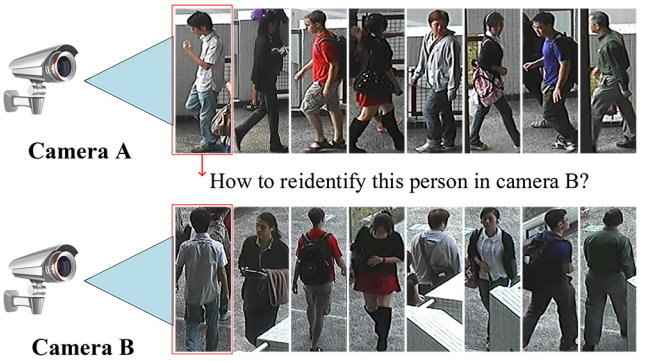

Fig. 1 Illustration of the basic task of PR-ID.

that in pairwise-based methods. However, existing triplet-based methods cannot solve difficulties caused by IC samples, such as they are transformed to new impostors, or they would be dispersed after projection. They cannot fully use the different discriminant information contained in IC samples. To address this problem, two aspects are needed to be considered in triplet-based methods. (i) Existing triplet-based methods[12–14] exploit information in impostors alone without IC samples. (ii) Impostor and its congeners maybe dispersed after projections, which must reduce the matching accuracy for PR-ID. (iii) Most deep PR-ID models are limited to handcrafted features in images by DML instead of the convolution of original images.

1.2 *Contributions*

The major contributions of this study are summarized as follows.

1. We propose a deep triplet-group network that fully employs symmetric and asymmetric information (DSAN) for triplets and IC samples (denoted as triplet group), which learns a deep neural network by the convolution of the original images of a person and trains the network with a symmetric and asymmetric constraint loss function to ensure the clustering effect of impostor and its congeners and make them more efficient and discriminable.
2. We design a triplet-group constraint objective function that requires the distance between a negative pair to be larger than that between a positive pair, and the distances between impostor and its congeners (denoted as impostor-group) are minimized simultaneously.
3. We conduct a number of matching accuracy experiments in this study. The experimental results show that our DSAN approach outperforms various triplet-based methods and other deep-learning methods.

2 Preliminary Knowledge

The corresponding relationships between an impostor and its relevant positive sample pair can be classified into two cases: a symmetric correspondence relationship and an asymmetric correspondence relationship (ACR). Given an impostor x_k and the corresponding positive sample pair $\langle x_i, x_j \rangle$, if x_k is an impostor of x_i with respect to x_j and an impostor of x_j with respect to x_i, the corresponding relationship between x_k and $\langle x_i, x_j \rangle$ is symmetric, as shown in Fig. 2(a). Otherwise, the correspondence relationship is asymmetric, as shown in Fig. 2(b). The ratio of impostors in some PR-ID datasets is presented in Ref. 13, and we can see the importance of impostors for PR-ID. For the distance between two samples $\langle x_i, x_j \rangle$, we compute the Euclidean distance $d(i,j)$ as follows:

$$d(i,j) = \|x_i - x_j\|_F^2, \tag{1}$$

where $\| * \|_F$ is the Fibonacci normalization.

2.1 *Existing Triplet-Based Methods*

The impostor-based metric learning method[15–17] exploits the impostors with a "normal" triplet constraint [i.e., for a triplet $\langle i,j,k \rangle$, it requires $d(i,j) < d(i,k)$, where $d(*)$ is a distance function], meaning that they cannot effectively remove the impostors in the case of an ACR. For this reason, Zhu et al.[13] proposed LISTEN; it requires that $d(i,k) \gg d(i,j)$ and $d(j,k) \gg d(i,j)$ simultaneously. However, LISTEN does not consider the relationship between $d(i,k)$ and $d(j,k)$ and other samples in a same class with k. This may lead to producing another impostor when removing the existing impostors, as in Figs. 2(a) and 2(b).

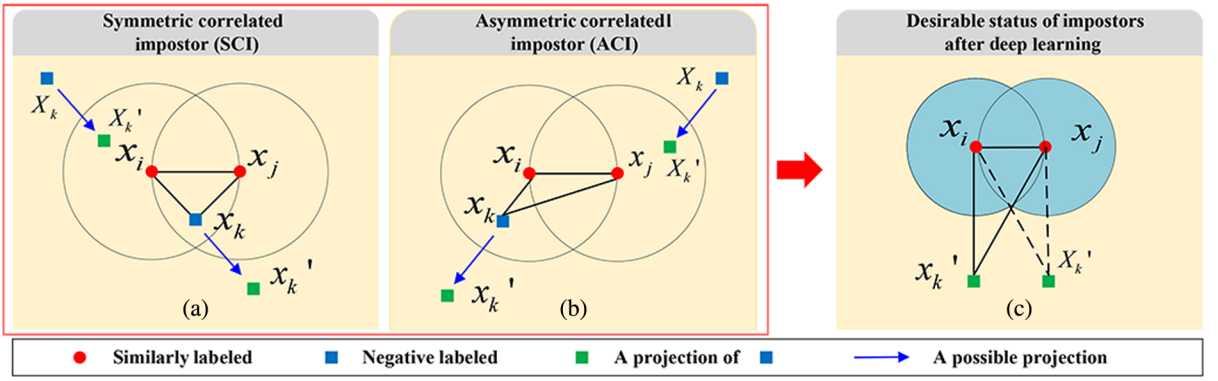

Fig. 2 Nonlinear projection of triplet samples and the desired status. $\langle x_i, x_j \rangle$ is a positive pair while x_k is a impostor and X_k is a collection of samples in a same class with x_k. X_k' and x_k' are projections of X_k and x_k, respectively.

2.2 *Our Overasymmetric and Oversymmetric Relationship Constraints on Triplet*

In our method, we transform the symmetric correlated impostor and asymmetric correlated impostor (Fig. 2) in two cases when an overasymmetric relationship (OAR) and an oversymmetric relationship (OSR) meet on positive pair and IC samples. Given a impostor x_k with its congeners $X_k = \{x_k^1, x_k^2, \cdots, x_k^{N_k}\}$ in a same class and the corresponding positive sample pair $\langle x_i, x_j \rangle$, we want them to become the desirable status as Fig. 2(c) regardless of their previous status, which make d_{ij} a very short distance as much as possible, and d_{ik} and d_{jk} are very long distance as much as possible. To some extreme degree, the correlation in a triplet $\{i, j, k\}$ can be considered as symmetric relationship because d_{ik} and d_{jk} are extremely longer than d_{ij}. Meanwhile, we make X_k' be clustering to x_k' for better classification in class k to avoid circumstances in Figs. 2(a) and 2(b).

3 Proposed Method

We proposed our deep triplet-group network and a person reidentification method for our proposed and details will be described below.

3.1 *Deep Triplet-Group Network*

For our deep triplet-group network, we use a deep convolutional network inspired by Schroff et al.[18] The network architecture is outlined in Fig. 3. We use $M+1$ layers, where the last layer is our OAR and OSR loss function. The input of the network is the triplet samples with impostor's congeners, and for image x_i, the output of the first layer is $h_i^1 = \sigma(W^1 x_i + b^1)$, where W^1 is the projection matrix, b^1 is the bias vector to be learned in the first layer of our network, and σ is a nonlinear active function that is applied in a component-wise manner. $h_i^2 = \sigma(W^2 h_i^2 + b^2)$, where W^2 is the projection matrix and b^2 is the bias vector to be learned in the second layer of our network. Similarly, the output for the m'th layer $(1 \le m \le M)$ is $h_i^m = \sigma(W^m h_i^{m-1} + b^m)$, and that for the top layer is

$$h_i^M = \sigma(W^M h_i^{M-1} + b^M), \tag{2}$$

where W^M is the projection matrix and b^M is the bias vector to be learned in the top layer of our network.

According to Eq. (1), we compute the distance between the outputs of the M'th layer from x_i and x_j as follows:

$$d(h_i^M, h_j^M) = \|h_i^M - h_j^M\|_F^2, \tag{3}$$

where h_i^M and h_j^M are the outputs of the network with inputs of x_i and x_j, respectively.

To increase the image classification performance, we expect all positive pair and IC-sample outputs through the network will simultaneously satisfy the OAR and OSR constraints. Assume a desired status, the impostor x_k should leave x_i and x_j, a maximal distance simultaneously, and we can consider there will be a symmetric relationship between x_i, x_j, and x_k. However, it is hard to meet this symmetric relationship in reality, and we develop this symmetric relationship on a cluster center u_k of impostor x_k and its congeners (denoted impostor group as X_k), which could not only maintain the asymmetric relationship in triplet but also exploit some discriminative information in its congeners to make impostor group more discriminative. In other words, our developed strategy ensures X_k meets OAR constraint and OSR constraint between x_i and x_j. In our network, for each triplet group $\langle x_i, x_j, x_k \rangle$ and congeners X_k of x_k, the outputs $\langle h_i^M, h_j^M, h_k^M \rangle$ and u_k^M satisfy the following objective function:

$$\begin{aligned} \min J = & \|d(h_i^M, u_k^M) - d(h_j^M, u_k^M)\|_F^2 \\ & - \|d(h_i^M, u_k^M) - d(h_i^M, h_j^M)\|_F^2 \\ & + \alpha d(h_i^M, h_j^M) - \beta d(h_i^M, h_k^M), \end{aligned} \tag{4}$$

where u_k^M is the cluster center of all samples in class k, including x_k, and $\|d(h_i^M, u_k^M) - d(h_j^M, u_k^M)\|_F^2$ is the OSR term. OSR term makes the distance between u_k and x_i and the distance between u_k and x_j equal to meet OSR constraint. $\|d(h_i^M, u_k^M) - d(h_i^M, h_j^M)\|_F^2$ is the OAR term. OAR term makes the distance between u_k and x_i larger than the distance between x_i and x_j to meet OAR constraint. In addition, $d(h_i^M, h_j^M)$ is the intraclass term to minimize the distance between samples in the same class, and

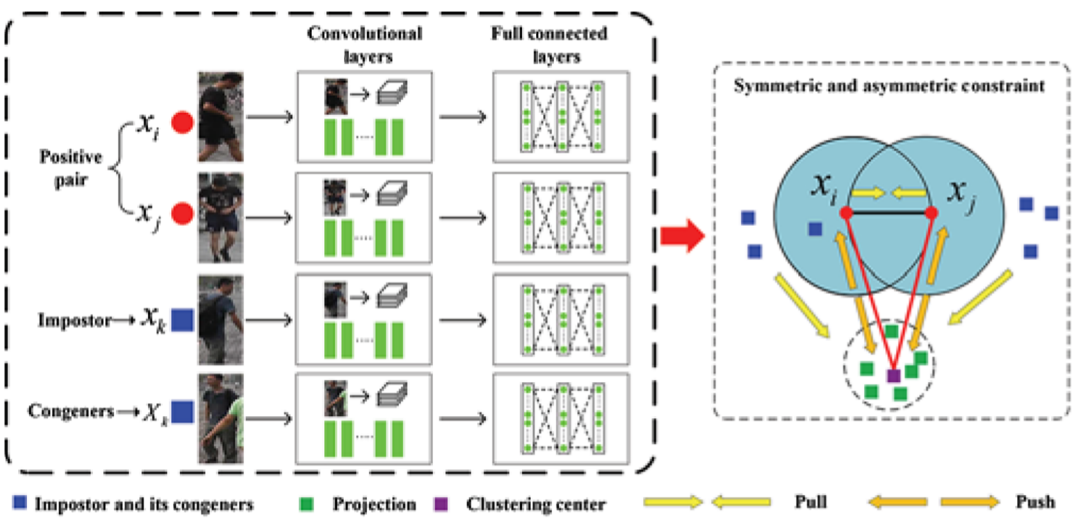

Fig. 3 Basic idea of our DSAN.

$-d(h_i^M, h_k^M)$ is the interclass term to maximize the distance between samples in different classes. α and β are the balance parameters of $d(h_i^M, h_j^M)$ and $d(h_i^M, h_k^M)$. Let $f = \{W^1, W^2, \ldots, W^M, b^1, b^2, \ldots, b^M\}$ be the parameters of our network. We formulate the following optimization problem to maximize the margin between the all triplet samples:

$$\min_f H = g\left(\sum_{\langle i,j,k\rangle \in T} J\right) + \frac{\gamma}{2}\sum_{m=1}^{M}(\|W^m\|_F^2 + \|b^m\|_2^2), \quad (5)$$

where T is the collection of triplet-group samples, γ is a parameter for balancing the contributions of different terms, and $g(a)$ is the generalized logistic loss function that smoothly approximates the hinge loss function $a = \max(a, 0)$ and is defined as follows:

$$g(a) = \frac{1}{\rho}\log[1 + \exp(\rho a)], \quad (6)$$

where ρ is the sharpness parameter. Details of our algorithm are demonstrated in Algorithm 1.

Algorithm 1 Our DSAN algorithm

Input: Training set X, number of network layers $M + 1$, learning rate μ, parameters α and β, and convergence error ε;

Output: Parameters W^m and b^m, $1 \le m \le M$.

Initialization: Initialize W^m and b^m with appropriate values

for $k = 1, 2, \cdots, K$ **do**

Compute the triple-group collection T

for $l = 1, 2, \cdots, M$ **do**

Compute h_i^l, h_j^l, and h_k^l-group using the deep network.

end

for $l = M, M-1, \cdots, 1$ **do**

Obtain the gradients according to backpropagation algorithm.

end

for $l = 1, 2, \cdots, M$ **do**

Update W^m and b^m according to forward propagation algorithm

end

Calculate H_k using Eq. (5).

If $k > 1$ and $\|H_k - H_{k-1}\| < \varepsilon$, go to **Return**.

end

Return: W^m and b^m, where $1 \le m \le M$.

3.2 *Person Reidentification Method*

For the image y of a pedestrian in probe from testing image set, we use y as the input of our network with the learned parameter f and obtain its deep feature representation h_y^M. Then, we compute the distances between h_y^M and each image in the gallery from testing image set by Eq. (3). Finally, we choose the smallest distance in every distance, including h_y^M, and obtain the label of the sample that has the smallest distance with h_y^M as follows:

$$\text{Label}_y = \arg\min_c(y, x_c) \cdot 1 \le c \le C, \quad (7)$$

where c is the class of x_c and C is the total number of classes in the training image set.

4 Experiments

We conducted extensive experiments using five widely used datasets: CUHK03,[19] CUHK01,[20] VIPeR,[21] iLIDS-VID,[22] and PRID2011.[23] Here, we compare the performance of our approach with triplet-based state-of-the-art approaches.

4.1 *Datasets and Experimental Settings*

Experiments are conducted with one large dataset and four small datasets. The large dataset is the CUHK03 dataset, which contains 13,164 images from 1360 persons. We randomly selected 1160 persons for training, 100 persons for validation, and 100 persons for testing, following exactly the same settings in Refs. 19 and 24. The four small datasets are the CUHK01, VIPeR, iLIDS, and PRID2011 datasets. For these four datasets, we randomly divided the individuals into two equal parts, with one used for training and the other for testing. Moreover, we created triplet collections following the method by Schroff et al.[18]

To validate the effectiveness of our DSAN approach, we compare the DSAN model with several state-of-the-art metric-learning-based methods: keep it simple and straightforward metric learning (KISSME)[25] and relaxed pairwise metric learning (RPML).[26] In addition, our DSAN model was compared with several state-of-the-art deep-learning-based methods: the improved deep-learning architecture (IDLA),[24] deep ranking PR-ID (DRank),[27] and an MDN (MTDnet).[10] Moreover, our DSAN model was compared with some state-of-the-art triplet-based networks: efficient impostor-based metric learning (EIML),[17] LISTEN,[13] an improved triplet loss network (ImpTrLoss),[28] and a spindle Net.[29]

4.2 *Implementation Details*

For evaluating our DSAN, we use TensorFlow[30] framework to train our DASN. Note that we used network configuration as in Ref. 18. For all datasets, our network contains six convolutional layers, four max polling layers, and one fully connected (FC) layers for each images. These layers configured as below.(1) Conv. 7×7, stride = 2, feature maps = 64; (2) Max pool 3×3, stride = 2; (2) Max pool 3×3, stride = 2; (3) Conv.3×3, stride = 1, feature maps = 192; (4) Max pool 3×4, stride = 2; (5) Conv.3×3, stride = 1, feature maps = 384; (6) Max pool 3×3, stride = 2; (7) Conv.3×3, stride = 1, feature maps = 256; (8) Conv.3×3, stride = 1, feature maps = 256;

(9) Conv.3 × 3, stride = 1, feature maps = 256; and (10) FC, output dimension = 128.

For small datasets, we adopt an unsupervised image generating strategy[31] to solve the problem of lacking training samples. In detail, we use small dataset as source domain and map 10,000 images in CUHK03 dataset into source domain. This strategy makes the 10,000 images follow distribution of target small dataset. Then, we used these generated images to train our model and fine-tune with target small datasets.

4.3 *Results and Analysis*

Table 1 shows our rank-1 matching accuracies, and Figs. 4–8 describe cumulative match characteristic (CMC) curves in different ranks on five datasets. We will describe evaluations on five datasets.

Table 1 Top-ranked matching rates (%) for five datasets.

Method	CUHK03	CUHK01	VIPeR	iLIDS	PRID2011
KISSME	14.17	18.25	19.61	28.58	15.75
RPML	18.67	20.14	23.93	31.97	18.69
IDLA	54.74	65.00	45.90	58.15	43.18
DRank	45.75	70.94	38.37	52.82	45.67
MTDnet	74.68	77.50	45.89	41.04	32.03
EIML	20.18	21.34	22.04	21.75	18.06
LISTEN	23.71	32.77	39.62	32.81	53.75
ImpTrLoss	75.37	53.70	47.8	60.45	22.00
Spindle	88.5	79.9	53.8	66.3	67
DSAN	77.35	82.15	54.85	66.70	68.2

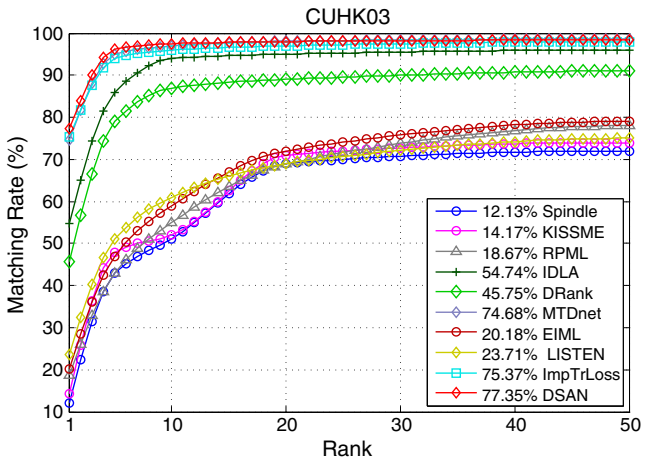

Fig. 4 CMC curves of the average matching rates for the CUHK03 dataset.

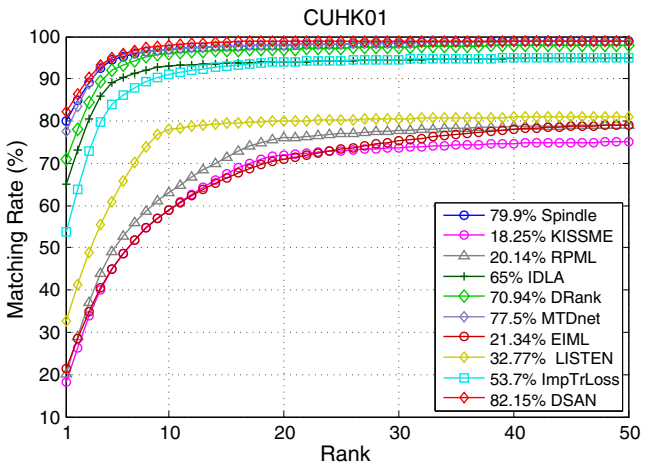

Fig. 5 CMC curves of the average matching rates for the CUHK01 dataset.

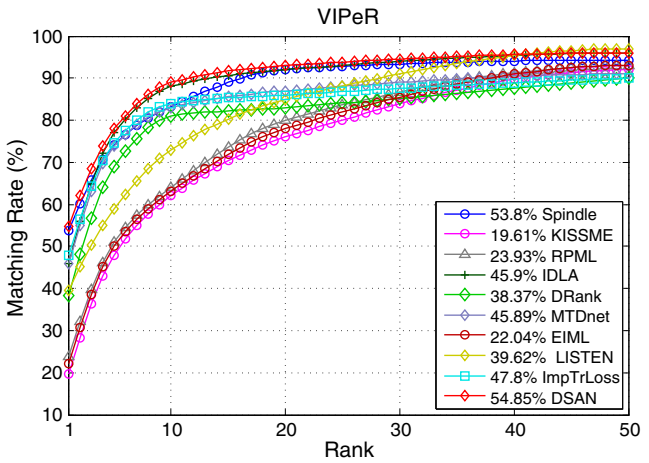

Fig. 6 CMC curves of the average matching rates for the VIPeR dataset.

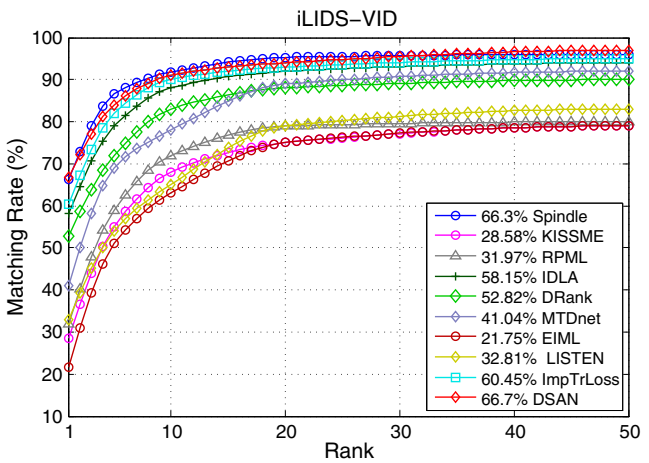

Fig. 7 CMC curves of the average matching rates for the iLIDS-VID dataset.

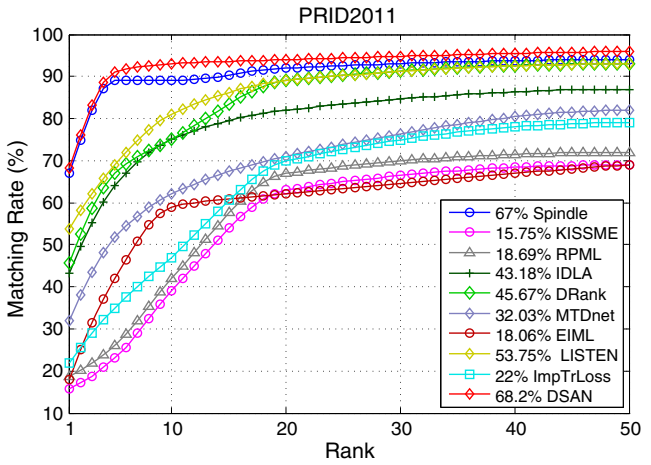

Fig. 8 CMC curves of the average matching rates for the PRID2011 dataset.

4.3.1 *Evaluation with the CUHK03 dataset*

The CUHK03 dataset contains 13,164 images of 1360 pedestrians captured by six surveillance cameras. Each identity is observed by two disjoint camera views. On average, there are 4.8 images per identity for each view. This dataset provides both manually labeled pedestrian bounding boxes and bounding boxes automatically obtained by running a pedestrian detector.[32] We report results for both versions of the data (labeled and detected). Following the protocol used in Ref. 19, we randomly divided the 1360 identities into nonoverlapping training (1160), test (100), and validation (100) sets. This yielded about 26,000 positive pairs before data augmentation. We used a minibatch size of 150 samples and trained the network for 200,000 iterations. We used the validation set to design the network architecture. In Table 1 and Fig. 4, we compare our method against KISSME, IDLA, MTDnet, ImpTrLoss and Spindle net, and it is observed that DSAN outperforms these methods with regards to the rank-1 matching accuracy except for Spindle. We achieve a rank-1 accuracy of 77.35% with the parameters $\alpha = 0.35$ and $\beta = 0.25$.

4.3.2 *Evaluation with the CUHK01 dataset*

The CUHK01 dataset has 971 identities, with two images per person for each view. Most previous papers have reported results using the CUHK01 dataset by considering 486 identities for testing. With 486 identities in the test set, only 485 identities remain for training. This leaves only 1940 positive samples for training, which makes it practically impossible for a deep architecture with a reasonable size to not overfit if trained from scratch with these data. One way to solve this problem is to use a model trained with the transformed CUHK03 dataset and then test the 486 identities of the CUHK01 dataset. This is unlikely to work well since the network does not know the statistics of the tests with the CUHK01 dataset. In fact, our model was trained with the transformed CUHK03 dataset and adapted for the CUHK01 dataset by fine-tuning it with the CUHK01 dataset with 485 training identities (nonoverlapping with the test set). Table 1 and Fig. 5 compare the performance of our approach with that of other methods. We used a minibatch size of 150 samples and trained the network for 180,000 iterations. Our method obtains a rank-1 accuracy of 79.35% with the parameters $\alpha = 0.15$ and $\beta = 0.45$, surpassing all other methods individually.

4.3.3 *Evaluation with the VIPeR dataset.*

The VIPeR dataset contains 632 pedestrian pairs with two views, with only one image per person for each view. The testing protocol is to split the dataset in half: 316 pairs for training and 316 pairs for testing. This dataset is extremely challenging for a deep neural network architecture for two reasons: (a) there are only 316 identities for training with one image per person for each view, giving a total of just 316 positives, and (b) the resolution of the images is lower (48×128 as compared to 60×160 for the CUHK01 dataset). We trained a model using the transformed CUHK03 dataset and then adapted the trained model to the VIPeR dataset by fine-tuning it with 316 training identities. Since the number of negatives is small for this dataset, hard negative mining does not improve results after fine-tuning because most of the negatives were already used during fine-tuning. The results in Table 1 and Fig. 6 show that DSAN outperforms the state-of-the-art methods by a large margin. We used a minibatch size of 150 samples and trained the network for 130,000 iterations. Our rank-1 accuracy is 49.05%, surpassing all other methods for the parameters $\alpha = 0.25$ and $\beta = 0.15$.

4.3.4 *Evaluation with the iLIDS dataset*

The iLIDS-VID dataset has 300 different pedestrians observed across two disjoint camera views in a public open space. This dataset is very challenging owing to the clothing similarities among people, the lighting, and the viewpoint variations across camera views. There are two versions: a static-image-based version and image-sequence-based version, and we chose the static images for use in our experiments. This version contains 600 images of 300 distinct individuals, with one pair of images from two camera views for each person. We divided the set into 150 individuals for training and the others for testing. In the iLIDS-VID dataset, we also encounter a similar problem, as for the CUHK01 and VIPeR datasets. We used the pretrained model using the transformed CUHK03 dataset and fine-tuned it for training with the iLIDS-VID dataset. From Table 1 and Fig. 7, DSAN outperforms the state-of-the-art methods. We used a minibatch size of 150 samples and trained the network for 180,000 iterations. Our rank-1 accuracy is 62.55% for the parameters $\alpha = 0.25$ and $\beta = 0.15$.

4.3.5 *Evaluation with the PRID2011 dataset*

This dataset has 385 trajectories from camera A and 749 trajectories from camera B. Among them, only 200 people appear in both cameras. This dataset also has a single hot version, which consists of randomly selected snapshots. The division and pretraining procedure is similar to that for the iLIDS-VID dataset: half for training and the others for testing. Furthermore, the transformed CUHK03 dataset is used to pretrain and fine-tune with the PRID2011 dataset. In our experiments, we used a minibatch size of 150 samples and trained the network for 160,000 iterations. We obtained a rank-1 accuracy of 55.86% with $\alpha = 0.25$ and $\beta = 0.15$, and the detailed results are presented in Table 1 and Fig. 8.

4.4 *Discussion*

In this section, we discuss several effects of OAR and OSR constraints, clustering enter symmetric constraint, and parameter analysis.

4.4.1 *Effects of the OAR and OSR constraints*

To evaluate the effects of the OAR and OSR constraints, we perform experiments with three datasets with or without utilization of the OAR and OSR constraints. The results obtained using DSAN without the OAR or OSR constraint are denoted as DSN and DAN, respectively. Table 2 reports the rank-1 matching rates of DSAN, DSN, and DAN for the five datasets. We can see that using OAR and OSR constraints improves the rank-1 matching rate by at least 3.55%, which indicates that our OAR and OSR constraints can exploit some discriminative information that is useful for PR-ID.

4.4.2 *Effects of our clustering center symmetric constraint*

To evaluate effects of our clustering center symmetric constraint, we conduct several experiments without clustering center symmetric constraint, which only use impostor into triplet constraint denoted as DTN. Table 1 reports the top-rank matching accuracy of our experiment and triplet-based methods (LISTEN and ImpTrLoss). It can be shown that our clustering center symmetric constraint improves by 7.081% on average

Table 2 Effects of the OAR and OSR constraints.

Method	CUHK03	CUHK01	VIPeR	iLIDS	PRID2011
DAN	62.80	74.84	39.27	49.36	48.18
DSN	59.35	67.95	34.84	50.80	47.52
DSAN	77.35	82.15	54.85	66.70	68.2

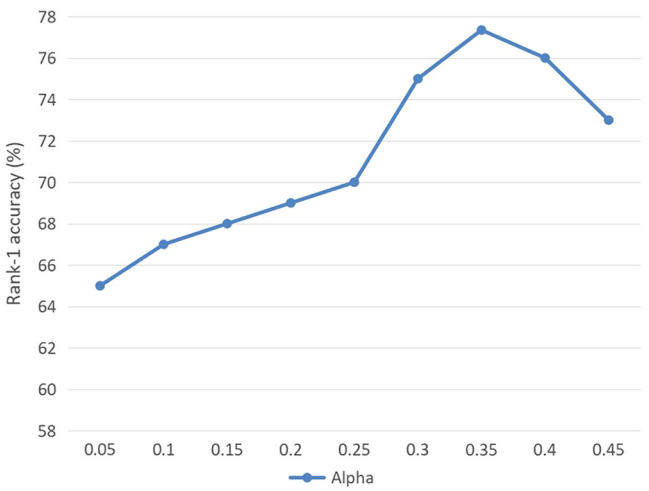

Fig. 9 Rank-1 results of DSAN with different α on CUHK03 dataset.

Table 3 Training time and testing time.

Method	DSAN	CUHK01	VIPeR	iLIDS	PRID2011
Training	62.80	74.84	39.27	49.36	48.18
Testing	59.35	67.95	34.84	50.80	47.52

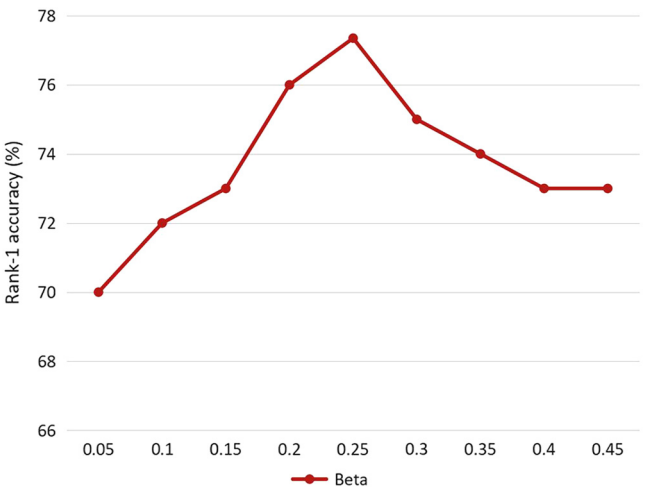

Fig. 10 Rank-1 results of DSAN with different β on CUHK03 dataset.

4.4.3 *Parameter analysis*

In this experiment, we investigate the effect of parameters, including α and β. Parameter α balances the effect of intra-class term. Parameter β controls the effect of interclass term. When one of the parameters is evaluated, the other is fixed as the values given in evaluation of datasets.

We take the experiment on CUHK03 dataset as an example. Figures 9 and 10 show the rank-1 matching rates of our approach versus different values of α and β on CUHK03 dataset. We can observe that: (1) DSAN is not sensitive to the choice of α in the range of [0.10, 0.30]; (2) DSAN achieves the best performance when α and β are set as 0.35 and 0.25, respectively; and (3) DSAN can obtain relatively good performance when β is in the range of [0.20, 0.30]. Similar effects can be observed on other datasets (Besides, the training and testing time are described in Table 3).

5 Conclusion

We have developed a deep triplet-group network by exploiting symmetric and asymmetric information on clustering center of impostor and its congeners. It differs from existing methods in that it can use the OAR and OSR constraints to exploit more discriminative information from the relationships between positive samples and its impostor clustering center. From the results of extensive experiments, we can draw the following conclusions. (1) DSAN outperforms several state-of-the-art DL-based methods in terms of the matching rate. (2) With the designed OAR and OSR constraints, DSAN can more effectively exploit discriminative information. (3) There exists some useful information in impostor-based clustering center, and the proper utilization of this information can improve performance.

References

1. W. R. Schwartz and L. S. Davis, "Learning discriminative appearance-based models using partial least squares," in *Proc. of the XXII Brazilian Symp. on Computer Graphics and Image Processing (SIBGRAPI)*, Rio de Janeiro, Brazil, pp. 322–329 (2009).
2. D. Baltieri, R. Vezzani, and R. Cucchiara, "Learning articulated body models for people re-identification," in *ACM Multimedia Conf.*, Barcelona, Spain, pp. 557–560 (2013).
3. D. Cheng et al., "Discriminative dictionary learning with ranking metric embedded for person re-identification," in *Proc. of the Twenty-Sixth Int. Joint Conf. on Artificial Intelligence (IJCAI)*, Melbourne, pp. 964–970 (2017).
4. X.-Y. Jing et al., "Super-resolution person re-identification with semi-coupled low-rank discriminant dictionary learning," *IEEE Trans. Image Process.* **26**(3), 1363–1378 (2017).
5. W.-S. Zheng, S. Gong, and T. Xiang, "Reidentification by relative distance comparison," *IEEE Trans. Pattern Anal. Mach. Intell.* **35**(3), 653–668 (2013).
6. S. Li, M. Shao, and Y. Fu, "Person re-identification by cross-view multi-level dictionary learning," *IEEE Trans. Pattern Anal. Mach. Intell.* PP (99), 1 (2017).
7. W. Li and X. Wang, "Locally aligned feature transforms across views," in *IEEE Conf. on Computer Vision and Pattern Recognition*, Portland, Oregon, pp. 3594–3601 (2013).
8. K. Li et al., "Discriminative semi-coupled projective dictionary learning for low-resolution person re-identification," in *Proc. of the Thirty-Second AAAI Conf. on Artificial Intelligence*, New Orleans, Louisiana (2018).
9. D. Yi et al., "Deep metric learning for person reidentification," in *22nd IEEE Int. Conf. on Pattern Recognition (ICPR)*, pp. 34–39 (2014).
10. W. Chen et al., "A multi-task deep network for person re-identification," in *AAAI Conf. on Artificial Intelligence* (2017).
11. L. Wu, C. Shen, and A. Van Den Hengel, "Deep linear discriminant analysis on fisher networks: a hybrid architecture for person re-identification," *Pattern Recognit.* **65**, 238–250 (2017).
12. J. Liu et al., "Multi-scale triplet CNN for person re-identification," in *Proc. of the ACM Conf. on Multimedia Conf. (MM)*, Amsterdam, The Netherlands, pp. 192–196 (2016).
13. X. Zhu et al., "Distance learning by treating negative samples differently and exploiting impostors with symmetric triplet constraint for person re-identification," in *IEEE Int. Conf. on Multimedia and Expo (ICME)*, Seattle, Washington, pp. 1–6 (2016).
14. H. Liu and W. Huang, "Body structure based triplet convolutional neural network for person re-identification," in *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, New Orleans, Louisiana, pp. 1772–1776 (2017).
15. K. Q. Weinberger and L. K. Saul, "Distance metric learning for large margin nearest neighbor classification," *J. Mach. Learn. Res.* **10**, 207–244 (2009).
16. M. Dikmen et al., "Pedestrian recognition with a learned metric," *Lect. Notes Comput. Sci.* **6495**, 501–512 (2010).
17. M. Hirzer, P. M. Roth, and H. Bischof, "Person re-identification by efficient impostor-based metric learning," in *IEEE Ninth Int. Conf. on Advanced Video and Signal-Based Surveillance*, pp. 203–208 (2012).
18. F. Schroff, D. Kalenichenko, and J. Philbin, "Facenet: a unified embedding for face recognition and clustering," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR), Boston, Massachusetts*, pp. 815–823 (2015).
19. W. Li et al., "DeepReID: deep filter pairing neural network for person re-identification," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Columbus, Ohio, pp. 152–159 (2014).
20. W. Li, R. Zhao, and X. Wang, "Human reidentification with transferred metric learning," *Lect. Notes Comput. Sci.* **7724**, 31–44 (2012).
21. D. Gray, S. Brennan, and H. Tao, "Evaluating appearance models for recognition, reacquisition, and tracking," in *10th IEEE Int. Workshop on Performance Evaluation of Tracking and Surveillance (PETS)* (2007).
22. T. Wang et al., "Person re-identification by discriminative selection in video ranking," *IEEE Trans. Pattern Anal. Mach. Intell.* **38**(12), 2501–2514 (2016).
23. M. Hirzer et al., "Person re-identification by descriptive and discriminative classification," *Lect. Notes Comput. Sci.* **6688**, 91–102 (2011).
24. E. Ahmed, M. J. Jones, and T. K. Marks, "An improved deep learning architecture for person re-identification," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, Boston, Massachusetts, pp. 3908–3916 (2015).
25. M. Köstinger et al., "Large scale metric learning from equivalence constraints," in *IEEE Conf. on Computer Vision and Pattern Recognition*, Providence, Rhode Island, pp. 2288–2295 (2012).
26. M. Hirzer et al., "Relaxed pairwise learned metric for person re-identification," *Lect. Notes Comput. Sci.* **7577**, 780–793 (2012).
27. S. Z. Chen, C. C. Guo, and J. H. Lai, "Deep ranking for person re-identification via joint representation learning," *IEEE Trans. Image Process.* **25**(5), 2353–2367 (2016).
28. D. Cheng et al., "Person re-identification by multi-channel parts-based CNN with improved triplet loss function," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 1335–1344 (2016).
29. H. Zhao et al., "Spindle net: person re-identification with human body region guided feature decomposition and fusion," in *IEEE Conf. on Computer Vision and Pattern Recognition (CVPR)*, pp. 907–915 (2017).
30. M. Abadi et al., "Tensorflow: a system for large-scale machine learning," in *Operating Systems Design and Implementation (OSDI)*, Vol. 16, pp. 265–283 (2016).
31. Y. Taigman, A. Polyak, and L. Wolf, "Unsupervised cross-domain image generation," in *Int. Conf. on Learning Representations*, pp. 1–15 (2017).
32. P. F. Felzenszwalb et al., "Object detection with discriminatively trained part-based models," *Computer* **47**(2), 6–7 (2014).

Benzhi Yu is a PhD student at the School of Computer Science and Technology of Wuhan University of Technology. His research interests include image processing, computer vision, data mining, and machine learning.

Ning Xu received his PhD in electronic science and technology from the University of Electronic Science and Technology of China, Chengdu, in 2003. Later, he was a postdoctoral fellow with Tsinghua University, Beijing, from 2003 to 2005. Currently, he is a professor at the School of Computer Science and Technology of Wuhan University of Technology, Wuhan. He research interests include computer aided design of VLSI circuits and systems, computer architectures, data mining, and highly combinatorial optimization algorithm.