





# Top-Push Polynomial Ranking Embedded Dictionary Learning for Enhanced Re-Id

Ying Chen<sup>1</sup><sup>a</sup>, De Cheng<sup>2</sup><sup>b</sup>, Zihui Li<sup>3</sup><sup>c</sup> and Andy Song<sup>1</sup><sup>d</sup>

<sup>1</sup>RMIT University, Melbourne, Australia

<sup>2</sup>Xidian University, Xi'an, China

<sup>3</sup>University of Science and Technology of China, Hefei, China

**Keywords:** Person Re-Identification, Dictionary Learning, Top-Push Polynomial Ranking Metric.


**Abstract:** Person re-identification (Re-Id) aims to match pedestrians captured by multiple non-overlapping cameras. In this paper, we introduce a novel dictionary learning approach enhanced with a top-push polynomial ranking metric for improved Re-Id performance. A key feature of our method is the incorporation of a ranking graph Laplacian term, designed to minimize intra-class compactness and maximize inter-class dispersion. Specifically, we employ a polynomial distance function to evaluate similarity between person images and propose the Top-push Polynomial Ranking Loss (TPRL) function, which enforces a margin between positive matching pairs and their closest non-matching pairs. The TPRL is then embedded into the dictionary learning objective, enabling our method to capture essential ranking relationships among person images—a critical aspect for retrieval-focused tasks. Unlike traditional dictionary learning approaches, our method reformulates ranking constraints through a graph Laplacian, resulting in an approach that is both straightforward to implement and highly effective. Extensive experiments on four popular Re-Id benchmark datasets demonstrate that our method consistently outperforms existing approaches, highlighting its effectiveness and robustness.


## 1 INTRODUCTION


Person re-identification (Re-Id) focuses on maintaining a consistent identity for individuals as they move across non-overlapping surveillance cameras, playing a critical role in video surveillance and attracting significant research interest (Huang et al., 2024; Gao et al., 2024). Recent approaches to Re-Id can be broadly categorized into two main types: similarity learning methods and feature representation learning methods. Similarity learning methods focus on learning distance metrics that accurately measure similarity between images captured by different cameras (Yuan et al., 2021; Liu et al., 2020), while feature representation learning methods aim to extract features from human images that are robust to variations in illumination, viewpoints, and poses, while remaining distinct for different individuals (Chang et al., 2018; Sha et al., 2023). Despite advancements, Re-


Id remains challenging due to several key issues: 1) A person's appearance can vary substantially across camera views due to occlusions, lighting conditions, viewpoint shifts, and pose changes in real-world scenarios; 2) People in public spaces often wear similar clothing (e.g., dark coats, jeans), leading to visual similarities between different individuals.

This paper presents a novel dictionary learning approach that integrates two essential components: (1) a reconstruction-focused module, which minimizes the discrepancy between original image features and their corresponding projected coefficients, and (2) a top-push polynomial ranking metric designed to ensure a pronounced margin between feature coefficients of identical identities and those of different identities. Central to this approach is the Top-Push Polynomial Ranking Loss (TPRL), which utilizes a polynomial distance function combining Mahalanobis and bilinear metrics to evaluate image similarity. The TPRL maximizes the separation between positive matches and their closest negative counterparts, embedding this objective directly within the dictionary learning framework. Unlike prior works in metric learning, our method uniquely incorporates ranking constraints

<sup>a</sup> <https://orcid.org/0009-0006-2681-1479>

<sup>b</sup> <https://orcid.org/0000-0003-4603-847X>

<sup>c</sup> <https://orcid.org/0000-0001-9642-8009>

<sup>d</sup> <https://orcid.org/0000-0002-7579-7048>

into dictionary learning through a graph Laplacian formulation, providing a structured approach to optimizing these relationships across datasets. Additionally, the polynomial distance measurement matrix is jointly optimized during dictionary learning, further bolstering the method’s effectiveness. By representing both gallery and probe images, the learned dictionary remains robust to variations in viewpoint, enabling it to encode features that are both discriminative for individuals and consistent within identities. This approach strengthens intra-class cohesion while amplifying inter-class distinctions.

In summary, the key contributions of this work are as follows:

- We introduce the Top-Push Polynomial Ranking Loss, which leverages a polynomial distance function to measure similarity and enforces a significant margin between positive matching pairs and their nearest non-matching counterparts.
- We reformulate the top-push polynomial ranking constraints into a graph Laplacian framework and integrate this directly into the dictionary learning process, enhancing the adaptability of traditional dictionary learning methods for person Re-Id tasks.
- Extensive experiments on multiple benchmark datasets demonstrate the effectiveness of the proposed discriminative dictionary learning approach, achieving state-of-the-art performance in person Re-Id.

## 2 ALGORITHM DESCRIPTION

This section begins with a brief overview of the fundamentals of dictionary learning. We then introduce our proposed approach, which embeds a ranking metric to learn dictionaries that are both discriminative and robust to viewpoint variations. Finally, we detail the optimization strategy for the proposed method.

### 2.1 The Proposed Top-Push Polynomial Ranking Distance Metric

Person Re-Id involves identifying a specific individual from a vast collection of gallery images captured across multiple cameras. Since ranking information is crucial for this task, we naturally integrate triplet ranking constraints into the objective function to enable discriminative dictionary learning. This ranking metric is designed to reduce the distance between coefficient vectors of samples belonging to the same individual while increasing the distance between those

of visually similar samples from different individuals. For person Re-Id, our objective is to ensure that the distance between samples of different individuals is significantly larger than that between samples of the same individual, maintaining a substantial margin.

Let the coding coefficient matrix be  $\mathbf{A} = [\mathbf{a}_1, \dots, \mathbf{a}_N] \in R^{K \times N}$ , which corresponds to the original data matrix  $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_N] \in R^{M \times N}$ . Each column of  $\mathbf{A}$  represents the transformed embedding of the respective data point  $\mathbf{x}_i$  in the new feature space. Using the training data, our objective function aims to guide the dictionary toward embeddings where the distance between images of the same person is significantly smaller than that between images of different individuals, maintaining a margin  $\tau$ . This formulation is inspired by the triplet loss function. However, unlike traditional triplet loss, this paper introduces the top-push constraint, which maximizes the margin specifically between the matching image pair and its closest non-matching pair, expressed as:

$$\Gamma(\mathbf{A}, \mathbf{W}) = \sum_{l_i=l_j} [f_{\mathbf{W}}(\mathbf{a}_i, \mathbf{a}_j) - \min_{l_i \neq l_k} f_{\mathbf{W}}(\mathbf{a}_i, \mathbf{a}_k) + \tau]_+, \quad (1)$$

where  $[x]_+$  takes zero if  $x < 0$ , and equals  $x$  otherwise.  $l_i$  is the labeled identity of the  $i$ -th training sample,  $\tau$  is the predefined max-margin parameter in the proposed top-push ranking loss, and  $f_{\mathbf{W}}$  is the distance function between two examples. We can clearly see that the proposed top-push ranking constraint maximizes the margin between each matching positive pair and its corresponding hardest non-matching negative pair.  $f_{\mathbf{W}}$  is the polynomial distance function used in (Chen et al., 2016a), which is the combination of Mahalanobis and the Bilinear distances. We can define it as follows,

$$\begin{aligned} f_{\mathbf{W}}(\mathbf{a}_i, \mathbf{a}_j) &= \langle \phi(\mathbf{a}_i, \mathbf{a}_j), [\mathbf{W}_M; \mathbf{W}_B] \rangle \\ &= \langle \phi_M(\mathbf{a}_i, \mathbf{a}_j), \mathbf{W}_M \rangle + \langle \phi_B(\mathbf{a}_i, \mathbf{a}_j), \mathbf{W}_B \rangle, \end{aligned} \quad (2)$$

where  $\langle \cdot, \cdot \rangle_F$  is the Frobenius inner product, and  $\mathbf{W} = [\mathbf{W}_M, \mathbf{W}_B]$ ,  $\phi(\mathbf{a}_i, \mathbf{a}_j) = [\phi_M(\mathbf{a}_i, \mathbf{a}_j), \phi_B(\mathbf{a}_i, \mathbf{a}_j)]$ . More specifically,

$$\begin{aligned} \phi_M(\mathbf{a}_i, \mathbf{a}_j) &= (\mathbf{a}_i - \mathbf{a}_j)(\mathbf{x}_i - \mathbf{x}_j)^T, \\ \phi_B(\mathbf{a}_i, \mathbf{a}_j) &= \mathbf{a}_i \mathbf{a}_j^T + \mathbf{a}_j \mathbf{a}_i^T. \end{aligned} \quad (3)$$

The part  $\langle \phi_M(\mathbf{a}_i, \mathbf{a}_j), \mathbf{W}_M \rangle_F = \|\mathbf{W}_M(\mathbf{a}_i - \mathbf{a}_j)\|_2^2 = (\mathbf{a}_i - \mathbf{a}_j)^T \mathbf{M}(\mathbf{a}_i - \mathbf{a}_j)$ , is connected to Mahalanobis distance, where  $M = \mathbf{W}_M^T \mathbf{W}_M$ . As we want to achieve low score when  $\mathbf{a}_i$  and  $\mathbf{a}_j$  are similar,  $\mathbf{W}_M$  should be positive semi-definite. The part  $\langle \phi_B(\mathbf{a}_i, \mathbf{a}_j), \mathbf{W}_B \rangle = \mathbf{a}_i^T \mathbf{W}_B \mathbf{a}_j + \mathbf{a}_j^T \mathbf{W}_B \mathbf{a}_i$ , corresponds to bilinear similarity. In order to simplify this bilinear similarity, we define the Bilinear matrix  $\mathbf{W}_B$  to be symmetric.

Then this part can be defined as  $\langle \phi_B(\mathbf{a}_i, \mathbf{a}_j), \mathbf{W}_B \rangle = 2\mathbf{a}_i^T \mathbf{W}_B \mathbf{a}_j$ . Thus, the similarity function can be written as Eq. (4),

$$f_{\mathbf{W}}(\mathbf{a}_i, \mathbf{a}_j) = \langle \phi(\mathbf{a}_i, \mathbf{a}_j), [\mathbf{W}_M; \mathbf{W}_B] \rangle \\ = \|\mathbf{W}_M(\mathbf{a}_i - \mathbf{a}_j)\|_2^2 + 2 \cdot \mathbf{a}_i^T \mathbf{W}_B \mathbf{a}_j. \quad (4)$$

We can clearly see that  $\phi_M(\mathbf{a}_i, \mathbf{a}_j)$  focus on measuring the similarity for descriptors at the same position.  $\phi_B(\mathbf{a}_i, \mathbf{a}_j)$  matches each patch in one image with all patches in the other image, and all the cross-patch similarities are attained as  $\mathbf{a}_i \mathbf{a}_j^T$  and  $\mathbf{a}_j \mathbf{a}_i^T$ . Both parts ensure the effectiveness of  $f(\mathbf{a}_i, \mathbf{a}_j)$ .

## 2.2 The Proposed Dictionary Learning Objective with TPRL Embedded

In the following, we reformulate the proposed TPRL into the metric form. As illustrated in Eq. 1, the proposed top-push ranking constraint  $\Gamma(\mathbf{A}, \mathbf{W})$  is constituted by all the positive sample pairs and their corresponding hardest negative sample pairs in the training dataset, where the distance between two samples  $(\mathbf{a}_i, \mathbf{a}_j)$  can be computed by the polynomial distance function  $f_{\mathbf{W}}(\mathbf{a}_i, \mathbf{a}_j)$ . Then, we innovatively reformulate Eq. 1 into the following metric form,

$$\Gamma(\mathbf{A}, \mathbf{W}) = \sum_{i,j=1}^N s_{ij} f_{\mathbf{W}}(\mathbf{a}_i, \mathbf{a}_j) + C(\tau) \\ = \sum_{i,j=1}^N s_{ij} [\|\mathbf{W}_M(\mathbf{a}_i - \mathbf{a}_j)\|_2^2 + 2 \cdot \mathbf{a}_i^T \mathbf{W}_B \mathbf{a}_j] + C(\tau) \\ = 2Tr(\mathbf{W}_M \mathbf{A} \Psi_M \mathbf{A}^T \mathbf{W}_M^T) + 2Tr(\mathbf{A} \Psi_B \mathbf{A}^T \mathbf{W}_B) + C(\tau) \\ = 2Tr(\mathbf{W}_M \mathbf{A} \Psi_M \mathbf{A}^T \mathbf{W}_M^T + \mathbf{A} \Psi_B \mathbf{A}^T \mathbf{W}_B) + C(\tau). \quad (5)$$

where  $C(\tau)$  is constant depending on parameter  $\tau$ ,  $s_{ij}$  is the adjacent weight for the pair-wise sample distance  $f_{\mathbf{W}}(\mathbf{a}_i, \mathbf{a}_j)$ . In Eq.(5), the parameter  $\mathbf{W} = [\mathbf{W}_M, \mathbf{W}_B]$  is to be learnt, and  $\Psi_M = \mathbf{G} - (\mathbf{S} + \mathbf{S}^T)/2$ ,  $\mathbf{G} = \text{diag}(g_{11}, \dots, g_{NN})$ ,  $g_{ii} = \sum_{j=1, j \neq i}^N \frac{s_{ij} + s_{ji}}{2}$ ,  $j = 1, 2, \dots, N$ , and  $\Psi_B = (\mathbf{S} + \mathbf{S}^T)/2$ .  $\Psi_M$  and  $\Psi_B$  are the Laplacian matrix of  $\mathbf{S}$ , and  $Tr(\cdot)$  denotes the trace of a matrix. The deduction from line 2 to 3 in Eq. 5 can refer to (Shi et al., 2016). The element  $s_{ij}$  of the adjacent matrix  $\mathbf{S}$  in Eq. 5 can be deduced from Eq. (5) and Eq. (1) as follows:

$$s_{ij} = \begin{cases} \delta[f_{\mathbf{W}}(\mathbf{a}_i, \mathbf{a}_j) - \min_{k=1, \dots, n} f_{\mathbf{W}}(\mathbf{a}_i, \mathbf{a}_k) + \tau], & l_i = l_j \neq l_k, i \neq j, \\ - \sum_{\substack{k=1, \\ l_i = l_k \neq l_j}}^N \varepsilon(j == j_{i.min}) \delta[f_{\mathbf{W}}(\mathbf{a}_i, \mathbf{a}_k) \\ - f_{\mathbf{W}}(\mathbf{a}_i, \mathbf{a}_j) + \tau], i \neq j, 0, & i = j. \end{cases} \quad (6)$$

where the function  $\delta[\cdot]$  is an indicator function which takes one if the argument is bigger than zero, and zeros otherwise.  $\varepsilon(x)$  is another indicator function which takes one if the argument inside the brackets is true, and zero otherwise.  $j_{i.min} = \text{argmin}_{j=1, \dots, N, l_j \neq l_i} f_{\mathbf{W}}(\mathbf{a}_i, \mathbf{a}_j)$ .

Therefore, the proposed dictionary learning algorithm with polynomial ranking metric embedded arrives at:

$$\text{argmin}_{\mathbf{D}, \mathbf{A}, \mathbf{W}} \|\mathbf{X} - \mathbf{D}\mathbf{A}\|_F^2 + \frac{\beta}{N(\tau)} Tr(\mathbf{W}_M \mathbf{A} \Psi_M \mathbf{A}^T \mathbf{W}_M^T \\ + \mathbf{A} \Psi_B \mathbf{A}^T \mathbf{W}_B) + \lambda \|\mathbf{A}\|_F^2 + \alpha_1 \|\mathbf{W}_M\|_F^2 + \alpha_2 \|\mathbf{W}_B\|_F^2 \\ s.t. \quad \|\mathbf{d}_i\|_2^2 \leq 1, \forall i, \quad (7)$$

where  $C(\tau)$  has been ignored from Eq. 5 as the constant has no influence on the objective, and  $N(\tau)$  is the number of all the selected sample triplets constructed by the  $N$  training examples with hardest negative miming process. The parameters  $\lambda$ ,  $\alpha_1$ ,  $\alpha_2$  and  $\beta$  are used to control the contributions of the corresponding terms. In Eq. 7, the first term denotes the reconstruction error. The second term is the embedded TPRL which maintains the distance of similar sample pairs to be closer than that of the closest dissimilar pairs by a large margin in the learned dictionary space, thus reduce the intra-personal variations. The last three terms are the regularization terms to avoid over-fitting.

## 3 EXPERIMENTS

In this section, we use four widely used person Re-Id benchmark datasets, namely VIPeR (Gray et al., 2007), 3DPES (Baltieri et al., 2011), CUHK01 (Li et al., 2014) and CUHK03 (Li et al., 2014), for performance evaluation. All the datasets contain a set of individuals, each of whom has several images captured by different camera views.

### 3.1 Experimental Setup

**Feature Representation.** We have used two kinds of features in our experiments: One is the traditional handcraft features, which includes colour+HOG+LBP, HSV, LAB SILPT, and each of them is extracted both in the whole image and the image subregions. Details about the 7538-D feature representations can refer to (Peng et al., 2016; Chen et al., 2016a). Another is the 2048-D deep residual network features(ResNet152) (He et al., 2016). Note that the 2048-D deep feature is extracted by the original ResNet152 model trained on the ImageNet

dataset, it was not fine-tuned on any person Re-Id dataset. Then we mix them to form a single feature vector for each image.

**Parameter Setting.** We empirically set the dictionary size for  $\mathbf{D}$  in Eq. 7 as  $K = 250$ . The parameters  $\tau, \alpha_1, \alpha_2, \beta$  and  $\lambda$  are set to 1.0, 0.2, 0.2, 0.7 and 0.35, respectively. The learning rate for optimizing Eq. 7 starts with  $\eta = 0.01$ , then at each iteration, we increase  $\eta$  by a factor of 1.2 if the loss function decreased and decrease  $\eta$  by a factor of 0.8 if the loss increased.

**Evaluation Protocol.** Our experiments follow the evaluation protocol in (Peng et al., 2016). The dataset is separated into the training and test set, where images of the same person can only appear in either set. The test set is further divided into the probe and gallery set, and two sets contain different images of a same person. In the VIPeR, 3DPES and CUHK01 datasets, half of the identities are used as training or test set, while in the CUHK03 dataset, 100 pedestrians are used as the test set, and the rest are used as the training set. We match each probe image with every image in the gallery set, and rank the gallery images according to their distance. The results are evaluated by the widely used CMC (Cumulative Matching Characteristic) metric (Peng et al., 2016).

### 3.2 Experimental Evaluations

As illustrated in Eq. (7), the proposed method mainly contains two components: the first one is the traditional dictionary learning method, which minimizes the reconstruction error between the input image features and the learned coding vectors; the second term is the proposed top-push polynomial ranking distance metric, where the polynomial distance metric is the combination of Mahalanobis and bilinear distances. In order to reveal how each ingredient contributes to the performance improvement, we implemented the following six variants of the proposed method, and compared them with many representative works in the literature:

**Variant 1** (denoted as DictL). We implement the dictionary learning method with the previously used Laplacian matrix embeddings, which just used the same identity information, and the matrix is constructed in the following way:  $s_{ij} = 1$  only if  $l_i = l_j, i \neq j$ , otherwise  $s_{ij} = 0$ , and the distance between two sample images is denoted as  $f(\mathbf{a}_i, \mathbf{a}_j) = \|\mathbf{a}_i - \mathbf{a}_j\|_2^2$ . This is our baseline method.

**Variant 2** (denoted as DictR). We implement the dictionary learning method as illustrated in Eq. 7,

but with the projection matrix  $\mathbf{W} = [\mathbf{W}_M; \mathbf{W}_B]$  removed (equal to set  $\mathbf{W}_M = \mathbf{I}$ , and  $\mathbf{W}_B = -\mathbf{I}$ , where  $\mathbf{I}$  is the identity matrix).

**Variant 3** (denoted as DictRWM). We implement the dictionary learning method as illustrated in Eq. 7, but we only use the Mahalanobis distance to measure the similarity between two images. That is to say, we just get rid of the term “ $Tr(\mathbf{A}\Psi_B\mathbf{A}^T\mathbf{W}_B)$ ” and “ $\alpha_2\|\mathbf{W}_B\|_F^2$ ” in Eq. 7 to train the model.

**Variant 4** (denoted as DictRWB). We implement the dictionary learning method as illustrated in Eq. 7, but we only use the bilinear distance to measure the similarity between two images. That is to say, we just get rid of the term “ $Tr(\mathbf{W}_M\mathbf{A}\Psi_M\mathbf{A}^T\mathbf{W}_M^T)$ ” and “ $\alpha_1\|\mathbf{W}_M\|_F^2$ ” in Eq. 7 to train the model.

**Variant 5** (denoted as Ours(DictRWMB)). This is our proposed final dictionary learning based method as illustrated in Eq. 7.

Table 1, 2, 3 and 4 show the evaluation results on VIPeR, 3DPES, CUHK01 and CUHK03 datasets, respectively, using the rank 1, 5, 10, 20, 30 accuracies. Each table includes the recently reported evaluation results. The compared methods include the approaches based on metric learning (Jose and Fleuret, 2016; Chen et al., 2016a; Bai et al., 2017; Zhou et al., 2017), common subspace based methods (Chen et al., 2015; Prates et al., 2015; Liao et al., 2015; Lisanti et al., 2014; Prates et al., 2016; Zhang et al., 2016b; Barman and Shah, 2017), and the deep learning based methods (Chen et al., 2016b; Varior et al., 2016a; Ahmed et al., 2015; Wang et al., 2016). Compared with all the aforementioned representative works, our model(DictRWMB) has achieved the top performances on the four person Re-Id benchmark datasets, with all the five ranking measurements. We achieve the rank-1 accuracy to 56.9%, 61.2%, 61.5% and 77.1% on VIPeR, 3DPES, CUHK01 and CUHK03 datasets, respectively. The evaluation results shown in Table 1,2,3 and 4 can be summarized as follows,

- Compared with many recently reported representative works, our method(DictRWMB) outperforms all the compared metric learning based methods on all the datasets by a margin of 2.0% at top 1 accuracy on average. We can also outperform the deep learning based methods on relatively small datasets, while get comparable results with some deep learning based methods on the relatively large datasets.
- With the proposed TPRL embedded, the performance accuracies can get up to 4.3% – 11.9%

Table 1: Experimental results on VIPeR dataset( $p=316$ ).

Method	r=1	r=5	r=10	r=20	r=30
(Prates et al., 2016)	35.8	69.1	80.8	89.9	93.8
(Chen et al., 2016b)	38.4	69.2	81.3	90.4	94.1
(Xiong et al., 2014)	39.2	71.8	81.3	92.4	94.9
(Lisanti et al., 2014)	37.0	--	85.0	93.0	--
(Liao et al., 2015)	40.0	68.0	80.5	91.1	95.5
(Jose and Fleuret, 2016)	40.2	68.2	80.7	91.1	--
(Yang et al., 2016)	41.1	71.7	83.2	91.7	--
(Zhang et al., 2016b)	42.3	71.5	82.9	92.1	--
(Chen et al., 2015)	43.0	75.8	87.3	94.8	--
(Ahmed et al., 2015)	45.9	77.5	88.9	95.8	--
(Matsukawa et al., 2016)	49.7	79.7	88.7	94.5	--
(Chen et al., 2016a)	53.5	82.6	91.5	96.6	--
(Barman and Shah, 2017)	34.2	57.3	67.6	80.7	--
(Zhou et al., 2017)	44.9	74.4	84.9	93.6	--
(Bai et al., 2017)	53.5	82.6	91.5	96.6	--
DictL(baseline)	52.6	77.5	85.9	91.8	94.6
DictR	55.1	82.7	90.7	95.8	97.3
DictRWM	56.3	82.9	91.5	96.9	97.8
DictRWB	55.7	81.9	90.9	96.1	97.3
DictRWMB	<b>56.9</b>	<b>83.8</b>	<b>92.5</b>	<b>97.4</b>	<b>98.3</b>

improvement compared with the baseline dictionary learning method. By comparing the method “DictR” and “DictL”, we can clearly see that the ranking information is better than only using the same identity information.

- We can clearly see that embedding either the Mahalanobis or bilinear distance function into the dictionary learning objective can contribute to the performance improvements. When the proposed polynomial distance function is embedded, much better performance improvements can be obtained.

Since we have used two kinds of features in our experiments (the handcraft and the ResNet152 features), we also did experiments to reveal their performances in Table 5, respectively. We can clearly see that combining the traditional handcraft features with the deep learning based features can further improve the Re-Id performances.

### 3.3 Parameter Analysis of the Method

As defined in Eq. 7, there are two important parameters in our proposed ranking metric embedded dictionary learning method, one is the dictionary size  $K$  of  $\mathbf{D}$ , and the other is the parameter  $\beta$ , which controls the balance between the dictionary construction loss and the ranking graph laplacian cost. To investigate

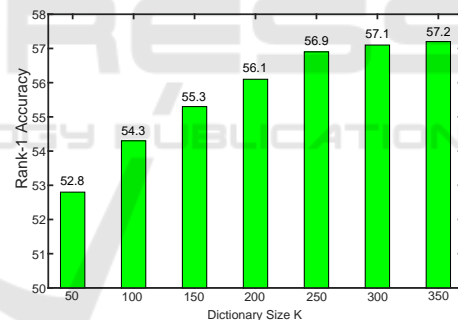


Figure 1: Parameter Analysis: We report how the rank-1 accuracy changes with the dictionary size  $K$ .

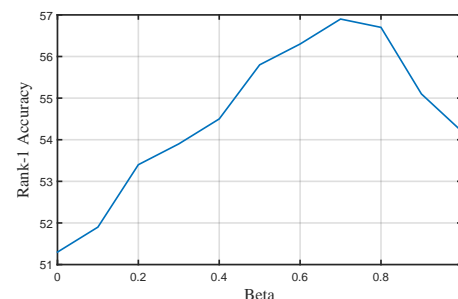


Figure 2: Parameter Analysis: We report how the rank-1 accuracy changes with the parameter  $\beta$  which controls the balance between the dictionary construction loss and the TPRL component.

Table 2: Experimental results on 3DPES dataset( $p=92$ ).

Method	r=1	r=5	r=10	r=20	r=30
(Koestinger et al., 2012)	34.2	58.7	69.6	80.2	--
(Mignon and Jurie, 2012)	43.5	71.6	81.8	91.0	--
(Pedagadi et al., 2013)	45.5	69.2	70.1	82.1	88.2
(Xiong et al., 2014)	54.0	77.7	85.9	92.4	--
(Paisitkriangkrai et al., 2015)	53.3	76.8	85.7	91.4	--
(Xiao et al., 2016)	55.2	76.4	84.9	91.9	94.1
(Chen et al., 2016a)	57.3	78.6	86.5	93.6	95.2
DictL(baseline)	55.3	76.3	83.7	91.4	94.2
DictR	59.0	80.7	87.6	94.1	95.8
DictRWM	60.5	81.7	89.6	96.1	96.8
DictRWB	59.8	81.3	88.7	95.2	96.7
DictRWMB	<b>61.2</b>	<b>82.4</b>	<b>91.7</b>	<b>96.8</b>	<b>97.7</b>

Table 3: Experimental results on CUHK01 dataset( $p=486$ ).

Method	r=1	r=5	r=10	r=20	r=30
(Prates et al., 2016)	38.3	66.8	77.7	86.8	90.5
(Chen et al., 2015)	40.4	64.6	75.3	84.1	--
(Ahmed et al., 2015)	47.5	71.6	80.3	87.5	--
(Xiong et al., 2014)	49.6	74.7	83.8	91.2	94.3
(Ahmed et al., 2015)	53.4	76.4	84.4	90.5	--
(Chen et al., 2016a)	56.8	87.6	89.5	92.3	94.7
(Matsukawa et al., 2016)	57.8	79.1	86.2	92.1	--
(Li et al., 2015)	59.5	81.3	89.7	93.1	--
(Prates et al., 2016)	61.2	80.9	87.3	93.2	95.6
DictL(baseline)	56.2	79.5	84.7	90.6	93.0
DictR	59.7	81.5	89.0	92.8	96.2
DictRWM	61.1	82.8	90.1	94.3	96.5
DictRWB	60.6	82.3	89.7	93.8	95.2
DictRWMB	<b>61.5</b>	<b>83.6</b>	<b>90.7</b>	<b>94.4</b>	<b>96.7</b>

the effect of the dictionary size  $K$  and the parameter  $\beta$  on the rank-1 accuracy, we conduct experiments using cross validation method on VIPeR dataset, and the rank-1 results are shown in Fig. 1 and Fig. 2.

Figure 1 illustrates the rank-1 accuracy with different dictionary size  $K$  from 50 to 350. We can see that firstly as the dictionary size becomes larger, the performance increases continuously. After the dictionary size  $K$  larger than 250, the performance increases very slowly. Although higher performance can also be obtained with larger dictionary size, we choose  $K = 250$  in all our experiments, since larger dictionary size requires more training and testing time.

Figure 2 shows the rank-1 accuracy with different parameter  $\beta$  from 0 to 1.0. We can clearly see that our proposed method yields the best rank-1 performance when  $\beta = 0.7$ . Thus, we set  $\beta$  to 0.7 in all our experimental evaluations.

## 4 CONCLUSION

In this paper, we present a novel dictionary learning method with the TPRL embedded, for person Re-Id. Specially, we first adopt the polynomial distance function to measure the similarity between two different person images; and then we propose the top-push polynomial ranking loss function, which maximizes the margin between the positive matching image pair and its closest non-matching image pair; Finally, we reformulate the TPRL into the graph Laplacian form, and then embedded it into the dictionary learning objectives. Experiment results illustrate the effectiveness of the proposed method. It shows that the proposed ranking graph Laplacian term is very essential for such retrieval related tasks, especially for person Re-Id. Overall, our proposed method have made the traditional dictionary learning method more suitable

Table 4: Experimental results on CUHK03 labeled dataset(p=100).

Method	r=1	r=5	r=10	r=20	r=30
(Zhao et al., 2013)	8.8	24.1	38.3	53.4	--
(Koestinger et al., 2012)	14.2	48.5	52.6	---	---
(Li et al., 2014)	20.6	51.5	66.5	80.0	---
(Liao et al., 2015)	52.2	82.2	92.1	96.2	---
(Xiong et al., 2014)	48.2	59.3	66.4	---	---
(Wang et al., 2016)	52.2	83.7	89.5	94.3	96.5
(Ahmed et al., 2015)	54.7	86.5	94.0	96.1	98.0
(Varior et al., 2016b)	57.3	80.1	88.3	---	---
(Paisitkriangkrai et al., 2015)	62.1	89.1	94.3	97.8	---
(Zhang et al., 2016a)	58.9	85.6	92.5	96.3	---
(Wu et al., 2016)	63.2	90.0	92.7	97.6	---
(Varior et al., 2016a)	68.1	88.1	94.6	---	---
(Zhou et al., 2017)	61.6	88.3	95.2	98.4	---
(Bai et al., 2017)	76.6	94.6	<b>98.0</b>	---	---
DictL(baseline)	65.2	83.5	88.7	93.6	96.0
DictR	73.2	90.3	93.3	96.8	98.0
DictRWM	75.3	93.3	94.5	97.8	98.0
DictRWB	74.2	91.4	93.6	97.5	98.0
DictRWMB	<b>77.1</b>	<b>94.7</b>	97.9	<b>98.0</b>	<b>99.0</b>

Table 5: Experiments comparison with handcraft(HC), ResNet152 features and its combination on CUHK03 datasets, respectively.

Method	r=1	r=5	r=10	r=20	r=30
DictRWMB(ResNet152)	47.7	79.8	88.7	95.1	97.2
DictRWMB(HC)	72.3	91.7	94.9	97.1	98.0
DictRWMB(HC+ResNet152)	<b>77.1</b>	<b>94.7</b>	<b>97.9</b>	<b>98.0</b>	<b>99.0</b>

for the retrieval related tasks. In the future, we will deploy our approach to other tasks, such as image and video retrieval.

## REFERENCES

Ahmed, E., Jones, M., and Marks, T. K. (2015). An improved deep learning architecture for person re-identification. In *CVPR*.

Bai, S., Bai, X., and Tian, Q. (2017). Scalable person re-identification on supervised smoothed manifold. In *CVPR*.

Baltieri, D., Vezzani, R., and Cucchiara, R. (2011). 3dpes: 3d people dataset for surveillance and forensics. In *ACM workshop on Human gesture and behavior understanding*.

Barman, A. and Shah, S. K. (2017). Shape: A novel graph theoretic algorithm for making consensus-based decisions in person re-identification systems. In *ICCV*.

Chang, X., Huang, P., Shen, Y., Liang, X., Yang, Y., and Hauptmann, A. G. (2018). RCAA: relational context-aware agents for person search. In Ferrari, V., Hebert, M., Sminchisescu, C., and Weiss, Y., editors, *ECCV*.

Chen, D., Yuan, Z., Chen, B., and Zheng, N. (2016a). Similarity learning with spatial constraints for person re-identification. In *CVPR*.

Chen, S.-Z., Guo, C.-C., and Lai, J.-H. (2016b). Deep ranking for person re-identification via joint representation learning. *IEEE TIP*.

Chen, Y.-C., Zheng, W.-S., and Lai, J. (2015). Mirror representation for modeling view-specific transform in person re-identification. In *IJCAI*.

Gao, J., Jiang, X., Dou, S., Li, D., Miao, D., and Zhao, C. (2024). Re-id-leak: Membership inference attacks against person re-identification. *Int. J. Comput. Vis.*, 132(10):4673–4687.

Gray, D., Brennan, S., and Tao, H. (2007). Evaluating appearance models for recognition, reacquisition, and tracking. In *ECCV*.

He, K., Zhang, Ren, S., and Sun, J. (2016). Deep residual learning for image recognition. In *CVPR*.

Huang, Y., Zhang, Z., Wu, Q., Zhong, Y., and Wang, L. (2024). Attribute-guided pedestrian retrieval: Bridging person re-id with internal attribute variability. In *CVPR*.

Jose, C. and Fleuret, F. (2016). Scalable metric learning via weighted approximate rank component analysis. In *ECCV*.

Koestinger, M., Hirzer, M., Wohlhart, P., Roth, P. M., and Bischof, H. (2012). Large scale metric learning from equivalence constraints. In *CVPR*.

Li, S., Shao, M., and Fu, Y. (2015). Cross-view projective dictionary learning for person re-identification. In *IJCAI*.

Li, W., Zhao, R., Xiao, T., and Wang, X. (2014). Deep-reid: Deep filter pairing neural network for person re-identification. In *CVPR*.

Liao, S., Hu, Y., Zhu, X., and Li, S. Z. (2015). Person re-identification by local maximal occurrence representation and metric learning. In *CVPR*.

Lisanti, G., Masi, I., and Del Bimbo, A. (2014). Matching people across camera views using kernel canonical correlation analysis. In *ICDSC*.

Liu, W., Chang, X., Chen, L., Phung, D., Zhang, X., Yang, Y., and Hauptmann, A. G. (2020). Pair-based uncertainty and diversity promoting early active learning for person re-identification. *ACM Trans. Intell. Syst. Technol.*, 11(2):21:1–21:15.

Matsukawa, T., Okabe, T., Suzuki, E., and Sato, Y. (2016). Hierarchical gaussian descriptor for person re-identification. In *CVPR*.

Mignon, A. and Jurie, F. (2012). Pcca: A new approach for distance learning from sparse pairwise constraints. In *CVPR*.

Paisitkriangkrai, S., Shen, C., and van den Hengel, A. (2015). Learning to rank in person re-identification with metric ensembles. In *CVPR*.

Pedagadi, S., Orwell, J., Velastin, S., and Boghossian, B. (2013). Local fisher discriminant analysis for pedestrian re-identification. In *CVPR*.

Peng, P., Xiang, T., Wang, Y., Pontil, M., Gong, S., Huang, T., and Tian, Y. (2016). Unsupervised cross-dataset transfer learning for person re-identification. In *CVPR*.

- Prates, R., Felipe, and Schwartz, W. R. (2015). Appearance-based person re-identification by intra-camera discriminative models and rank aggregation. In *International Conference on Biometrics*.
- Prates, R., Oliveira, M., and Schwartz, W. R. (2016). Kernel partial least squares for person re-identification. In *AVSS*.
- Sha, B., Li, B., Chen, T., Fan, J., and Sheng, T. (2023). Rethinking pseudo-label-based unsupervised person re-identification with hierarchical prototype-based graph. In *ACM MM*.
- Shi, W., Gong, Y., and Jinjun (2016). Improving cnn performance with min-max objective. In *IJCAI*.
- Variator, R. R., Haloi, M., and Wang, G. (2016a). Gated siamese convolutional neural network architecture for human re-identification. In *ECCV*.
- Variator, R. R., Shuai, B., Lu, J., Xu, D., and Wang, G. (2016b). A siamese long short-term memory architecture for human re-identification. In *ECCV*.
- Wang, F., Zuo, W., Lin, L., Zhang, D., and Zhang, L. (2016). Joint learning of single-image and cross-image representations for person re-identification. In *CVPR*.
- Wu, L., Shen, C., and van den Hengel, A. (2016). Deep linear discriminant analysis on fisher networks: A hybrid architecture for person re-identification. *Pattern Recognition*.
- Xiao, T., Li, H., Ouyang, W., and Wang, X. (2016). Learning deep feature representations with domain guided dropout for person re-identification. *CVPR*.
- Xiong, F., Gou, M., Camps, O., and Szaier, M. (2014). Person re-identification using kernel-based metric learning methods. In *ECCV*.
- Yang, Y., Lei, Z., Zhang, S., Shi, H., and Li, S. Z. (2016). Metric embedded discriminative vocabulary learning for high-level person representation. In *AAAI*.
- Yuan, D., Chang, X., Huang, P., Liu, Q., and He, Z. (2021). Self-supervised deep correlation tracking. *IEEE Trans. Image Process.*, 30:976–985.
- Zhang, L., Xiang, T., and Gong, S. (2016a). Learning a discriminative null space for person re-identification. In *CVPR*.
- Zhang, Y., Li, B., Lu, H., Irie, A., and Ruan, X. (2016b). Sample-specific svm learning for person re-identification. In *CVPR*.
- Zhao, R., Ouyang, W., and Wang, X. (2013). Unsupervised salience learning for person re-identification. In *CVPR*.
- Zhou, J., Yu, P., Tang, W., and Wu, Y. (2017). Efficient online local metric adaptation via negative samples for person re-identification. In *CVPR*.