

Graph Convolution for Re-ranking in Person Re-identification

Yuqi Zhang¹ Qian Qi^{1*} Chong Liu^{1,2,3} Weihua Chen¹ Fan Wang¹ Hao Li¹
 Rong Jin¹

¹ Machine Intelligence Technology Lab, Alibaba Group

² State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences, China

³ University of Chinese Academy of Sciences, Beijing 100049, China

{gongyou.zyq, qi.qian, kugang.cwh, fan.w, lihao.lh, jinrong.jr}@alibaba-inc.com

liuchong@ios.ac.cn

Abstract

Nowadays, deep learning is widely applied to extract features for similarity computation in person re-identification (re-ID) and have achieved great success. However, due to the non-overlapping between training and testing IDs, the difference between the data used for model training and the testing data makes the performance of learned feature degraded during testing. Hence, re-ranking is proposed to mitigate this issue and various algorithms have been developed. However, most of existing re-ranking methods focus on replacing the Euclidean distance with sophisticated distance metrics, which are not friendly to downstream tasks and hard to be used for fast retrieval of massive data in real applications. In this work, we propose a graph-based re-ranking method to improve learned features while still keeping Euclidean distance as the similarity metric. Inspired by graph convolution networks, we develop an operator to propagate features over an appropriate graph. Since graph is the essential key for the propagation, two important criteria are considered for designing the graph, and three different graphs are explored accordingly. Furthermore, a simple yet effective method is proposed to generate a profile vector for each tracklet in videos, which helps extend our method to video re-ID. Extensive experiments on three benchmark data sets, e.g., Market-1501, Duke, and MARS, demonstrate the effectiveness of our proposed approach.

1. Introduction

Person re-identification (re-ID) aims to retrieve images of the same person from the gallery set given a query image [54]. A standard pipeline is to extract features for

*Equal contribution

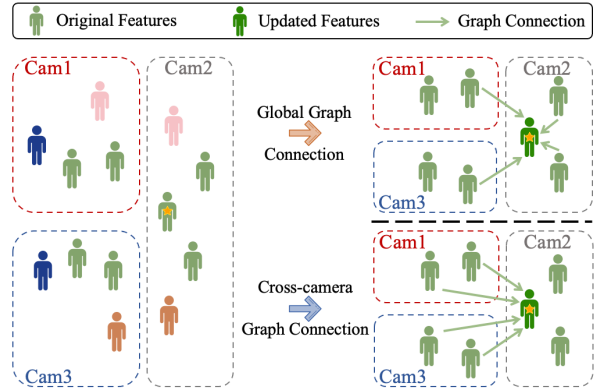


Figure 1: Illustration of graphs with two proposed criteria. The person with the star denotes the target image and the arrows indicate its k -nearest neighbors. People with the same color hold the same ID. Corresponding to the two criteria, we generate two graphs (i.e., Global graph: connecting the k -nearest neighbors in all cameras, and Cross-camera graph: connecting the k -nearest neighbors from different cameras of the target person, excluding those from the same camera).

images in both the gallery set and the query based on a pre-trained deep model, and then return the top-ranked images in the gallery, where the similarity is measured by the Euclidean distance [53]. However, due to the difference between the distribution of the training set from the deep model and that of the testing set, directly generating features based on the pre-trained model may result in a sub-optimal performance. Many post-process methods have been proposed to mitigate the challenge while re-ranking is one of the most effective approaches for outstanding performance [2, 28, 54].

Given features from the deep model, re-ranking is to recalculate the similarity of images by introducing other information and use sophisticated similarity metrics [1, 2, 3, 28, 47, 54] to rearrange the ranking list. For example, k -reciprocal encoding [54] obtains additional k -nearest neighbors (k -NN) for each image as its context information, which are utilized to recalculate the similarity between query and gallery images based on the Jaccard distance. ECN [28] shares the similar idea that leverages the nearest neighbors for top-ranked images but re-computes the similarity with a rank-list distance metric [13]. Both of these methods can surpass the performance of original features by a large margin. Despite the success, the sophisticated distance metrics adopted by these re-ranking methods are much more complicated than Euclidean distance, which are not friendly to downstream tasks and hard to be used for fast retrieval of massive data in real applications. Therefore, some work [21] tries to optimize the original features based on Euclidean distance. But their performance still cannot catch up with k -reciprocal encoding.

Instead of figuring out an appropriate and sophisticated distance metric, in this work, we aim to modify the original features while Euclidean distance can still be directly used as the similarity measure. Inspired by graph convolution networks (GCN) [15], we adopt the graph convolution operator to propagate features over a graph, so as to improve the representation of each image. More specifically, we construct our graphs for feature propagation with two criteria. First, the changes in features should be moderate after re-ranking to preserve the knowledge learned in the pre-trained feature representation model. Therefore, only features from nearest neighbors can be propagated to the target image. This criterion essentially shares a similar idea with other successful re-ranking methods [54, 28]. Second, features propagated from different cameras should be emphasized. This criterion has been rarely investigated but it is helpful to eliminate the bias from cameras. With these criteria, we develop a feature propagation method that obtains features from two graphs simultaneously.

Fig. 1 illustrates the proposed graphs with our two criteria. Both of two graphs take the k -nearest neighbors into account for each image. The difference is that in the global graph, the k -nearest neighbors of each image are from all cameras, while in the cross-camera graph, the k -nearest neighbors are from only different cameras of a given image. Then, we apply a graph convolution operator on these two graphs. After obtaining propagated features from two graphs, their weighted combination is treated as the final feature representation to re-compute the ranking list based on Euclidean distance. To the best of our knowledge, this is the first work that achieves state-of-the-art performance in re-ranking with Euclidean distance.

The main contributions of our work can be summarized

as follows.

- We propose the criteria of feature propagation for re-ranking and develop a graph convolution based re-ranking (GCR) method accordingly. The features obtained from our method are still in the Euclidean space, which can be easily used in downstream tasks and available for fast retrieval of massive data in real applications.
- Along with the GCR, to take full advantage of multi-frame information in video re-ID task, we further present a simple yet effective method to generate a profile vector for each tracklet in video re-ID, called profile vector generation (PVG).
- As the image-level re-ID task can be considered as a video re-ID with only one image in each tracklet, we combine GCR and PVG together to build our final solution, *i.e.* Graph Convolution Re-ranking for Video (GCRV), which achieves state-of-the-art performance on the ReID benchmarks in both image-level and video-level re-ID tasks.

The rest of this paper is organized as follows. Section 2 reviews the related work. Section 3 analyzes feature propagation for re-ranking and proposes our graph convolution based re-ranking method. Section 4 presents our method on how to generate profile vectors to represent tracklets for video re-ID. Section 5 provides the experiment analysis of our methods on the benchmark data sets. Section 6 concludes this work with the future direction.

2. Related Work

2.1. Person Re-Identification

Recent person re-identification methods focus on global or local feature learning. The global feature learning methods [37, 53, 5, 22, 23] are straightforward for both training and evaluation. The model applies on the entire image rather than a part of it. Apart from global features, local features [38, 35, 16, 48, 50] are also studied based on the assumption that person images are roughly aligned. The local models could alleviate the impact of occlusion or inaccurate detection, and achieve better performance.

In spite of the above spatial operations, some methods also study different loss functions. Learning features purely by softmax loss may lose discrimination [42] and thus a large amount of modifications [19, 40, 8] have been made based on softmax loss. Besides the feature learning methods, metric learning methods are also studied for re-ID including contrastive loss [10] and triplet loss [29]. The widely-used open-source re-ID strong baseline [22] uses both softmax loss and triplet loss. Equipped with some tricks including random erasing [55], dropout [32] and so on, the baseline sets competitive results in the recent years. We use this baseline for re-ID feature extraction to make a

fair comparison with other re-ranking methods.

2.1.1 Re-ranking based Methods

Re-ranking is often used as a post-processing step to improve the initial ranking results. Since the original top-ranked data might be polluted by false positives, k -reciprocal nearest neighbors [14, 26] can be regarded as highly relevant candidates. The concept of k -reciprocal nearest neighbor was introduced to re-ID by Zhong et al. [54]. The Jaccard distance of k -reciprocal encodings can be used as a strong complementary ranking cue. Although k -reciprocal provides state-of-the-art performance, it has to recompute the reciprocal rank lists for each image pair. Sarfraz et al. [28] reinforced the original pairwise distance by aggregating distances between expanded neighbors of image pairs. This results in a more effective re-ranking framework since no re-computation for each image pair is required.

There are a series of similarity propagation methods for re-ranking. Bai et al. [3] proposed Regularized Ensemble Diffusion (RED) to maximize the smoothness of multiple graph-based manifolds by performing similarity learning and weight learning simultaneously. Yu et al. [47] fused distances between different sub-features in a “Divide and Fuse” framework. Bai et al. [1] proposed Supervised Smoothed Manifold (SSM) to tackle person re-identification task on the data manifold. The similarity value between two instances could be estimated in the context of other pairs of instances. Recently, Bai et al. [2] designed Unified Ensemble Diffusion (UED) for metric fusion. UED optimizes a new objective function and derivation, maintaining the advantages of three existing fusion algorithms.

Besides refining rank lists by original re-ID features, researchers also consider attribute [33, 34, 49, 17], space-time [39] and face [20] to improve person re-identification. These methods can also be regarded as a re-ranking stage. Su et al. [33] embedded original binary attributes to a continuous attribute space, where incorrect and incomplete attributes are rectified and recovered to better describe people. Lin et al. [17] systematically investigated how person re-ID and attribute recognition benefit from each other. Wang et al. [39] eliminated lots of irrelevant images with the help of the spatial-temporal constraint and thus narrowed the gallery database.

2.1.2 Graph based Methods

There are also graph-based feature learning works [30, 31, 21] in recent years. They applied similarity transformation on the graph to model the similarity relationship among the

query image and the gallery image. Shen et. al. [31] proposed Similarity-Guided Graph Neural Network (SGGNN) to incorporate the pair-wise gallery-gallery similarity information into training process of person re-identification. However, these methods are often embedded into the training phase and thus could not serve as a re-ranking method in post-process stage. What’s more, in their methods, they do not take the cross-camera relationship into account. While in our method, we design a special graph to handle the cross-camera relationship between images.

2.2. Graph Convolution Network

Due to the strong power of modelling relations, graph convolution networks (GCNs) [4, 11, 12, 15] have been successfully applied to many computer vision tasks, including skeleton-based action recognition [44], video classification [41], and multi-label image recognition [6]. Many works [43, 46] also apply GCNs on person re-ID in the recent years. Specifically, Yan et al. [46] consider the relations among images by building the graph model on image samples. Wu et al. [43] use a graph neural network to learn the contextual interactions between the relevant regional features. However, these methods are mainly designed for feature learning and thus could not serve as an easy-to-use post processor.

3. Graph Convolution For Re-ranking

Different from previous re-ranking methods, which compute similarity with the sophisticated distance metric rather than Euclidean distance, we aim to improve representations while simply using Euclidean distance for retrieval. These features can be more flexible for downstream tasks.

The key challenge in the proposed method is to have the appropriate criterion for generating new features. Considering that the provided features are well trained, the changes in the features should be mild. Besides, features of the same person can be from different cameras and it is important to align features across multiple cameras to eliminate the bias from cameras. Therefore, we propose to propagate features over a graph with following criteria.

1. Given an image, only features from its k -nearest neighbors should be propagated.
2. Nearest neighbors from different cameras should be emphasized.

The first criterion implies a sparse graph which tries to mitigate the noisy features by taking their neighbors into account. The second criterion is to align features from different cameras, which is rarely investigated and important for reducing the gap between training and testing data. In the following sections, we will illustrate the details of our

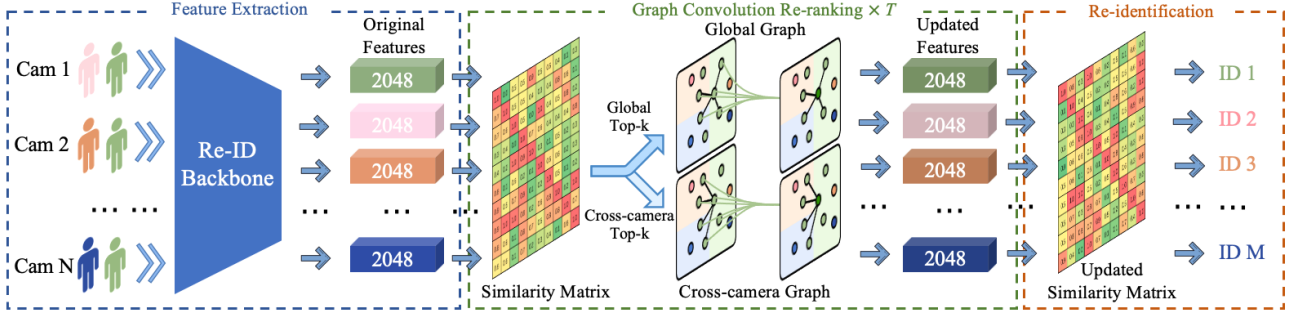


Figure 2: The pipeline of the proposed graph convolution based re-ranking (GCR) method.

graph convolution based re-ranking (GCR) method, especially how to build graphs with these two criteria.

3.1. Feature Propagation on Graph

Inspired by the graph convolution networks (GCN) [15], we try to propagate features with a convolution operator. The convolution operator in a standard GCN layer [15] can be written as

$$\tilde{X} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} X W \quad (1)$$

where $X \in \mathbb{R}^{n \times d}$ is input features, n is the number of features and d is the feature dimensionality. A is the $n \times n$ similarity matrix and D is the degree matrix that is a diagonal matrix with $D_{i,i} = \sum_j A_{i,j}$. $W \in \mathbb{R}^{d \times d'}$ denotes the parameters of convolution and \tilde{X} is the output after convolution. Therefore, the graph convolution can be considered as convolution on the graph characterized by the similarity matrix of A .

Without training in re-ranking, we adopt W as the identity matrix and $d' = d$. Hence, the graph convolution operator for re-ranking can be simplified as

$$\tilde{X} = D^{-\frac{1}{2}} A D^{-\frac{1}{2}} X \quad (2)$$

where X contains image features from both the query set and gallery set. Explicitly, the operator is to propagate features over a given graph and the key is to design an appropriate similarity matrix A .

3.1.1 Non-symmetric k -Nearest Neighbor Graph

Considering the first proposed criterion, we apply a global graph with k -nearest neighbor to keep features stable after re-ranking as follows.

1. For the i -th image, obtain its k -nearest neighbors \mathcal{N}_i^k with the original features.

2. For the i -th row of A , we compute the similarity as

$$A_{i,j} = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 / \gamma) & j \in \mathcal{N}_i^k \\ 1 & j = i \\ 0 & o.w. \end{cases} \quad (3)$$

where $\mathbf{x}_i \in \mathbb{R}^d$ is the input features for the i -th image and γ is the temperature parameter. The similarity matrix obtained from Eq. 3 is denoted as A_{nonsym}^{global} since it is a non-symmetric matrix. The degree matrix can be computed as

$$D_{row:global}(i, i) = \sum_j A_{i,j} \quad D_{col:global}(j, j) = \sum_i A_{i,j} \quad (4)$$

and the propagation criterion becomes

$$\tilde{X} = D_{row:global}^{-\frac{1}{2}} A_{nonsym}^{global} D_{col:global}^{-\frac{1}{2}} X \quad (5)$$

The method in Eq. 5 considers the nearest neighbors from all cameras.

3.1.2 Non-symmetric k -Nearest Neighbor Cross-camera Graph

To make sure that there are samples from different cameras for propagation, which is suggested in the second criterion, we introduce an cross-camera graph with k -nearest neighbors from different cameras as follows.

1. For the i -th image, obtain its k -nearest neighbors $\mathcal{N}_i^{diff:k}$ from different cameras with the original features.
2. For the i -th row of A , we compute the similarity as

$$A_{i,j} = \begin{cases} \exp(-\|\mathbf{x}_i - \mathbf{x}_j\|_2^2 / \gamma) & j \in \mathcal{N}_i^{diff:k} \\ 1 & j = i \\ 0 & o.w. \end{cases} \quad (6)$$

We denote the resulting similarity matrix as A_{nonsym}^{cross} , which is the similarity matrix across different cameras. Note that we include the i -th image itself in the similarity graph to calibrate the feature after propagation and make it comparable to the one from the global propagation.

Propagation with the cross-camera graph emphasizes the relationship between the image and its k -nearest neighbors from different cameras. It helps to eliminate the bias from cameras in the similarity matrix and align features across multiple cameras. With two obtained similarity matrices, we have our final propagation criterion as

$$\begin{aligned} \tilde{X} = & \alpha D_{row:global}^{-\frac{1}{2}} A_{nonsym}^{global} D_{col:global}^{-\frac{1}{2}} X + \\ & (1 - \alpha) D_{row:cross}^{-\frac{1}{2}} A_{nonsym}^{cross} D_{col:cross}^{-\frac{1}{2}} X \end{aligned} \quad (7)$$

where α is the parameter to balance the weights between two propagation procedures. Note that the parameter k can be different when generating these two similarity matrix, we denote them as k_g and k_c , respectively. Finally, the obtained features can be iteratively updated with the same criterion in Eq. 7 as

$$\begin{aligned} X_{t+1} = & \alpha D_{row:global}^{-\frac{1}{2}} A_{nonsym}^{global} D_{col:global}^{-\frac{1}{2}} X_t + \\ & (1 - \alpha) D_{row:cross}^{-\frac{1}{2}} A_{nonsym}^{cross} D_{col:cross}^{-\frac{1}{2}} X_t \end{aligned} \quad (8)$$

where t indicates the iteration index, from 1 to T . T is the total number of iterations and $X_1 = X$. The similarity matrices A_{nonsym}^{global} and A_{nonsym}^{cross} change during iterations. The whole pipeline is shown in Fig. 2.

3.2. Different Graphs

In the above subsection, we investigate the proposed two graphs used for propagation under our criteria. In this subsection, we introduce more graphs for potential applications.

3.2.1 Symmetric k -Nearest Neighbor Graph

First, the standard k -NN graph is non-symmetric, which can be converted to a symmetric one as follows.

$$A_{sym} = (A_{nonsym} + A_{nonsym}^\top) / 2 \quad (9)$$

For the symmetric matrix, the row degree matrix and the column degree matrix are identical and the propagation criterion becomes

$$\begin{aligned} X_{t+1} = & \alpha D_{row:global}^{-\frac{1}{2}} A_{sym}^{global} D_{col:global}^{-\frac{1}{2}} X_t + \\ & (1 - \alpha) D_{row:cross}^{-\frac{1}{2}} A_{sym}^{cross} D_{col:cross}^{-\frac{1}{2}} X_t \end{aligned} \quad (10)$$

3.2.2 Local k -Nearest Neighbor Graph

Besides non-symmetry, the conventional k -NN graph is constructed with all images, which is inefficient to update features on a large-scale data set. Consequently, we can have a local similarity matrix with size of $k \times k$ for parallel propagation. The local k -NN graph can be obtained as follows.

1. For the i -th image, obtain its k -nearest neighbors \mathcal{N}_i^k with the provided features.
2. Generate the symmetric $k \times k$ local similarity matrix A_{local}^i for the i -th image as
$$\forall u, v \in \mathcal{N}_i^k \cup \{i\}, A_{u,v}^i = \exp(-\|\mathbf{x}_u^i - \mathbf{x}_v^i\|_2^2 / \gamma)$$

We can obtain the local similarity matrices $A_{cross'}^i$ from different cameras with the similar procedure. After obtaining two compact similarity matrices, we can propagate features for each image as

$$\begin{aligned} \mathbf{x}_{t+1}^i = & \alpha D_{row:local}^{i, \frac{1}{2}} A_{local}^i D_{col:local}^{i, -\frac{1}{2}} X_t^i + \\ & (1 - \alpha) D_{row:cross'}^{i, \frac{1}{2}} A_{cross'}^i D_{col:cross'}^{i, -\frac{1}{2}} X_t^i \end{aligned} \quad (11)$$

where $X_t^i \in \mathbb{R}^{k \times d}$ contains images in \mathcal{N}_i^k .

Compared with the conventional k -NN graph, local k -NN graph only propagates the information from nearest neighbors and can be more efficient for large-scale applications.

4. Profile Vector Generation for Video Re-ID

Besides re-ranking for images, its application for video re-ID attracted much attention recently. Compared with image-based re-ID, each sample in video re-ID is a tracklet, which consists of a set of images rather than a single image. It's important to take full advantage of these multiple images in the tracklet to build a robust feature vector of this tracklet. Therefore, we propose a profile vector generation (PVG) method to extract a profile vector for each tracklet. And then our GCR method from image-level re-ID task can be extended to be applied in the video re-ID task.

Given images $\{\mathbf{x}_i, y_i\}$ from tracklets in the z -th camera, a simple profile for the c -th tracklet can be the mean vector

$$\hat{\mathbf{x}}_c = \frac{1}{n_z^c} \sum_{y_i=c} \mathbf{x}_i \quad (12)$$

where y_i indicates the id of the tracklet and n_z^c is the number of images in the c -th tracklet from the z -th camera.

In this paper, we expect the new profile vector $\hat{\mathbf{x}}_c$ of the c -th tracklet should be near to the features of images in the

the c -th tracklet, and meanwhile far away from the other features in the same camera. Hence, a ridge regression is involved to achieve this constraint. For each $\hat{\mathbf{x}}_c$, the optimization problem becomes

$$\min_{\hat{\mathbf{x}}_c} \frac{1}{n_z} \sum_{i=1}^n (\mathbf{x}_i^\top \hat{\mathbf{x}}_c - z_i^c)^2 + \frac{\lambda_p}{2} \|\hat{\mathbf{x}}_c\|_2^2 \quad (13)$$

where n_z is the total number of images in the z -th camera, and the z_i^c is the binary label whether the feature \mathbf{x}_i comes from the c -th tracklet. The $\|\hat{\mathbf{x}}_c\|_2$ is a regularization term. The challenge is how to design an appropriate label z_i^c . In a standard classification problem, z_i^c can be a binary variable as $z_i^c \in \{1, -1\}$ to indicate if the image is from the c -th tracklet. However, these labels imply a large margin between positive and negative images, which can be inapplicable for the linear projection defined by the profile vector. On the contrary, we consider a data-dependent margin as

$$z_i^c = \begin{cases} \frac{1}{n_z^c} - \frac{1}{n_z} & y_i = c \\ -\frac{1}{n_z} & o.w. \end{cases} \quad (14)$$

It is obvious that the margin between positive images and negative ones is $1/n_z^c$. The more frames the tracklet contains, the harder to split positive images from negative ones. As a result, in Eq. 14, the larger tracklet would holds a smaller margin. Moreover, when n_z is large, the label for negative images is close to 0, which is sufficient to identify irrelevant images in the high-dimensional space.

With the proposed labels, for each tracklet, the profile vector can be calculated with the closed-form solution as

$$\hat{\mathbf{x}}_c = \text{norm}((X_z^\top X_z + n_z \lambda_p I)^{-1} (\frac{1}{n_z^c} \sum_{i: y_i=c} x_i - \frac{1}{n_z} \sum_{i=1}^{n_z} x_i)) \quad (15)$$

where I is the identity matrix and X_z consists of all images from the z -th camera. $\text{norm}(\cdot)$ is a l2-norm operator. Compared with the mean vector in Eq. 12, the profile in Eq. 15 eliminates the mean vector $\frac{1}{n_z} \sum_{i=1}^{n_z} x_i$ of images from the same camera to reduce the bias from different cameras and leverages the geometric information from the covariance matrix $X_z^\top X_z$.

Although designed for video-based re-ID, the profile vector is also available for image-based re-ID, where each image could be viewed as a tracklet with only one frame. After obtaining the profile vector, the proposed GCR method can be further applied on re-ranking video tracklets. The whole process is called Graph Convolution Re-ranking for Video (GCRV) in our experiments. The whole method is summarized in Alg. 1.

Algorithm 1 Graph Convolution Re-ranking in Video (GCRV)

Input: features X , global neighbour k_g , cross neighbour k_c
 For each tracklet, generate the profile vector as in Eq. 15
for $t = 1$ **to** T **do**
 Generate similarity matrices A_{global} and A_{cross} from X
 Build graphs as in Section 3.1 and 3.2.
 Update features with the corresponding criterion as in Eq. 8 10 11
 Normalize new features to unit length (optional)
end for

5. Experiments

5.1. Datasets

In our experiments, we evaluate the proposed GCR on both image-based including Market-1501 [52] and Duke-MTMC-re-ID (Duke) [27], and video-based re-ID data sets, e.g. MARS [51].

Market-1501 [52] is a widely-used benchmark for person re-id with 1,501 identities from 6 cameras in total 750 identities (12,936 images) are used for training, 751 identities (19,732 images) are used for testing. Although high accuracy has been achieved in recent years, we demonstrate that the proposed re-ranking method can further boost the performance.

Duke-MTMC-re-ID (Duke) [27] dataset consists of 1,812 people from 8 cameras. Training and test sets both consist of 702 persons. The training set includes 16,522 images, the gallery 17,661 images, and the query set 2,228 image. Person bounding boxes in this dataset are manually annotated.

MARS [51] is used as a large-scale video-based person re-ID datasets in our experiments. It consists of 17,503 tracks and 1,261 identities. Each track has 59 frames on average. Similar to image-based Market-1501, MARS includes 3,248 distractor tracks to make the person re-ID more challenging.

5.2. Implementation Details

We extract re-ID features with models from strong baseline [18, 22, 23]. [22] provides public pre-trained models for Market and Duke while [18] provides models for MARS. All these baseline models achieve competitive performance and the dimension of features is 2,048. As current re-ranking methods always provide their results on different baselines and thus hard to compare with each other, we use an open-source state-of-the-art baseline and hope other researchers could be convenient to compare results with ours under the same baseline in future.

Method	Reference	Market		Duke		MARS	
		Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
PSE+ KR [28]	CVPR18	90.2	83.5	84.4	78.9	74.9	70.7
PSE+ ECN [28]	CVPR18	90.3	84.0	85.2	79.8	76.7	71.8
RED [3]	CVPR19	94.7	91.0	-	-	-	-
UED [2]	CVPR19	95.9	92.8	-	-	-	-
SFT+KR [21]	ICCV19	93.5	90.6	88.3	83.3	-	-
SFT+LBR [21]	ICCV19	94.1	87.5	90.0	79.6	-	-
ISP [56]	ECCV20	95.3	88.6	89.6	80.0	-	-
MPN [9]	TPAMI20	96.3	89.4	91.5	82.0	-	-
CircleLoss [36]	CVPR20	96.1	87.4	-	-	-	-
AGRL [43]	TIP20	-	-	-	-	89.5	81.9
VKD [25]	ECCV20	-	-	-	-	89.4	83.1
MGH [45]	CVPR20	-	-	-	-	90.0	85.8
BoT	CVPRW19	94.5	85.9	86.5	76.4	85.8	79.7
BoT+KR [54]	CVPR17	95.4	94.2	90.2	89.1	84.3	85.2
BoT+ECN [28]	CVPR18	95.8	93.2	90.9	87.0	88.1	85.6
BoT+LBR [21]	ICCV19	95.7	91.3	89.7	84.4	87.4	82.5
BoT+GCRV	-	96.1	94.7	91.5	89.7	89.0	87.0
BoT+GCRV+KR	-	96.8	94.9	92.0	89.9	89.7	87.1
SOTA features	CVPR20	96.3	89.4	91.5	82.0	90.0	85.8
SOTA+KR	CVPR17	95.6	94.5	90.5	89.6	88.8	90.7
SOTA+ECN	CVPR18	95.1	94.0	90.8	88.3	92.7	90.5
SOTA+GCRV	-	96.6	95.1	92.9	91.3	93.8	92.8

Table 1: Comparison with state-of-the-art methods on Market-1501, Duke and MARS. The **bold** indicates the best performance.

		T=1		T=2		T=3		T=4		T=5	
		Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
$k=5$	$\gamma=0.1$	95.0	88.1	95.3	89.7	95.5	90.9	95.4	90.7	95.3	90.5
	$\gamma=0.2$	95.4	91.0	95.7	93.0	95.6	93.5	95.4	93.4	95.5	93.5
	$\gamma=0.3$	95.5	92.0	95.3	93.4	95.0	93.4	94.9	93.2	94.7	93.1
$k=15$	$\gamma=0.1$	95.0	89.0	95.6	91.3	95.9	92.7	95.8	92.6	95.6	92.6
	$\gamma=0.2$	95.5	92.6	95.6	94.3	95.9	94.6	95.7	94.5	95.5	94.4
	$\gamma=0.3$	95.6	93.3	95.6	94.3	95.2	93.9	95.1	93.6	95.0	93.6
$k=50$	$\gamma=0.1$	95.0	88.8	95.3	91.0	95.3	92.3	95.1	92.2	95.0	92.1
	$\gamma=0.2$	95.2	91.4	94.5	91.8	92.4	89.2	90.2	88.1	88.0	87.3
	$\gamma=0.3$	94.8	91.1	92.6	88.3	88.5	81.9	85.5	74.9	79.5	69.5

Table 2: The performance comparison (%) of GCR parameters on the Market-1501. The **bold** indicates the best performance.

5.3. Comparison with State-of-the-Art Methods

Table 1 compares the proposed method to state-of-the-art re-ranking methods. To make a fair comparison, we reproduce the results of the most commonly used re-ranking methods, including k-reciprocal (KR) [54], ECN [28] and LBR [21] and compare with them under the same baseline, *i.e.* BoT [22]. It can be found that our *BoT + GCRV* achieves a much better performance than the origin base-

line *BoT* by a large margin. Meanwhile the performance of *BoT + GCRV* is also higher than that of other re-ranking methods in both Rank-1 and mAP, especially on the video re-ID dataset MARS, which thanks to the proposed PVG method. Moreover, when we integrate our *GCRV* with *KR*, the result *BoT + GCRV + KR* is further improved, higher than *BoT + KR*. This phenomenon implies that *GCRV* can be used to complement *KR*, because our *GCRV* involves the cross-camera information in the graph

that is ignored in KR .

We also list current state-of-the-art re-ID methods in Table 1. Some of them involve post-processing in their methods, such as ISP [56] for image-level re-ID and MGH [45] for video-level re-ID. And some of them import off-the-shelf re-ranking methods, which is marked as +. We can see that most of these state-of-the-art methods with different re-ranking methods are lower than our $BoT+GCRV$ result, even our baseline BoT is lower than the state of the arts. If we use the same SOTA features, the proposed method outperforms KR and ECN by a large margin. Also $GCRV$ method could be further improved if combined with KR . It is worth noticing that after re-ranking with our $GCRV$, the feature is still in the Euclidean space which can be easily used in downstream tasks and available for fast retrieval of massive data in real applications.

5.4. Ablation Study

5.4.1 Effect of Parameters in Graphs

There are three parameters in the non-symmetric k -NN graph, *i.e.* the temperature parameter γ and the number of nearest neighbor k in Eq. 3, the number of iterations in Eq. 8. We conduct ablation study on Market-1501.

Table 2 summarizes the results with different parameters. First, we observe that k is important for the performance. When k is small, the information from nearest neighbor is limited and the features may not be aligned well. However, if k is too large, the additional noise from images with different labels will be introduced, which can result in the degraded performance. Hence, choosing an appropriate number of nearest neighbors is essential for our graph-based method. Then, it is evident that γ is related to the choice of k . When k is small, a large γ is preferred. And when k is large, a small γ can help to focus on more similar images. Finally, with multiple iterations, the performance of features can be further improved as illustrated. Since γ and T are not sensitive as shown in Table 2, we keep them as $\gamma = 0.2$ and $T = 3$ in all other experiments.

To investigate the changes between different iterations, we visualize the updated features using t-sne [24] in Fig. 3. It can be seen that the features from the same person become more aggregated with the propagation of features. With more iterations $T > 3$, the samples in light blue are over-clustered into three points in (c). Such over-clustering might cause matching mistakes. Similar results can be found in Tab. 2, where more iterations ($T > 3$) no longer improves the accuracy.

5.4.2 Effect of Global and Cross-camera Graphs

We evaluate the effectiveness of the two graphs pro-

Setting	Market		Duke		MARS	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
global only	95.9	94.6	90.1	89.4	84.9	85.2
cross only	96.2	93.1	91.6	86.8	89.8	84.7
global+cross	96.0	94.7	91.5	89.7	89.0	87.0

Table 3: The performance comparison (%) between GCR global graph and cross-camera graph on Market-1501.

Setting	Market	
	Rank-1	mAP
non-sym	96.0	94.7
sym	96.4	94.6
local- k	96.0	94.3

Table 4: The performance comparison (%) of GCR with different graphs on the Market-1501.

posed in Section 3.1, *i.e.* non-symmetric k -NN graph and non-symmetric k -NN cross-camera graph, denoted as *globalonly* and *crossonly* respectively in Table 3. It is obvious that each of graph can propagate appropriate features and achieve the applicable performance on Market-1501. By combining the two graphs, we can achieve the ideal performance that enjoys the good mAP from the global graph and the better Rank-1 with the help of the cross-camera graph. Note that here we set $k_g = 15$ and $k_c = 3$ in the experiments for all re-ID datasets.

The trade-off hyper-parameter between two graphs is fixed as $\alpha = 0.7$. We plot accuracy curves with respect to the different α in Fig. 4. Rank-1 saturates for $\alpha < 0.7$ while mAP reaches the peak at $\alpha = 0.7$. Since mAP is often more important for retrieval cases, we select the hyper-parameter for the sake of better mAP.

5.4.3 Effect of Different Graphs

There can be different graphs adopted in our method and we evaluate graphs discussed in Sec. 3.2. Specifically, 3 graphs are compared in the experiments as follows.

- **Non-symmetric formulation (non-sym)** The non-symmetric formulation in Eq. 8, which serves as the default graph.
- **Symmetric formulation (sym)**: The symmetric formulation as in Eq. 10.
- **Local k -nearest neighbor (local- k)**: local k -nearest neighbor graph as described in Eq. 11.

Table 4 demonstrates the performance of different graphs. First, it can be concluded that *non-sym* and *sym* share the similar performance while *sym* is slightly better than *non-sym*. It shows that a symmetric graph can be

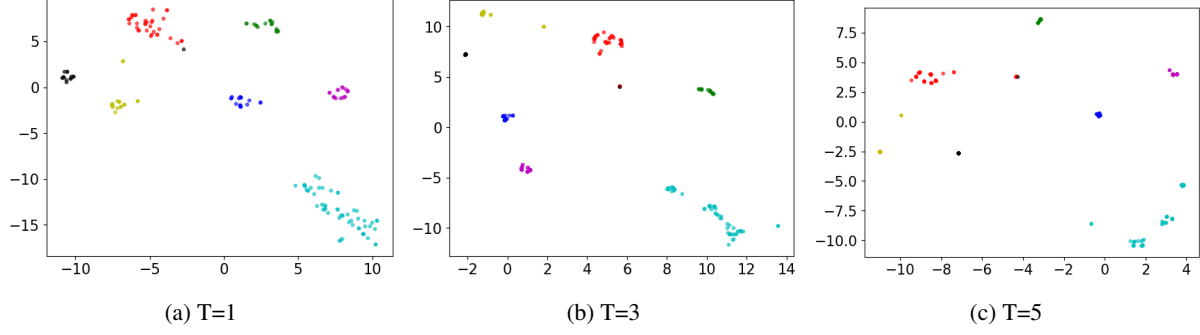


Figure 3: Feature distribution for GCR with different iterations on MARS.

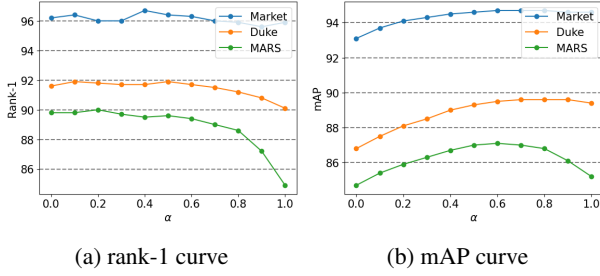


Figure 4: The performance curve under different α .

Setting	MARS		Market	
	Rank-1	mAP	Rank-1	mAP
Mean	85.8	79.7	94.5	85.9
PVG ($\lambda_p=0.01$)	82.9	75.6	94.7	84.6
PVG ($\lambda_p=0.1$)	85.9	79.3	94.7	86.3
PVG ($\lambda_p=1.0$)	88.1	80.3	94.6	86.3
PVG ($\lambda_p=10.0$)	88.6	80.6	94.6	86.3

Table 5: The performance comparison (%) of the parameter λ_p in profile vector generation (PVG).

Method	Market		Duke		MARS	
	Rank-1	mAP	Rank-1	mAP	Rank-1	mAP
baseline	94.5	85.9	86.5	76.4	85.8	79.7
+GCR	96.0	94.7	91.5	89.6	86.6	85.3
+PVG	94.6	86.3	86.9	76.0	88.6	80.6
+GCRV	96.1	94.7	91.6	89.2	89.0	87.0

Table 6: Comparison of GCR, PVG and GCRV on Market-1501, Duke and MARS.

more appropriate for propagation. Then, the local k -NN graph performs 0.3% worse than *sym* on mAP. Considering the efficiency from the parallel implementation, local k -NN graph is still potential useful for large-scale data set.

5.4.4 Effect of Profile Vector Generation (PVG)

When applying for video re-ID, we will first generate profile vectors for tracklets. We compare the proposed profile vector generation (PVG) method to the baseline of mean vectors as in Eq. 12. The only parameter in PVG is the weight of the regularizer λ_p . We vary it from 0.01 to 10 and evaluate it on the video data set MARS and the image data set Market.

From Table 5, we can find that when setting the regularizer appropriately, the performance on video re-ID can be significantly improved. On the video data set, PVG with $\lambda_p = 10$ outperforms mean vector by 0.9% on mAP and about 3% on Rank-1. Considering our strategy of generating profile vectors is extremely simple with the closed-form solution, it can improve the performance nearly without any additional cost. Even on the image data set, where mean vector becomes the original image vector, the mAP can still gain from the proposed PVG. It demonstrates the effectiveness of the proposed PVG method. We will keep $\lambda_p = 10$ in the rest experiments.

Then, we incorporate PVG to GCR and compare the performance of GCR and GCRV in Table 6. It is not surprising to observe that GCR achieves dramatic improvement on different data sets compared to the baseline. It is because re-ranking can effectively mitigate the challenge from different cameras. On the image-based re-ID, GCRV achieves similar result with GCR. But on the video-based re-ID dataset MARS, GCRV demonstrates a better performance than GCR. It confirms that GCRV is more appropriate for the video-based re-ID.

5.5. Visualization

To better study different feature propagation methods, we visualize the propagated features in Fig. 5. Naive query expansion [7] treats each neighbour equally and thus produces unsatisfying results. LBR [21] performs local blurring on the gallery features while keeping query features

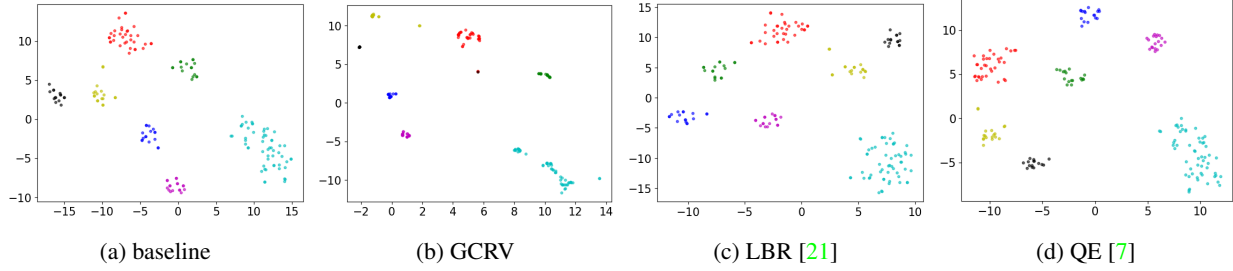


Figure 5: Visualization of different re-ranking methods on MARS.

Method	KR [54]	ECN [28]	proposed
Time(s)	76	72	24

Table 7: The computation time of re-ranking methods on Market-1501.

unchanged. Our proposed method relying on carefully designed graphs, better propagation strategy and profile vector generation produces the best visualization results.

5.6. Efficiency

In real-world applications where the gallery size is very large, traditional re-ranking methods such as k-reciprocal may suffer from heavy computation. Table 7 lists the computation time of different re-ranking methods on the same Market-1501 dataset with the same hardware settings of 24 cores Platinum 8163 CPU. The similarity matrix size is 3368 queries * 15913 galleries, and our time complexity is $\mathcal{O}(N^2 \log N)$. As can be seen, K-reciprocal (KR) and ECN suffer from low computation speed due to complex set operations. On the other hand, the proposed method relies only on simple matrix operations and achieves better efficiency.

6. Conclusion

In this paper we propose a graph convolution based re-ranking method for person re-ID. Unlike previous methods, we propose to learn features with propagation over graphs and re-compute similarity with the standard Euclidean distance. By investigating the criteria for propagation, we develop different similarity graphs and propagate features from both graphs for a single image. Empirical study with strong baseline verifies the effectiveness of the proposed method.

In our method, the convolution parameter of W is set to be an identity matrix. With a small set of labeled images from the target domain, we can improve the re-ranking method with a learnable W . Applying our method for semi-supervised re-ranking can be our future work.

References

- [1] Song Bai, Xiang Bai, and Qi Tian. Scalable person re-identification on supervised smoothed manifold. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2530–2539, Honolulu, HI, USA, 2017. IEEE. [2](#), [3](#)
- [2] Song Bai, Peng Tang, Philip HS Torr, and Longin Jan Latecki. Re-ranking via metric fusion for object retrieval and person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 740–749, Long Beach, CA, USA, 2019. IEEE. [1](#), [2](#), [3](#), [7](#)
- [3] Song Bai, Zhichao Zhou, Jingdong Wang, Xiang Bai, Longin Jan Latecki, and Qi Tian. Ensemble diffusion for retrieval. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 774–783, Venice, Italy, 2017. IEEE. [2](#), [3](#), [7](#)
- [4] Jie Chen, Tengfei Ma, and Cao Xiao. Fastgcn: fast learning with graph convolutional networks via importance sampling. In *International Conference on Learning Representations, ICLR*, pages 1–1, Vancouver, BC, Canada, 2018. OpenReview.net. [3](#)
- [5] Weihua Chen, Xiaotang Chen, Jianguo Zhang, and Kaiqi Huang. A multi-task deep network for person re-identification. In *Proceedings of the Thirty-First AAAI Conference on Artificial Intelligence*, pages 3988–3994, San Francisco, California, USA, 2017. AAAI. [2](#)
- [6] Zhao-Min Chen, Xiu-Shen Wei, Peng Wang, and Yanwen Guo. Multi-label image recognition with graph convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5177–5186, Long Beach, CA, USA, 2019. IEEE. [3](#)
- [7] Ondrej Chum, James Philbin, Josef Sivic, Michael Isard, and Andrew Zisserman. Total recall: Automatic query expansion with a generative feature model for object retrieval. In *2007 IEEE 11th International Conference on Computer Vision*, pages 1–8, Rio de Janeiro, Brazil, 2007. IEEE. [9](#), [10](#)
- [8] Jiankang Deng, Jia Guo, Niannan Xue, and Stefanos Zafeiriou. Arcface: Additive angular margin loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 4690–4699, Long Beach, CA, USA, 2019. IEEE. [2](#)
- [9] Changxing Ding, Kan Wang, Pengfei Wang, and Dacheng Tao. Multi-task learning with coarse priors for robust part-aware person re-identification. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, Early(Access):1–1, 2020. [7](#)
- [10] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality reduction by learning an invariant mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR’06)*, volume 2, pages 1735–1742, New York, NY, USA, 2006. IEEE. [2](#)
- [11] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In *Advances in neural information processing systems*, pages 1024–1034, Long Beach, CA, USA, 2017. Curran Associates, Inc. [3](#)
- [12] Wenbing Huang, Tong Zhang, Yu Rong, and Junzhou Huang. Adaptive sampling towards fast graph representation learning. In *Advances in neural information processing systems*, pages 4558–4567, Montréal, Canada, 2018. Curran Associates, Inc. [3](#)
- [13] Raymond Austin Jarvis and Edward A. Patrick. Clustering using a similarity measure based on shared near neighbors. *IEEE Trans. Computers*, 22(11):1025–1034, 1973. [2](#)
- [14] Herve Jegou, Hedi Harzallah, and Cordelia Schmid. A contextual dissimilarity measure for accurate and efficient image search. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8, Minneapolis, MN, USA, 2007. IEEE. [3](#)
- [15] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. In *International Conference on Learning Representations*, pages 1–1, Toulon, France, 2016. OpenReview.net. [2](#), [3](#), [4](#)
- [16] Dangwei Li, Xiaotang Chen, Zhang Zhang, and Kaiqi Huang. Learning deep context-aware features over body and latent parts for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 384–393, Honolulu, HI, USA, 2017. IEEE. [2](#)
- [17] Yutian Lin, Liang Zheng, Zhedong Zheng, Yu Wu, Zhilan Hu, Chenggang Yan, and Yi Yang. Improving person re-identification by attribute and identity learning. *Pattern Recognition*, 95:151–161, 2019. [3](#)
- [18] Chih-Ting Liu, Chih-Wei Wu, Yu-Chiang Frank Wang, and Shao-Yi Chien. Spatially and temporally efficient non-local attention network for video-based person re-identification. In *British Machine Vision Conference*, page 243, Cardiff, UK, 2019. BMVA Press. [6](#)
- [19] Weiyang Liu, Yandong Wen, Zhiding Yu, Ming Li, Bhiksha Raj, and Le Song. Sphreface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 212–220, Honolulu, HI, USA, 2017. IEEE. [2](#)
- [20] Yuanliu Liu, Bo Peng, Peipei Shi, He Yan, Yong Zhou, Bing Han, Yi Zheng, Chao Lin, Jianbin Jiang, Yin Fan, et al. iqi-vid: A large dataset for multi-modal person identification. *arXiv preprint arXiv:1811.07548*, abs/1811.07548:1–1, 2018. [3](#)
- [21] Chuanchen Luo, Yuntao Chen, Naiyan Wang, and Zhaoxiang Zhang. Spectral feature transformation for person re-identification. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 4976–4985, Seoul, Korea, 2019. IEEE. [2](#), [3](#), [7](#), [9](#), [10](#)
- [22] Hao Luo, Youzhi Gu, Xingyu Liao, Shenqi Lai, and Wei Jiang. Bag of tricks and a strong baseline for deep person re-identification. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, Long Beach, CA, USA, 2019. IEEE. [2](#), [6](#), [7](#)
- [23] Hao Luo, Wei Jiang, Youzhi Gu, Fuxu Liu, Xingyu Liao, Shenqi Lai, and Jianyang Gu. A strong baseline and batch normalization neck for deep person re-identification. *IEEE Transactions on Multimedia*, 22(10):2597–2609, 2019. [2](#), [6](#)
- [24] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008. [8](#)

- [25] Angelo Porrello, Luca Bergamini, and Simone Calderara. Robust re-identification by multiple views knowledge distillation. In *European Conference on Computer Vision*, pages 93–110, Glasgow, UK, 2020. Springer. 7
- [26] Danfeng Qin, Stephan Gammeter, Lukas Bossard, Till Quack, and Luc Van Gool. Hello neighbor: Accurate object retrieval with k-reciprocal nearest neighbors. In *CVPR 2011*, pages 777–784, Colorado Springs, CO, USA, 2011. IEEE. 3
- [27] Ergys Ristani, Francesco Solera, Roger Zou, Rita Cucchiara, and Carlo Tomasi. Performance measures and a data set for multi-target, multi-camera tracking. In *European Conference on Computer Vision*, pages 17–35, Amsterdam, The Netherlands, 2016. Springer. 6
- [28] M Saquib Sarfraz, Arne Schumann, Andreas Eberle, and Rainer Stiefelwagen. 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*, pages 420–429, Salt Lake City, UT, USA, 2018. IEEE. 1, 2, 3, 7, 10
- [29] Florian Schroff, Dmitry Kalenichenko, and James Philbin. Facenet: A unified embedding for face recognition and clustering. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 815–823, Boston, MA, USA, 2015. IEEE. 2
- [30] Yantao Shen, Hongsheng Li, Tong Xiao, Shuai Yi, Dapeng Chen, and Xiaogang Wang. Deep group-shuffling random walk for person re-identification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2265–2274, Salt Lake City, UT, USA, 2018. IEEE. 3
- [31] Yantao Shen, Hongsheng Li, Shuai Yi, Dapeng Chen, and Xiaogang Wang. Person re-identification with deep similarity-guided graph neural network. In *Proceedings of the European conference on computer vision (ECCV)*, pages 486–504, Munich, Germany, 2018. Springer. 3
- [32] Nitish Srivastava, Geoffrey Hinton, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov. Dropout: a simple way to prevent neural networks from overfitting. *The journal of machine learning research*, 15(1):1929–1958, 2014. 2
- [33] Chi Su, Fan Yang, Shiliang Zhang, Qi Tian, Larry S Davis, and Wen Gao. Multi-task learning with low rank attribute embedding for person re-identification. In *Proceedings of the IEEE international conference on computer vision*, pages 3739–3747, Santiago, Chile, 2015. IEEE. 3
- [34] Chi Su, Shiliang Zhang, Junliang Xing, Wen Gao, and Qi Tian. Deep attributes driven multi-camera person re-identification. In *European conference on computer vision*, pages 475–491, Amsterdam, The Netherlands, 2016. Springer. 3
- [35] Yumin Suh, Jingdong Wang, Siyu Tang, Tao Mei, and Kyoung Mu Lee. Part-aligned bilinear representations for person re-identification. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 402–419, Munich, Germany, 2018. Springer. 2
- [36] Yifan Sun, Changmao Cheng, Yuhang Zhang, Chi Zhang, Liang Zheng, Zhongdao Wang, and Yichen Wei. Circle loss: A unified perspective of pair similarity optimization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6398–6407, Seattle, WA, USA, 2020. IEEE. 7
- [37] Yifan Sun, Liang Zheng, Weijian Deng, and Shengjin Wang. Svdnet for pedestrian retrieval. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3800–3808, Venice, Italy, 2017. IEEE. 2
- [38] Yifan Sun, Liang Zheng, Yi Yang, Qi Tian, and Shengjin Wang. Beyond part models: Person retrieval with refined part pooling (and a strong convolutional baseline). In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 480–496, Munich, Germany, 2018. Springer. 2
- [39] Guangcong Wang, Jianhuang Lai, Peigen Huang, and Xiaohua Xie. Spatial-temporal person re-identification. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 33, pages 8933–8940, Honolulu, Hawaii, USA, 2019. AAAI. 3
- [40] Hao Wang, Yitong Wang, Zheng Zhou, Xing Ji, Dihong Gong, Jingchao Zhou, Zhifeng Li, and Wei Liu. Cosface: Large margin cosine loss for deep face recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5265–5274, Salt Lake City, UT, USA, 2018. IEEE. 2
- [41] Xiaolong Wang and Abhinav Gupta. Videos as space-time region graphs. In *Proceedings of the European conference on computer vision (ECCV)*, pages 399–417, Munich, Germany, 2018. Springer. 3
- [42] Yandong Wen, Kaipeng Zhang, Zhifeng Li, and Yu Qiao. A discriminative feature learning approach for deep face recognition. In *European conference on computer vision*, pages 499–515, Amsterdam, The Netherlands, 2016. Springer. 2
- [43] Yiming Wu, Omar El Farouk Bourahla, Xi Li, Fei Wu, Qi Tian, and Xue Zhou. Adaptive graph representation learning for video person re-identification. *IEEE Transactions on Image Processing*, 29:8821–8830, 2020. 3, 7
- [44] Sijie Yan, Yuanjun Xiong, and Dahua Lin. Spatial temporal graph convolutional networks for skeleton-based action recognition. In *Conference on Artificial Intelligence*, pages 7444–7452, New Orleans, Louisiana, USA, 2018. AAAI Press. 3
- [45] Yichao Yan, Jie Qin, Jiabin Chen, Li Liu, Fan Zhu, Ying Tai, and Ling Shao. Learning multi-granular hypergraphs for video-based person re-identification. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2899–2908, Seattle, WA, USA, 2020. IEEE. 7, 8
- [46] Yichao Yan, Qiang Zhang, Bingbing Ni, Wendong Zhang, Minghao Xu, and Xiaokang Yang. Learning context graph for person search. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2158–2167, Long Beach, CA, USA, 2019. IEEE. 3
- [47] Rui Yu, Zhichao Zhou, Song Bai, and Xiang Bai. Divide and fuse: A re-ranking approach for person re-identification. In *British Machine Vision Conference*, pages 135.1–135.13, London, UK, 2017. BMVA Press. 2, 3

- [48] Xuan Zhang, Hao Luo, Xing Fan, Weilai Xiang, Yixiao Sun, Qiqi Xiao, Wei Jiang, Chi Zhang, and Jian Sun. Alignedreid: Surpassing human-level performance in person re-identification. *arXiv preprint arXiv:1711.08184*, arXiv(preprint):1–1, 2017. [2](#)
- [49] Yuqi Zhang, Yongzhen Huang, Liang Wang, and Shiqi Yu. A comprehensive study on gait biometrics using a joint cnn-based method. *Pattern Recognition*, 93:228–236, 2019. [3](#)
- [50] Yuqi Zhang, Yongzhen Huang, Shiqi Yu, and Liang Wang. Cross-view gait recognition by discriminative feature learning. *IEEE Transactions on Image Processing*, 29:1001–1015, 2019. [2](#)
- [51] Liang Zheng, Zhi Bie, Yifan Sun, Jingdong Wang, Chi Su, Shengjin Wang, and Qi Tian. Mars: A video benchmark for large-scale person re-identification. In *European Conference on Computer Vision*, pages 868–884, Amsterdam, The Netherlands, 2016. Springer. [6](#)
- [52] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian. Scalable person re-identification: A benchmark. In *Proceedings of the IEEE international conference on computer vision*, pages 1116–1124, Santiago, Chile, 2015. IEEE. [6](#)
- [53] Zhedong Zheng, Liang Zheng, and Yi Yang. A discriminatively learned cnn embedding for person reidentification. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, 14(1):1–20, 2017. [1](#), [2](#)
- [54] Zhun Zhong, Liang Zheng, Donglin Cao, and Shaozi Li. Re-ranking person re-identification with k-reciprocal encoding. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1318–1327, Honolulu, HI, USA, 2017. IEEE. [1](#), [2](#), [3](#), [7](#), [10](#)
- [55] Zhun Zhong, Liang Zheng, Guoliang Kang, Shaozi Li, and Yi Yang. Random erasing data augmentation. In *AAAI*, pages 13001–13008, New York, NY, USA, 2020. AAAI. [2](#)
- [56] Kuan Zhu, Haiyun Guo, Zhiwei Liu, Ming Tang, and Jinqiao Wang. Identity-guided human semantic parsing for person re-identification. In *Proceedings of the European conference on computer vision (ECCV)*, pages 346–363, Glasgow, UK, 2020. Springer. [7](#), [8](#)