# GRAPH CONVOLUTION FOR RE-RANKING IN PERSON RE-IDENTIFICATION

*Yuqi Zhang*[⋆] *Qi Qian*[⋆§] *Chong Liu*[†‡] *Weihua Chen*[⋆] *Fan Wang*[⋆] *Hao Li*[⋆] *Rong Jin*[⋆]

[†] State Key Laboratory of Computer Science, Institute of Software, Chinese Academy of Sciences
[‡] University of Chinese Academy of Sciences, Beijing China
[⋆] Machine Intelligence Technology Lab, Alibaba Group

## ABSTRACT

Nowadays, deep learning is widely applied to extract features for similarity computation in person re-identification (re-ID). However, the difference between the training data and 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 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.

***Index Terms*—** Reranking, graph neural networks, person re-identification

## 1. INTRODUCTION

Person re-identification (re-ID) aims to retrieve images of the same person from the gallery set given a query image [1]. A standard pipeline is to extract features for 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 [2, 3, 4, 5]. 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 [6, 7, 1].

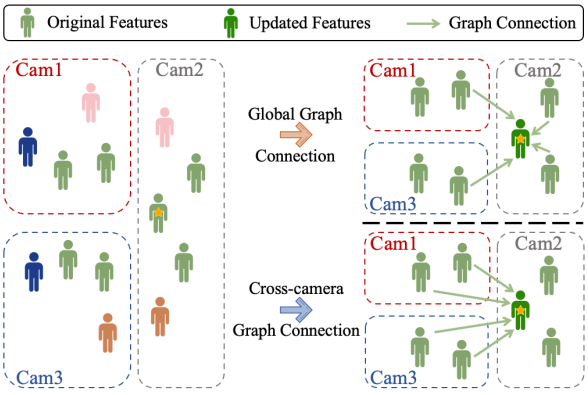



**Fig. 1**: Illustration of graphs with two proposed criteria. The person with the star denotes the target image and the arrows indicate its $k$-nearest neighbors. People with the same color hold the same ID. Corresponding to the two criteria, we generate two graphs (i.e., Global graph: connecting the $k$-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).

Given features from the deep model, re-ranking is to re-calculate the similarity of images by introducing other information and use sophisticated similarity metrics [8, 6, 9, 7, 10, 1] to rearrange the ranking list. Current SOTA methods $k$-reciprocal encoding [1] or ECN [7]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 [11] 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

§ Equal contribution
The work was done when Chong Liu was intern at Alibaba Group

features while Euclidean distance can still be directly used as the similarity measure. Inspired by graph convolution networks (GCN) [12], 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 [1, 7]. 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.

## 2. GRAPH CONVOLUTION FOR RE-RANKING

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

### 2.1. $k$-Nearest Cross-camera Graph

Considering the first proposed criterion, we propose a global graph first. To make sure that there are samples from different cameras for propagation, which is suggested in the second criterion, we also 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 $\mathcal{N}_i^{\text{diff}:k}$ from different cameras with the original features.
2. For the $i$-th row of $A$, we compute the similarity as

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

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

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

where $\alpha$ 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

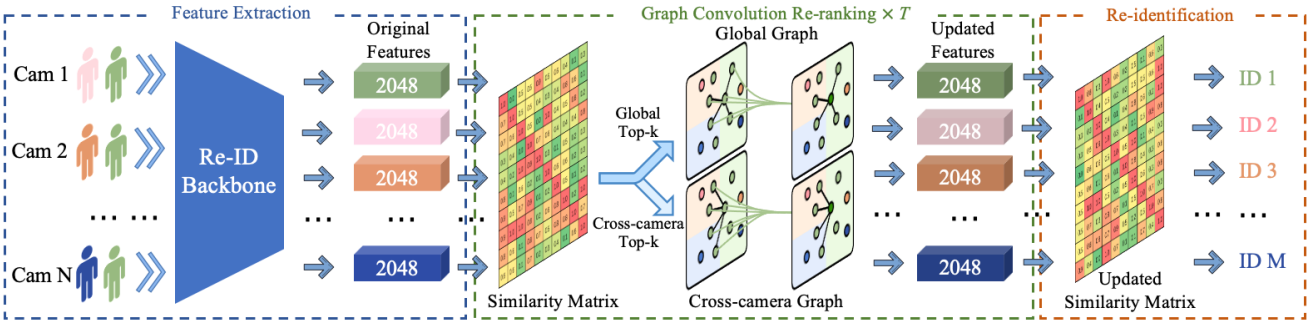



**Fig. 2**: The pipeline of the proposed graph convolution based re-ranking (GCR) method.

them as $k_g$ and $k_c$, respectively. Finally, the obtained features can be iteratively updated with the same criterion in Eq.2 as

$$\begin{aligned} X_{t+1} =& \alpha D_{row:global}^{-\frac{1}{2}} A_{nonsym}^{global} D_{col:global}^{-\frac{1}{2}} X_t + \\ & (1-\alpha) D_{row:cross}^{-\frac{1}{2}} A_{nonsym}^{cross} D_{col:cross}^{-\frac{1}{2}} X_t \end{aligned} \tag{3}$$

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. PROFILE VECTOR GENERATION FOR VIDEO RE-ID

Besides re-ranking for images, its application for video re-ID attracted much attention recently.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.

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

$$\min_{\hat{\mathbf{x}}_c} \frac{1}{n_z} \sum_{i=1}^{n} (\mathbf{x}_i^\top \hat{\mathbf{x}}_c - z_i^c)^2 + \frac{\lambda_p}{2} \|\hat{\mathbf{x}}_c\|_2^2 \tag{4}$$

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. For each tracklet, the profile vector can be calculated with the closed-form solution as

$$\hat{\mathbf{x}}_c = norm\big((X_z^\top X_z + n_z \lambda_p I)^{-1} (\frac{1}{n_z^c} \sum_{i:y_i=c} x_i - \frac{1}{n_z} \sum_{i=1}^{n_z} x_i)\big) \tag{5}$$

where $I$ is the identity matrix and $X_z$ consists of all images from the $z$-th camera. $norm(\cdot)$ is a l2-norm operator. Compared with the mean vector, the profile in Eq. 5 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.

## 4. EXPERIMENTS

### 4.1. Datasets

In our experiments, we evaluate the proposed GCR on both image-based including Market-1501 [16] and Duke-MTMC-re-ID (Duke) [17], and video-based re-ID data sets, *e.g.* MARS [18].

**Market-1501** [16] 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.

**Duke-MTMC-re-ID (Duke)** [17] dataset consists of $1,812$ people from 8 cameras. Training and test sets both consist of 702 persons.

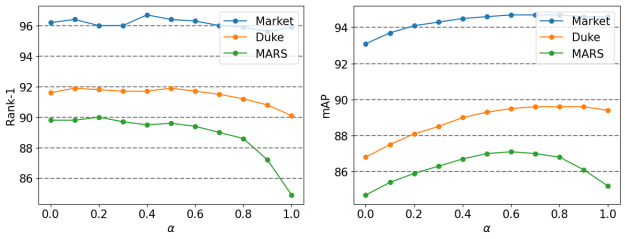
**MARS** [18] is used as a large-scale video-based person re-ID datasets in our experiments. It consists of $17,503$ tracks and $1,261$ identities.

### 4.2. 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 under the same features. The proposed method outperforms reranking methods KR, ECN and LBR by a large margin. 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.

| Method | Reference | Market | | Duke | | MARS | |
|---|---|---|---|---|---|---|---|
| | | Rank-1 | mAP | Rank-1 | mAP | Rank-1 | mAP |
| ISP [13] | ECCV20 | 95.3 | 88.6 | 89.6 | 80.0 | - | - |
| MPN [14] | TPAMI20 | 96.3 | 89.4 | 91.5 | 82.0 | - | - |
| MGH [15] | CVPR20 | - | - | - | - | 90.0 | 85.8 |
| SOTA features | CVPR20 | 96.3 | 89.4 | 91.5 | 82.0 | 90.0 | 85.8 |
| SOTA+KR [1] | CVPR17 | 95.6 | 94.5 | 90.5 | 89.6 | 88.8 | 90.7 |
| SOTA+ECN [7] | CVPR18 | 95.1 | 94.0 | 90.8 | 88.3 | 92.7 | 90.5 |
| SOTA+LBR [11] | ICCV19 | 95.0 | 92.3 | 89.7 | 85.8 | 91.4 | 87.5 |
| 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.




(a) rank-1 curve (b) mAP curve

**Fig. 3**: The performance curve under different $\alpha$.

| 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 2**: Comparison of GCR, PVG and GCRV on Market-1501, Duke and MARS.

### 4.3. Ablation Study

To make a fair comparison, we use BoT [19] features in the ablation study. The trade-off hyper-parameter between two graphs is fixed as $\alpha = 0.7$. We plot accuracy curves with respect to the different $\alpha$ in Fig. 3. 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.

Then, we incorporate PVG to GCR and compare the performance of GCR and GCRV in Table 2. 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.

| Method | KR [1] | ECN [7] | proposed |
|---|---|---|---|
| Time(s) | 76 | 72 | 24 |

**Table 3**: The computation time of re-ranking methods on Market-1501.

### 4.4. Efficiency

Table 3 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 $\mathcal{O}\left(N^2 \log N\right)$. 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.

## 5. 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.

## 6. REFERENCES


[1] 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*, Honolulu, HI, USA, 2017, pp. 1318–1327, IEEE.

[2] Zhedong Zheng, Liang Zheng, and Yi Yang, "A discriminatively learned cnn embedding for person reidentification," *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)*, vol. 14, no. 1, pp. 1–20, 2017.

[3] Yuqi Zhang, Yongzhen Huang, Liang Wang, and Shiqi Yu, "A comprehensive study on gait biometrics using a joint cnn-based method," *Pattern Recognition*, vol. 93, pp. 228–236, 2019.

[4] Kai Chen, Weihua Chen, Tao He, Rong Du, Fan Wang, Xiuyu Sun, Yuchen Guo, and Guiguang Ding, "Tagperson: A target-aware generation pipeline for person reidentification," *arXiv preprint arXiv:2112.14239*, 2021.

[5] Tongkun Xu, Weihua Chen, Pichao Wang, Fan Wang, Hao Li, and Rong Jin, "Cdtrans: Cross-domain transformer for unsupervised domain adaptation," *arXiv preprint arXiv:2109.06165*, 2021.

[6] 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*, Long Beach, CA, USA, 2019, pp. 740–749, IEEE.

[7] M Saquib Sarfraz, Arne Schumann, Andreas Eberle, and Rainer Stiefelhagen, "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*, Salt Lake City, UT, USA, 2018, pp. 420–429, IEEE.

[8] 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*, Honolulu, HI, USA, 2017, pp. 2530–2539, IEEE.

[9] 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*, Venice, Italy, 2017, pp. 774–783, IEEE.

[10] Rui Yu, Zhichao Zhou, Song Bai, and Xiang Bai, "Divide and fuse: A re-ranking approach for person reidentification," in *British Machine Vision Conference*, London, UK, 2017, pp. 135.1–135.13, BMVA Press.

[11] 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*, Seoul, Korea, 2019, pp. 4976–4985, IEEE.

[12] Thomas N Kipf and Max Welling, "Semi-supervised classification with graph convolutional networks," in *International Conference on Learning Representations*, Toulon, France, 2016, pp. 1–1, OpenReview.net.

[13] 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)*, Glasgow, UK, 2020, pp. 346–363, Springer.

[14] 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*, vol. Early, no. Access, pp. 1–1, 2020.

[15] Yichao Yan, Jie Qin, Jiaxin 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*, Seattle, WA, USA, 2020, pp. 2899–2908, IEEE.

[16] Liang Zheng, Liyue Shen, Lu Tian, Shengjin Wang, Jingdong Wang, and Qi Tian, "Scalable person reidentification: A benchmark," in *Proceedings of the IEEE international conference on computer vision*, Santiago, Chile, 2015, pp. 1116–1124, IEEE.

[17] 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*, Amsterdam, The Netherlands, 2016, pp. 17–35, Springer.

[18] 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*, Amsterdam, The Netherlands, 2016, pp. 868–884, Springer.

[19] 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*, Long Beach, CA, USA, 2019, pp. 0–0, IEEE.