Person re-identification via semi-supervised adaptive graph embedding

Jiao Liu[1,2] · Mingquan Lin[2] · Mingbo Zhao[2,3] · Choujun Zhan[2,3] · Bing Li[4] · John Kwok Tai Chui[5]

Abstract

Video surveillance is an indispensable part of the smart city for public safety and security. Person Re-Identification (Re-ID), as one of elementary learning tasks for video surveillance, is to track and identify a given pedestrian in a multi-camera scene. In general, most existing methods has firstly adopted a CNN based detector to obtain the cropped pedestrian image, it then aims to learn a specific distance metric for retrieval. However, unlabeled gallery images are generally overlooked and not utilized in the training. On the other hands, Manifold Embedding (ME) has well been applied to Person Re-ID as it is good to characterize the geometry of database associated with the query data. However, ME has its limitation to be scalable to large-scale data due to the huge computational complexity for graph construction and ranking. To handle this problem, we in this paper propose a novel scalable manifold embedding approach for Person Re-ID task. The new method is to incorporate both graph weight construction and manifold regularized term in the same framework. The graph we developed is discriminative and doubly-stochastic so that the side information has been considered so that it can enhance the clustering performances. The doubly-stochastic property can also guarantee the graph is highly robust and less sensitive to the parameters. Meriting from such a graph, we then incorporate the graph construction, the subspace learning method in the unified loss term. Therefore, the subspace results can be utilized into the graph construction, and the updated graph can in turn incorporate discriminative information for graph embedding. Extensive simulations is conducted based on three benchmark Person Re-ID datasets and the results verify that the proposed method can achieve better ranking performance compared with other state-of-the-art graph-based methods.

Keywords Semi-supervised learning · Dimensionality reduction · Adaptive graph embedding

1 Introduction

Smart Cities are evolving along a number of paths, such as social services and governance, water and power networks, healthcare and transportation systems et al. And with the increasing number of vehicles, buildings and citizens, many studies have been started to develop extensive big data processing and analysis methods for modeling human behavior, urban systems, et al., while some new technique, such as memetic computing, have been adopted or applied

on these challenging tasks. Among different real-world area for smart cities, video surveillance is an indispensable part to ganrantee the public security, especially for the traffic and safety management. As one of critical learning tasks in video surveillance, Person Re-Identification (Re-ID) is to track and identify a given pedestrian in a multi-camera scene in different challenging environments [1–5], such as camera views, person poses and changes in illumination, background clutter, and occlusion. In general, most currently state-of-the-art Person Re-ID is first to adopt a CNN based detector to obtain the cropped pedestrian image, it is then to learn a specific distance metric for ranking, where the distance is usually measured by the similarity between two low-dimensional image features extracted from CNN framework [4, 5]. For example, the work in [6] has developed FastReID to provide strong baselines that are capable of achieving the state-of-the-art performance, where the pedestrian images are firstly cropped from video sequence by human labeler or selected by CNN

Jiao Liu and Mingquan Lin contribute equally to this work.

✉ Mingbo Zhao
  mbzhao4@gmail.com

✉ Choujun Zhan
  zchoujun2@gmail.com

Extended author information available on the last page of the article.

based detector, and then extensive CNN backbones with certain aggregation strategies and loss functions are trained to extract the feature of pedestrian images for distance metric. In addition, to handle a real-world video Re-ID problem in surveillance applications, [1] has developed a video-based Person Re-ID by exploring useful information captured from sequences. In detail, it firstly crops the image sequences belonging to a certain pedestrian via certain detector. It then trains a special aggregation network with these sequential data to grasp both spatial and temporal information for feature extraction, where these features can be utilized for distance metric for realizing Re-ID task.

However, the distance-based ranking methods only focuses on the pair-wise similarity among training dataset but the distribution of the whole dataset is neglected and not considered. In some real-world application case, e.g., the pose of the pedestrian changes greatly, the pedestrian image with the front view may be quite different from the one with back view. In such case, the system may judge them belonging to uncorrelated pedestrians, since distance-based ranking method only consider pair-wise similarity. However, if the similarity is measured by involving the distribution information or intrinsic manifold, such problem may be avoided, as the pedestrian images belonging to the same person but with different view are in the same manifold. This is the main drawback of distance-based ranking method. Obviously, a good Re-ID method should consider image features as well as the intrinsic structure of the image database. To handle the above problem, a key issue is to how to involve the data distribution or manifold information into the calculating of similarity between any two pair-wise images, so that the extracted feature can better provide more geometrical structure for guiding the distance-based ranking method. Motivated by this end, we in this paper consider the graph embedding method by taking the underlying structure into account for distance metrics.

The Graph Embedding (GE) methods [7–10] are developed to preserve the properties of a graph for feature extraction, where these methods are firstly to construct a graph to approximate the geometrical structure of data and then to learn a project matrix to cast the high-dimensional dataset to low-dimensional subspace while preserving the geometrical information. From this point of view, ME can naturally preserve the distribution of dataset. many GE methods have been proposed during the past decades in order to discover the intrinsic geometrical structure of data manifold [11–19]. Typical methods include ISOMAP [20], Locally Linear Embedding (LLE) [21] and Laplacian Eigenmap (LE) [22]. But these methods cannot handle the new-coming samples, hence suffering out-of-sample problems. To solve this problem, He et al. has developed Locality Preserving Projections (LPP) [23] to project the original high-dimensional data into low-dimensional

subspace with a projection matrix. Therefore, the out-of-sample problem can be naturally handled as the projection matrix can be used for handling new-coming samples. Cai et al. [24, 25] further pointed that LPP is time-consuming especially for calculating the generalized eigenvalue problem given the data matrix is with high-dimensionality. They therefore proposed Spectral Regression (SR) to handle the above problem, in which they first to calculate the low-dimensional subspace, and then to calculate the projection via solving a regression problem. Recently, Nie et al. [26] have detailed analyzed the equivalent relationship between LPP and SR, through which they proposed a more efficient and effective method, namely Unsupervised Large Graph Embedding (ULGE), for dimensionality reduction and subspace clustering [27, 28].

On the other hand, graph neural network (GNN) is one of the most popular topic during the past few years, which has been proved to be effective for dealing with structured data [29–32]. For example, the work in [33, 34] have applied GCN for handling diagnosis problem of COVID-19. The goal of GNN [29, 30] is to integrate local geometrical structure and data features in the convolutional layers, so that both the connectivity patterns and feature attributes of graph-structured data can be preserved via the graph convolution and outperforms many state-of-the-art methods significantly. However, most of these methods could only be utilized to small-scale graph due to the computation limitation; another type of GNN is Spatial GNN [31, 32], which is to constructs a new feature vector for each vertex by aggregating the neighborhood information. But these methods need a predefined local graph so that they cannot learn the global information. Our work can be potentially viewed to make GNN to deal with large-scale graph without local sampling.

In general, the above graph construction methods are good to characterize the geometrical structure of data manifold. But these methods are all unsupervised, which means they do not utilize the discriminative information, such as class label or side information, which is good to enhance the clustering performance. Since the graph construction is unsupervised, it may confront the case that the data belonging to two different clusters or classes are connected. In such case, the graph construction tends to not be correct and the learned graph embedding results may cause some mistakes. Another problem is that these methods are usually two-stage approach, i.e. the graph construction and subspace learning are separated. The two stages do not have any information to improve graph embedding. In practice, we hope the geometrical information learned by the graph embedding procedure can well be used to enhance the local preserving ability of graph construction, and the modified graph construction can in turn promote the learning of graph embedding [35].

In this paper, to handle this problem, we develop a new spectral regression method for dimensionality reduction by integrating both graph construction and manifold regularized term into the same framework. In order to handle the case that nearby data points with different clusters may be mis-connected, we in this paper has adopted the side information for graph construction, which is utilized to make the constraint that the pair-wise data points have no link when they are clearly belonged to different clusters. In addition, we further involve the normalized and symmetrical constraint to make the graph doubly-stochastic. As mentioned by the work [36], the doubly-stochastic property can also guarantee the graph is highly robust and less sensitive to the parameters. Meriting from such a graph, we then incorporate the graph construction, the subspace learning method in the unified loss term. Therefore, the subspace results can be utilized into the graph construction, and the updated graph can in turn incorporate discriminative information for graph embedding. Extensive simulations have been conducted based on several benchmark datasets. Simulation results indicate that the proposed work can achieve much better clustering and image retrieval performance than other methods.

The main contribution of the proposed work are as follows:

1) we have proposed a discriminative and doubly-stochastic graph for characterizing the geometrical of data manifold. By adopting the side information into consideration, the proposed graph can guarantee the data points with different clusters will not be mis-connected. In addition, the normalized and symmetrical constraint can make the graph doubly-stochastic so that the graph is highly robust and less sensitive to the parameters;

2) we have incorporated the graph construction, the subspace learning method in the unified loss term. Therefore, the subspace results can be utilized into the graph construction, and the updated graph can in turn incorporate discriminative information for graph embedding;

3) we also developed an iterative solution to handle the above optimization problem. Theoretical analysis has shown the convergence.

4) we have applied the proposed method on Person Re-Identification. Extensive simulation results also verify the effectiveness of the proposed method.

This paper is structured as follows: in Section 2, we have briefly review some related work about graph embedding; in Section 3, we give detail description of the proposed work for graph embedding; in Section 4, we will conduct extensive simulations to show the effectiveness of the

proposed work on Person Re-Identification and final conclusion is drawn in Section 5.

2 Related work

In this paper, we first give some notations used in this work and review the work of LE, LPP, SR and ULSR.

2.1 Notations

Specifically, let $X = \{x_1, x_2, \cdots, x_l\} \in R^{D \times l}$ be the original-high dimensional dataset, each x_i belongs to a class $c_i = \{1, 2, \cdots, c\}$, where l_i be the number of data points in the ith class and l be the number of data points in all classes. We also denote that $Y = \{y_1, y_2, \cdots, y_l\} \in R^{d \times l}$ is the low-dimensional representation to X. In graph-based subspace learing framework, a similarity matrix is to defined for measuring the similarity of dataset. In detail, denote $G = (V, E)$ as the graph, where V is the vertex set of G representing the training set, E is the edge set of G related with W involving the geometrical information. In addition, let $L = D - W$ and $\widetilde{L} = D^{-1/2} L D^{-1/2}$ be the graph and normalized graph Laplacian matrix, which is to approximate the geometrical structure of dataset, D is a diagonal matrix satisfying $D_{jj} = \sum_{i=1}^{l} W_{ij}$.

2.2 Review of LE, LPP and SR

The goal of LE is to calculate the low-dimensional representation as follows:

$$\min_Y Tr \left(\left(Y D Y^T \right)^{-1} Y L Y^T \right). \tag{1}$$

where the optimal result for Y can be formed by the top p eigenvectors of $D^{-1}L$ or $D^{-1/2}LD^{-1/2} = \widetilde{L}$. However, LE cannot calculate the subspace graph embedding for new-coming data points hence cannot handle out-of-sample problem. To solve this problem, LPP further assume that the low-dimensional representation can be projected by high-dimensional data, i.e. $Y = V^T X$, where V is the projection matrix. it is then to calculate the projected matrix following the objective function as:

$$\min_V Tr \left(\left(V^T \left(X D X^T + \alpha I \right) V \right)^{-1} V^T X L X^T V \right). \tag{2}$$

where I is an identity matrix to avoid ill-posed problem, α is a small value. The optimal solution of V can be calculated by solving generalized eigenvalue decomposition (GEVD) of $\left(X D X^T + \alpha I \right)^{-1} X L X^T$. Then, the out-of-sample problem can be naturally handled by projecting the new-coming data points to the low-dimensional graph embedding by the projection matrix. However, the

computational cost for LPP is huge given that XDX^T and XLX^T are dense matrixes. To solve this problem, SR is developed, which is first to calculate the low-dimensional representation Y following (1), it then calculate the projection matrix via solving a regression problem as follows:

$$\min_{V} \left\| V^T X - Y \right\|_F^2 + \alpha \left\| V \right\|_F^2 . \tag{3}$$

The optimal solution of SR is $\left(XX^T + \alpha I\right)^{-1} X^T Y$, which can be efficiently solved by some well-studies methods, such as LSQR. The SR is more efficient than LPP it only needs to calculate the GEVD of sparser matrix $D^{-1}L$. However, SR is not equivalent to LPP. In addition, both SR and LPP cannot handle large-scale datasets. As pointed in [26], given the degree of similar matrix W equivalent to 1, i.e. $D = I$, so that the graph Laplacian matrix is equivalent to normalized one $L = I - W = \tilde{L}$, the optimal solution of SR can be equivalent to LPP.

2.3 Unsupervised large graph embedding (ULGE) with anchor graph construction

Though the conventional graph embedding methods have achieved satisfied results, they cannot be extended to large-scale dataset due to the computational complexites for both graph construction and graph embedding are not linear with the number of data points. To handle this problem, ULGE has developed an efficient SR method for graph embedding, which is based anchor graph construction [37].

In general, anchor graph first seeks m anchors, where $m \ll n$, and then calculate the weight matrix based on the anchor data and each data point. In detail, let $A = \{a_1, a_2, \ldots a_m\} \in R^{d \times m}$ represents the set of anchor points, $Z \in R^{m \times n}$ represents the adjancy matrix. Each element Z_{ij} is to evaluate the similarity between x_j and a_i with constraint $Z_{ij} \geq 0$, $\sum_{i=1}^{m} Z_{ij} = 1$. Then, the anchor graph W is contructed based Z, which can be shown as follows:

$$W = Z^T \Delta^{-1} Z = S^T S \in R^{n \times n}. \tag{4}$$

where $\Delta \in R^{m \times m}$ represents a diagonal matrix satisfying $\Delta_{ii} = \sum_{j=1}^{n} Z_{ij}$, $S = \Delta^{-1/2} Z \in R^{m \times n}$ is the bilinear decomposition of W. It can be easily proved that W^a is doubly-stochastic hence it has probability meaning. Compared with kNN graph, which needs $O(n^2 k)$ to contruct the graph, the anchor graph construction is more efficient since it only needs $O(m^3 + nm^2)$ computational complexity.

Finally, The ULGE is to formulate the graph embedding Y by the eigenvectors corresponding to the largest eigenvalues of S^T and to calculate the projection matrix by solving a regression problems as (3). Since $S \in R^{m \times n}$, the computational complexity for calculating the GEVD of S is linear with n and much smaller than the one for directly calculating the GEVD of $\tilde{L}$.

Notes: Cai et al. [27, 38] recently developed a graph embedding method for subspace clustering, namely, **L**arge-scale **S**pectral **C**lustering (SLC), which is also based on anchor graph construction and share similar concept with ULGE. However, SLC is only to formulate the graph embedding Y by the eigenvectors corresponding to the largest eigenvalues of S^T and then perform conventional k-means on Y for calculating the cluster labels for the dataset. Therefore, compared with ULGE, SLC cannot obtain the graph embedding for new-coming dataset so that it cannot handle out-of-sample problem. In other word, ULGE can be viewed as an extension work to SLC.

# 3 Semi-supervised adaptive graph embedding (SAGE)

3.1 Motivation and problem formulations

The above graph construction methods are good to grasp the geometrical information of dataset [26]. However, one problem is that they are all unsupervised and do not utilize the discriminative information provided by the class labels to enhance the clustering performance. A case in point is that if two data points x_i and x_j are in different cluster structure, they should not share any comment anchors so that $s_i^T s_j = 0$. An illustration can be shown in Fig. 1.

To solve the above problem, it is natural to integrate side information into the graph construction so that the samples with different clusters will has no link. Here, we induce a matrix $T \in R^{n \times n}$ in this work, where $T_{ij} = 1$ if x_i and x_j are not connected in the graph; $T_{ij} = 0$, otherwise. Then, we can constrain graph weight matrix W with $W_{ij} T_{ij} = 0$ for optimization. As can be seen in the simulation, more and more link is disconnected so that more clear clustering structure can be observed.

Another problem is that the graph construction and subspace learning are separated. The two stages do not share any information to enhance the performance of graph embedding. By this end, we then propose the proposed method by simutaneously calculating the optimal graph

Fig. 1 Motivations: the conventional anchor graph construction is unsupervised and do not utilize the discriminative information provided by the class labels. This will cause the case that nearby data points belonging to different classes will share the common anchors hance causing mis-connections between them. An ideal case should guarantee there is no connection between any pairwise data points that belong to different classes

(a) Anchor graph construction with mis-connections

(b) Anchor graph construction with ideal case
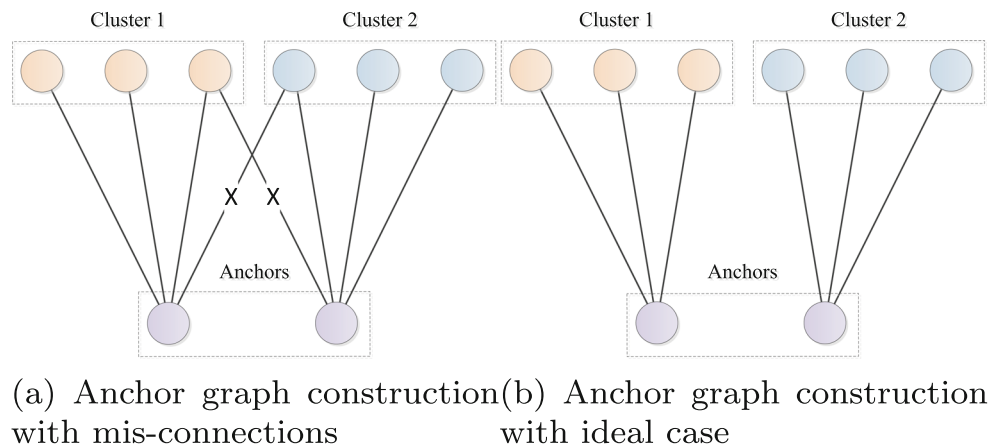
matrix W, the bilinear decomposition S and the graph embedding Y as follows:

$$W, S, Y = \arg_{W,S,Y} \min \left\| W - S^T S \right\|_F^2 + \gamma Tr\left(Y L Y^T\right)$$

$$s.t. \ S \geq 0, \ L = I - S^T S$$
$$W \geq 0, \ W = W^T, \ W 1^T = 1^T,$$
$$\sum_{ij} (T \odot W)_{ij} = Tr\,(TW) = 0 \tag{5}$$

where γ is a parameter balancing the tradeoff between two terms. We further add the constraint of degree normalization, i.e. $D_{ii} = 1$, i.e. $W 1^T = 1^T$, where $1 \in R^{1 \times n}$ is a one vector. Since W is both non-negative and symmetric, imposing $W 1^T = 1^T$ can make it doubly-stochastic [39] and less sensitive to the parameters. From (5), we can see that we have unified the graph construction and subspace learning into the same object function. Then, the subspace results can be used to enhance the geometrical preserving ability of the graph construction, and the modified graph can in turn incorporate discriminative information for graph embedding.

3.2 Solution

We next show how to solve the optimal solution of W, S and Y in (5), we use the alternative optimization approach to handle the problem. In detail, we first let S_t, W_t and Y_t be the t-th iteration of S, W and Y. To calculate graph embedding Y_{t+1}, we first fix W_t and S_t. Then (5) will be back to the conventional graph based subspace learning method, i.e.

$$Y_{t+1} = \arg_Y \min Tr\left(Y L Y^T\right) = \arg_Y \max Tr\left(Y S_t^T S_t Y^T\right)$$

$$s.t. \ Y Y^T = I, \tag{6}$$

The optimal Y_{t+1} is to be obtained by the largest eigenvectors of $S_t^T S_t$.

To calculate the updated S_{t+1}, we need to replace the objective function of (5) with its first order approximation. Then (5) can be rewritten as:

$$\left\| W_t - S^T S \right\|_F^2 + \gamma Tr\left(S W_t^Y S^T\right)$$
$$= \frac{\eta}{2} \|S - S_t\|_F^2 + \langle S - S_t, \nabla_S Q\,(S_t)\rangle + Q\,(S_t) \tag{7}$$
$$= \frac{\eta}{2} \left\| S - S_t - \frac{1}{\eta} \nabla_S Q\,(S_t) \right\|_F^2 + P\,(S_t)$$

where $W_t^Y \in R^{n \times n}$ with each element satisfying $W_t^Y|_{ij} = ||y_i - y_j||_F^2$, so that $Tr(Y L Y^T) = \sum_{ij} ||y_i - y_j||_F^2 S^T S = Tr(S W_t^Y S^T)$, $Q\,(S) = \left\| W_t - S^T S \right\|_F^2 + \gamma Tr\left(S W_t^Y S^T\right)$ and $\nabla_S Q\,(S_t) = S_t S_t^T S_t - S_t\left(W_t - \gamma W_t^Y\right)$ is the partially differential of $Q\,(S)$ w.r.t. S at S_t, η is the Lipschitz parameter and we choose it as the trace of second differential of $Q\,(S)$, i.e. $\eta = \left\| S^T S \right\|_F^2$, $P\,(S_t)$ is a fixed term with S_t that can be neglected. The updated S_{t+1} is equivalent to solve:

$$S_{t+1} = \arg\min_S \frac{\eta}{2} \left\| S - S_t - \frac{1}{\eta} \nabla_S Q\,(S_t) \right\|_F^2 \quad s.t. \ S \geq 0 \tag{8}$$

We next fix Y_t and S_t and update W_{t+1}. Then (5) falls into an instance of quadratic programming (QP), which can be shown as follows

$$W_{t+1} = \arg_W \min \left\| W - W_t^0 \right\|_F^2$$
$$W \geq 0, \ W = W^T, \ W 1^T = 1^T, \tag{9}$$
$$\sum_{ij} (T \odot W)_{ij} = Tr\left(W^T T\right) = 0$$

where $W_t^0 = S_t^T S_t$. For efficient computation, we divide the QP problem into two convex sub-problems:

$$W_{t+1} = \arg_W \min \left\| W - W_t^0 \right\|_F^2 \quad s.t. \ W \geq 0. \tag{10}$$

and

$$W_{t+1} = \arg_W \min \left\| W - W_t^0 \right\|_F^2$$
$$s.t.\ W = W^T,\ W1^T = 1^T,\ Tr\left(W^T T\right) = 0 \tag{11}$$

The W_{t+1} in (10) can be simply formulated by as the non-negative elements of W_t^0. On the other hand, (11) is solved to take the Lagrangian:

$$J(W) = \left\| W - W_t^0 \right\|_F^2 - t\, Tr\,(WT)$$
$$- \mu\left(W1^T - 1^T\right) - \mu\left(W^T 1^T - 1^T\right) \tag{12}$$

where t and $\mu \in R^{1 \times n}$ are the Lagrangian parameters. We then set the derivatives of $J(W)$ w.r.t. W to zero, i.e.

$$W = W_t^0 + t\,T + \mu^T 1 + 1^T \mu. \tag{13}$$

To fulfill $Tr\,(WT) = 0$, we have:

$$Tr\left(\left(W_t^0 + t\,T + \mu^T 1 + 1^T \mu\right) T\right) = 0$$
$$\Rightarrow Tr\left(W_t^0 T\right) + |T| t + 2\mu T 1^T = 0 \tag{14}$$
$$\Rightarrow t = \left(-Tr\left(W_t^0 T\right) - 2\mu T 1^T\right) \big/ |T|$$

Considering the normalized constraint $W1^T = 1^T$ or $1W = 1$, we have

$$\begin{cases} 1W1^T = n = 1W_t^0 1^T + |T| t + n\mu 1^T + n1\mu^T \\ 1WT1^T = |T| = 1W_t^0 T 1^T + 1T^2 1^T t + 1\mu^T |T| + n\mu T 1^T \\ W1^T = 1^T = W_t^0 1^T + t\,T1^T + n\mu^T + 1^T \mu 1^T \end{cases} \tag{15}$$

We then replace t in (15) with (14) and with some math derivations, we have

$$\begin{cases} \mu 1^T = 1\mu^T = \dfrac{1^T 1}{2n} R_t + \dfrac{1^T \mu T 1^T}{n} \\ \mu T 1^T = \dfrac{2n1\left(T - \frac{|T|}{2n}I\right) R_t}{2n^2 + 2|T| - 4n1T^2 1^T / |T|} \\ \mu^T = \dfrac{I}{n} R_t + \dfrac{2T1^T \mu T 1^T}{|T| n} - \dfrac{1^T \mu 1^T}{n} \end{cases}$$
$$\Rightarrow \mu^T = \left(\dfrac{I}{n} - \dfrac{1^T 1}{2n^2} - \dfrac{\left(T - \frac{|T|}{2n}I\right)1^T 1\left(T - \frac{|T|}{2n}I\right)}{n1T^2 1^T - n^2 |T|/2 - |T|^2/2}\right) R_t \tag{16}$$

where R is a fixed matrix as:

$$R_t = 1^T - W_t^0 1^T + \dfrac{Tr\left(W_t^0\right)}{|T|} T 1^T \tag{17}$$

The detail derivations of (16) from (15) can be seen in Appendix. Then by replacing μ in (13) with (16), we can obtain the updated W_{t+1}. This iterative procedure will be converged due to the Von-Neumann's successive projection lemma [36, 40, 41]. Therefore, we iteratively updated W,

S and Y following (6), (8) and (9), respectively, until the objective function in (5) is converged to a given small value. The detail algorithm for iteratively solving (5) can be seen in Algorithm 1.

Algorithm 1 Iterative solution for calculating the optimal solution of AWGE.

---

1 **Input**: Data matrix X.
2 Obtain the anchor data matrix A by k-means clustering and calculate similarity matrix Z between X and A based on Gaussian function;
3 Initialize the graph weighted matrix S and W according to (4);
4 **repeat**
5   Update Y_t as (6);
6   Update S_t as (8);
7   Update W_t as (11);
8 **until** $\left\| W_t - S_t^T S_t \right\|_F^2 + \gamma Tr\left(Y_t L Y_t^T\right)$ *does not change.*;
9 Calculate the projection matrix according to (3), i.e. $\min_V \left\| V^T X - Y_t \right\|_F^2 + \alpha \left\| V \right\|_F^2$.
10 **Output**: Projection Matrix V.

3.3 Computational complexity analysis

We next analyze the computational complexity of the proposed SAGE. Based on the basic steps in Algorithm 1, the computational complexity can be divided into three parts:

1) Obtain the m anchors by k-means. If we choose the Balanced and Hierarchical K-means (BAHK) methods as in [42], this will needs $O(nd\log(m)t)$, where t is the number of iterations;

2) Initialize the similarity matrix S_0 and formulate the original graph weight matrix W_0, where we need the computational complexity $O(ndm)$;

3) Update the low-dimensional representation Y, the similarity matrix S, and graph weight matrix W according to Algorithm 1. In detail, to update Y_{t+1} in each iteration, we needs to perform the SVD of matrix S_t, where the computational complexity is $O(nmk)$ and k is the number of nearest neighbors; to update S_{t+1}, we needs to calculate W_t^Y and $\delta_S Q(S_t)$. The computational complexity is $O(nd^2)$ and $O(nmd + nmq)$, respectively, where q is the average non-zero number of W_t or W_t^Y. Therefore the total complexity $O(nd^2 + nmd + nmq)$; to update W_{t+1}, we need to calculate the alternate projection procedure of (10) and (11), where the total complexity is $O(nqt_p)$ and t_p is the iteration number of the alternate projection procedure;

4) Calculate the projection matrix V by solving the regression problem of (3), where the computational complexity is $O(ndm)$.

Considering that $m \ll n$, the computational complexity of the proposed SAGE is $O(ndm)$. Compared with the conventional spectral based methods, such as LPP and SR, the proposed method is computationally efficient that can handle large-scale data.

4 Simulations

In this section, we will evaluate the proposed method based on one toy example and several real-world benchmark datasets, and compare the performance with other state-of-the-art methods. In toy example, we generalize a two-cycle dataset with two clusters, each follows a cycle distribution but with different radius; for real-world datasets, we choose three benchmark datasets, i.e. COIL100 dataset [43], CASIA-HWDB [44] and Yale-B dataset [45] and

Fashion-MNIST dataset [46] for evaluating the performance of graph embedding. To further show the effectiveness of the proposed work, we also utilize the proposed work for handling the task of content-based image retrieval. Our goal is to verify the distance calculated by the low-dimensional representation of datasets can well preserve the geometrical structure of data manifold, which is good to enhance some real-world applications.

4.1 Parameter analysis

We first verify the convergence of the proposed method. Specifically, we randomly initialize the similarity matrix S_0 and graph weight matrix W_0, we also fix the regularized parameter $\gamma = 10$ in (5) and the number of anchors as 500. For the parameter T, we randomly select 10% side information to form matrix T. We then show the convergence for the object function value in (5) based on CASIA-HWDB [44] dataset. Figure 2 shows the iterative procedure of the proposed method for updating the graph weight. From the simulation results, we can see that with

Fig. 2 Graph weight updated via iterative procedure

(a) original data

(b) 10 iterations

(c) 15 iterations
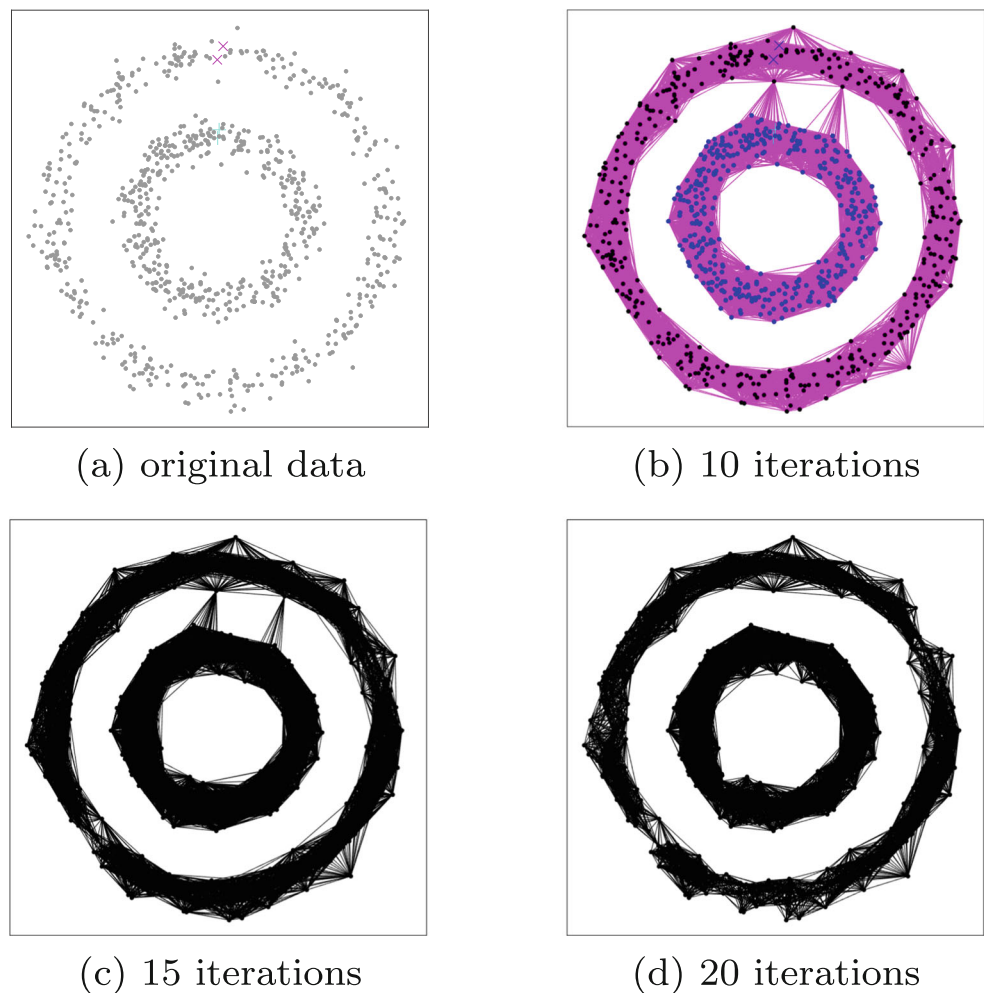
(d) 20 iterations

the iterations the missing link between different classes are graduate weakened until the structure of two classes are distinctively separated. This toy example well shows the effectiveness of the proposed method.

We next evaluate some important parameters of the proposed work. Here, it should be noted that there are three main parameters in our work including matrix T, anchor number and tradeoff parameter γ. To get the optimal values, we first fix matrix T, then set the candidates of anchor number as [500, 1000, 1500, 2000, 2500] and those of γ as $[10^{-3}, 10^{-2}, 10^{-1}, 1, 10, 10^2, 10^3]$, respectively. We then perform the proposed method and calculate the clustering accuracies based on certain value of T, anchor number and γ, through which the optimal values can be selected according to the best accuracies. Here, for the parameter T, 10% side information are randomly selected to form matrix T. The reason is that the proposed work mainly focuses on semi-supervised learning, which aims to utilize side information to enhance the clustering or ranking performance. If we choose 0% side information, it means the proposed work is totally an unsupervised learning method so that the parameter γ has no use as it is to balance the manifold term and side information regularization term; on the other hand, if we choose 20% side information, it means we use too much discriminative information. A good strategy is to use as fewer side information as possible and to achieve competitive results, so that we randomly select 10% side information to form matrix T. Figure 3 shows accuracies with varied number of anchors with range from 500 to 2500 and γ from 10^{-3} to 10^3. From Fig. 3, we can see that the results are satisfied and related stable when γ fall in the range of $[0.1, 10^3]$ and the number of anchors is larger than 1000.

4.2 Subspace embedding evaluations

In this section, we will evaluate the effectiveness of the proposed work based on one synthetic dataset and three benchmark datasets. In the synthetic dataset, we evaluate the proposed work based on a dataset with two classes. each one follows a cycle distribution. In this toy example, we will show how the proposed method can update the graph weight so that both geometrical structure and discriminative information can be perserved. In the real-world datasets, we evaluate the performance of subspace embedding based on COIL100 dataset [43], CASIA-HWDB [44], Yale-B dataset [45] and Fashion-MNIST [46].

COIL100 dataset [43] is an object image dataset where each object is viewed from varying angles at an interval of five degrees so that each object has 72 images. The original size of each cropped image is 128×128, with 24bit color levels per pixel. In our study, we down-sample
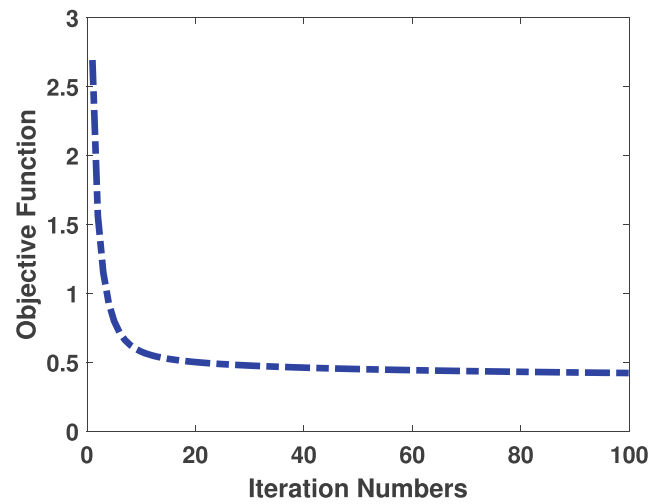
(a)
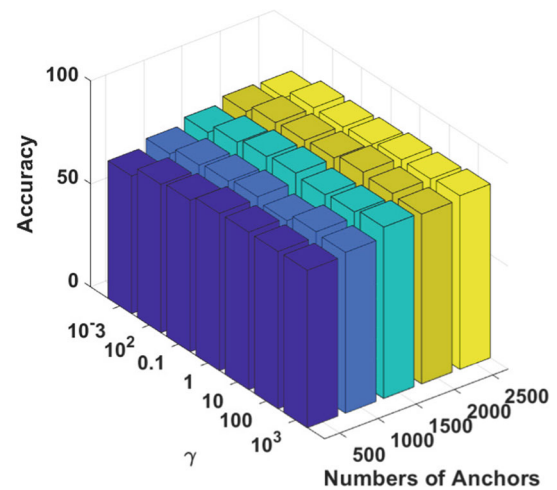
(b)

Fig. 3 Parameter analysis

the images to 32×32 and transfer them to 256 gray levels.

CASIA-HWDB dataset [44] is a handwritten image dataset which include both isolated characters (52 categories including 26 upper case letters and 26 lower case letters) and handwritten texts (10 categories including 0-9 digits). In our study, we choose a subset from it which includes images from 52 isolated characters. Then, the subset has 12456 samples with an image size of 16×16 in 256 gray levels.

Yale-B dataset [45] is a famous face dataset that contains 16128 images for 38 humans under 64 illumination conditions and 9 poses. In our study, we resize them to 32×32 pixels and choose approximately 64 near frontal images under different illuminations for each human.

Fashion-MNIST dataset [46] is a Zalando's clothes images dataset having 10 classes. It has 60000 and 10000 images for training set and test set, respectively. Each image is with 28×28 size in 256 gray level.

We first utilize one-cycle dataset to verify the effectiveness of the proposed method, where we randomly select 10% side information to form matrix T. Figure 2 shows the updated graph weight matrix during the iterative procedure in Algorithm 1, where the black line is the weight connection. From Fig. 2, we can observe that the graph weights between pairwise data points that belong to different clusters are gradually weakened, while those in the same cluster are strengthened. Therefore, the clustering structure of data manifold is much more distinctive. This illustrates the effectiveness of the proposed method.

We next evaluate the clustering results of the proposed method, where we use Clustering Accuracy to evaluate the performance for different methods [47]. The average results for ACC and NMI based on 20 random splits are given in Tables 1 and 2. Here, we evaluate the clustering results for a different number of clustering k. Tables 1 and 2 give the clustering results based on different datasets, where the

left columns are the results of LE, LPP, SR, ULGE and the unsupervised version of the proposed method, while the right two columns are the results of semi-supervised versions of SAGE with 10% and 20% side information. From simulation results, we can obtain the following observations:

1) the performance of ULGE and unsupervised version of SAGE outperform the conventional graph embedding methods such as LE, LPP and SR by approximately 4-5% improvement. The reason is that ULGE and the proposed method have utilized anchor graph so that the doubly-stochastic property can make the graph highly robust and less sensitive to the parameters;

2) the unsupervised version of SAGE is superior to other methods, by approximately 3% and 4% enhancement compared to ULGE and other methods. This is believable due to the reason that the proposed method has incorporated the graph construction and graph embedding learning into the unified framework;

3) by utilizing side information, the semi-supervised version of SAGE can achieve much better results than other state-of-the-art unsupervised methods. The improvement can achieve approximately 5% in COIL100 datasets given 20% side information. This

Table 1 Clustering performance based on different methods on four datasets (ACC)

Datasets	k	k-means	LE	LPP	SR	LSC	ULGE	SAGE/U	SAGE1	SAGE2
COIL100 [43]	20	56.5	56.8	63.3	63.6	65.9	68.1	**70.3**	72.7	75.2
	40	61.3	61.2	63.8	68.8	68.1	71.1	**72.8**	74.8	77.1
	60	62.5	66.0	70.6	73.1	72.7	75.1	**76.9**	79.5	81.6
	80	69.4	68.1	73.5	75.3	75.0	77.3	**77.9**	80.8	83.7
	100	70.8	72.2	76.8	77.0	78.7	80.8	**83.0**	85.3	87.7
	Avg.	64.1	64.8	69.6	71.5	72.0	74.4	**76.1**	78.6	81.0
CASIA-HWDB [44]	2	62.2	59.6	69.0	71.0	69.8	**72.5**	72.2	74.6	77.4
	4	65.7	69.1	71.1	71.4	72.7	**75.1**	74.6	77.0	79.6
	6	66.9	70.7	73.9	74.2	73.6	**75.9**	75.8	77.9	80.4
	8	72.2	76.3	81.0	82.3	78.0	80.3	**84.3**	87.0	89.2
	10	90.1	89.7	92.0	90.8	90.9	**93.3**	93.1	95.8	98.5
	Avg.	71.4	73.0	77.4	77.9	77.0	79.4	**80.0**	82.4	85.0
Yale-B [45]	10	13.4	48.3	55.8	55.5	54.2	57.2	**57.6**	60.2	63.0
	15	14.8	49.7	57.1	56.4	56.3	59.1	**61.3**	63.9	66.0
	20	22.3	50.3	57.7	57.0	58.1	61.1	**61.7**	63.9	66.1
	25	24.5	50.8	59.1	57.2	59.0	62.0	**63.8**	66.5	68.6
	30	31.2	57.2	65.6	67.5	67.9	70.2	**71.4**	74.1	76.6
	Avg.	21.2	51.2	59.0	58.7	59.1	61.9	**63.1**	65.7	68.0
Fashion-MNIST [46]	2	51.5	28.7	51.5	**54.0**	53.5	48.7	48.7	50.2	52.0
	4	55.7	30.7	56.2	**56.7**	56.2	52.7	52.7	54.2	56.1
	6	54.7	31.7	55.0	59.7	**62.7**	62.2	62.0	63.5	65.3
	8	63.7	29.5	63.0	55.2	55.5	64.5	**65.2**	66.7	68.5
	10	63.0	35.2	62.7	56.9	60.7	67.7	**68.2**	69.7	71.5
	Avg.	57.7	31.1	57.6	56.5	57.7	59.1	**59.3**	60.8	62.7

The bold entries represent the best accuracy result among different unsupervised methods

Table 2 Clustering performance based on different methods on four datasets (NMI)

Datasets	k	k-means	LE	LPP	SR	LSC	ULGE	SAGE/U	SAGE1	SAGE2
COIL100 [43]	20	54.3	53.8	60.9	60.6	63.3	65.3	**67.8**	70.5	72.4
	40	58.5	58.2	61.5	66.1	65.1	68.5	**70.0**	72.5	74.2
	60	60.4	63.5	68.0	70.3	69.9	72.1	**74.6**	77.1	78.8
	80	66.7	65.8	70.5	73.1	72.1	74.4	**75.0**	78.1	81.1
	100	67.9	70.1	74.0	74.4	75.8	78.7	**80.6**	83.1	85.0
	Avg.	61.5	62.2	66.9	68.9	69.2	71.8	**73.6**	76.2	78.3
CASIA-HWDB [44]	2	59.8	57.1	66.1	68.1	67.2	**70.0**	69.5	71.9	74.8
	4	63.1	66.5	68.9	68.6	69.7	72.1	**72.2**	74.9	76.7
	6	64.2	67.7	71.4	71.3	71.5	73.1	**73.5**	75.7	78.3
	8	70.0	73.6	78.9	79.4	75.9	77.6	**81.3**	84.5	87.1
	10	87.6	86.8	89.0	88.6	88.3	**91.1**	91.0	93.4	95.9
	Avg.	68.9	70.3	74.8	75.2	74.5	76.7	**77.5**	80.0	82.5
Yale-B [45]	10	10.6	45.8	53.7	52.9	51.6	**55.1**	54.6	57.9	60.7
	15	11.9	46.7	54.8	54.3	53.9	56.1	**58.5**	61.0	63.9
	20	19.5	47.5	55.1	54.5	55.7	58.9	**59.0**	61.5	63.9
	25	21.8	48.6	56.6	54.7	56.3	59.9	**61.4**	63.9	65.9
	30	28.6	54.2	62.8	64.7	65.2	67.9	**68.5**	71.8	74.2
	Avg.	18.4	48.5	56.6	56.2	56.5	59.5	**60.4**	63.2	65.7
Fashion-MNIST [46]	2	50	21.3	49.4	**54.6**	54.2	46.2	46.2	46.4	47.9
	4	54.4	23.1	53.8	**57.4**	57	51.2	51.2	51.4	52.9
	6	52.7	27.9	53.4	59.5	**63.2**	58	58.1	58.3	59.8
	8	60	23.4	58.6	55.6	56.3	62.4	**62.6**	62.8	64.3
	10	62	32.4	61.4	60	62.6	63.1	**63.8**	64.0	65.5
	Avg.	55.8	25.6	55.3	57.4	**58.6**	56.1	56.3	56.6	58.1

The bold entries represent the best normalized mutual information result among different unsupervised methods

indicates that the discriminative information is indeed good to improve the subspace clustering results.

4.3 Applications for person re-identification

In this section, we utilize the proposed method on Person Re-Identification [48, 49]. In detail, given a query person provided, we first calculate its low-dimensional representation. We then calculate the Euclidean Distance between the query person image and database person images via the low-dimensional subspace. Finally, the top person images in database with smallest distance to the query person image are chosen as the relevant results. The original query data combined with relevant data will be formulated as a new set of query data for another round retrieval procedure. The procedure will be repeated iteratively for several times after the retrieval performance is satisfied by the user. In this work, we select three benchmark Re-ID datasets, i.e. PKU [50], WARD (Wide Area Camera Network) [51] and RAiD (Re-Identification Across Indoor-Outdoor) Dataset [52] datasets for evaluations.

PKU-ReID dataset [50] is a Re-ID dataset. This dataset has 114 individuals including 1824 images captured from two disjoint camera views. For each person, eight images are captured from eight different orientations under one camera view and are normalized to pixels. In our study, we split it into two parts randomly, i.e., 57 individuals for training, and the other 57 ones for testing.

WARD dataset [51] is also a Re-ID dataset collected with three non-overlapping cameras. Each person has several images in each camera. As a result, a total of 4786 images of 70 different individuals has been extracted. The original size of images is with 320×240 pixels. In our work, we down-sample them to 96×48 pixels.

RAID dataset [52] is collected at the Winstun Chung Hall of UC Riverside. It is a 4-camera Re-ID dataset with 2 indoor and 2 outdoor cameras. A total of 43 peoples have walked into these camera views causing 6920 images. The size of each cropped image is originally 128×64 with 24bit color levels. In our work, we down-sample the size of images to 32×16 pixels.

4.3.1 Qualitative analysis

We first shown the qualitative analysis of the proposed method for Person Re-ID. Figure 4 shows the retrieval

Fig. 4 The retrieval results of the proposed work under different query data: the images in the first column in each subfigure are the query person images, the following ones are the top relevant feedback images that are closest to the query images
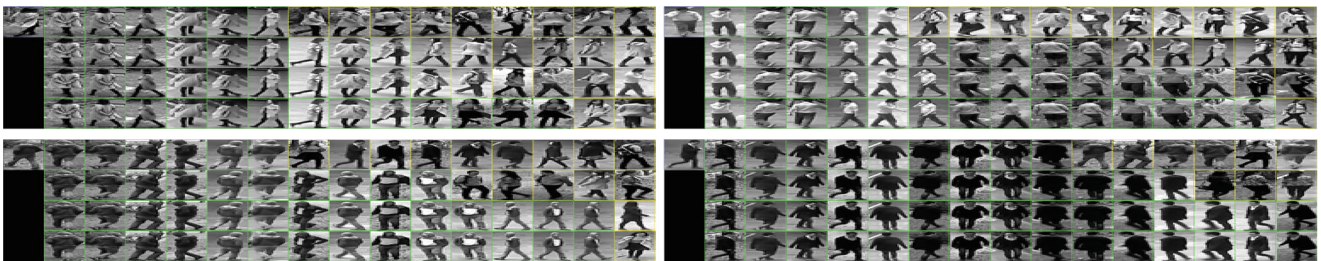

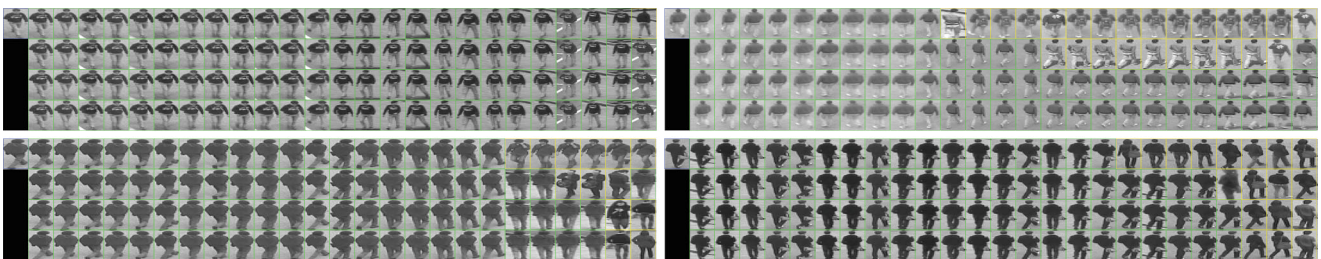(a) PKU dataset

(b) WARD dataset

(c) RAiD dataset

Fig. 5 The ranking results with different rounds of user revelance feedback: the first row to the fourth row are the ranking results associated with the zero, first, second and fourth round of user revelance feedback
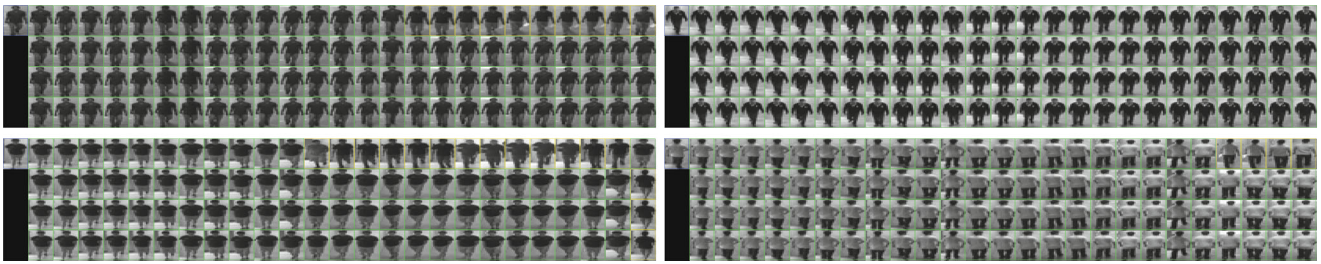

Fig. 6 The ranking results with different rounds of user revelance feedback: the first row to the fourth row are the ranking results associated with the zero, first, second and fourth round of user revelance feedback

Fig. 7 The ranking results with different rounds of user revelance feedback: the first row to the fourth row are the ranking results associated with the zero, first, second and fourth round of user revelance feedback

results of the proposed work under different query data for PKU [50], WARD [51] and RAiD [52] datasets. Simulation results can show that the proposed work can almost achieve

100% results given certain query data illustrating the effectiveness of the proposed work. Figures 5, 6, 7 show the image retrieval results by the proposed method with
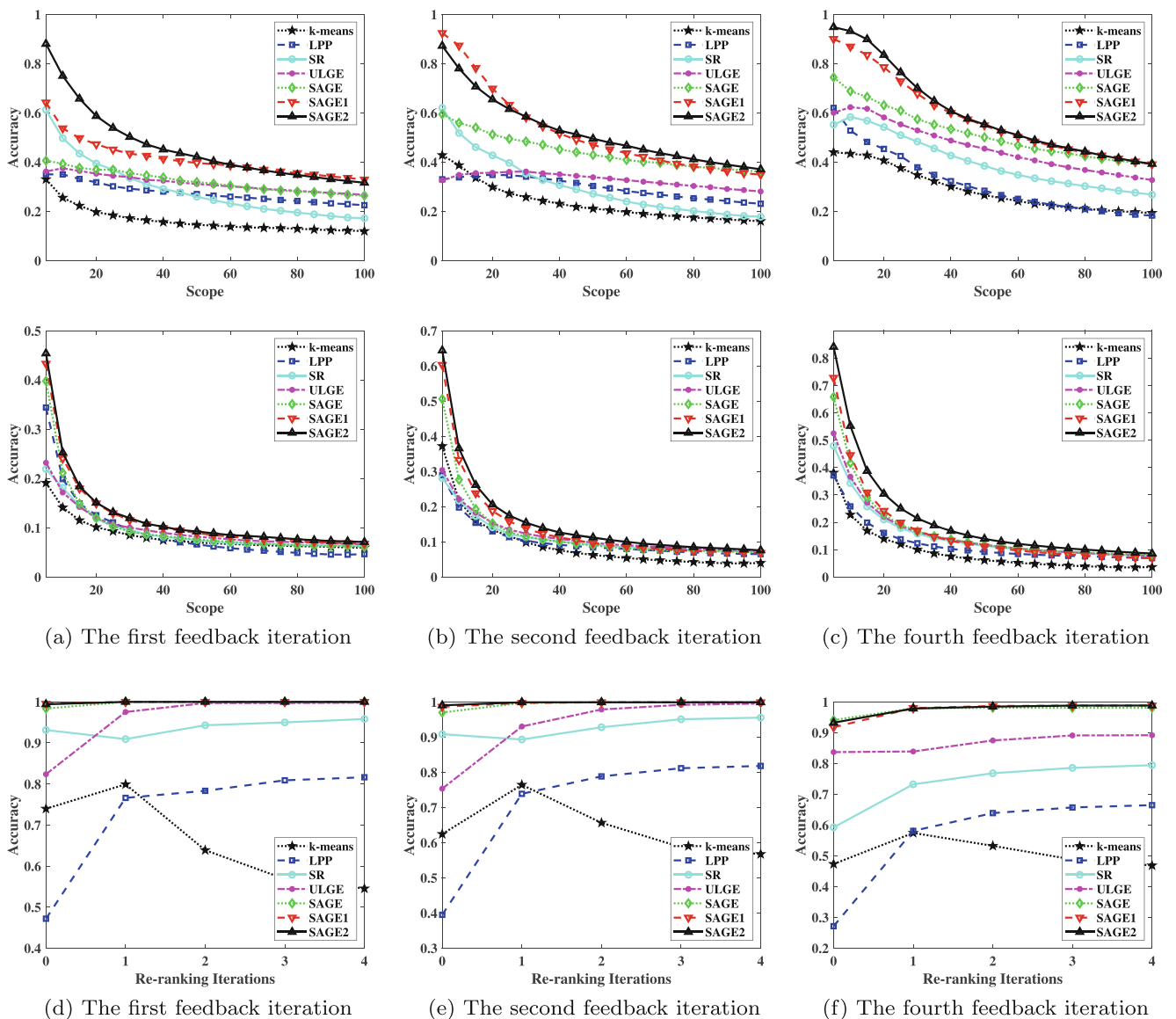

(a) The first feedback iteration

(b) The second feedback iteration

(c) The fourth feedback iteration

(d) The first feedback iteration

(e) The second feedback iteration

(f) The fourth feedback iteration

Fig. 8 MAP-scope curves based on all test set given the different feedback iterations: the up to the bottom row represent the results of PKU [50], WARD [51] and RAiD [52] datasets

10 scope. In this result, the images of the first column represent the queries and those in the following columns are the retrieval reslts with the top 15, 25 and 25 values under different iterations for PKU [50], WARD [51] and RAiD [52] datasets. The image with blue box in each group represents the query image randomly selected from WARD dataset, while those with green boxes represents the ones that are related to the query and those with yellow boxes represents the ones that are not related to the query. From Figs. 5, 6, 7, we can observe the retrieval performance given the queries are good, which can achieve almost up to 90% accuracies after the fourth iteration with user's provided revelant feedback. In addition, we can see that with the iteration increased, the retrieval results can be greatly enhanced showing that the user's revelant feedback information can provide useful supervised information that are good for the Person ReID.

4.3.2 Quantitative analysis

In this subsection, we compared the proposed method with other graph embedding methods for CBIR. The compared methods are the same as in data clustering evaluation. But we do not compare the performance with LE [22] and LSC [38] as they cannot calculate the low-dimensional representations of query data as it handle out-of-sample data. In our study, we use **MAP**-scope to evaluate the performance of the proposed work and to compare with
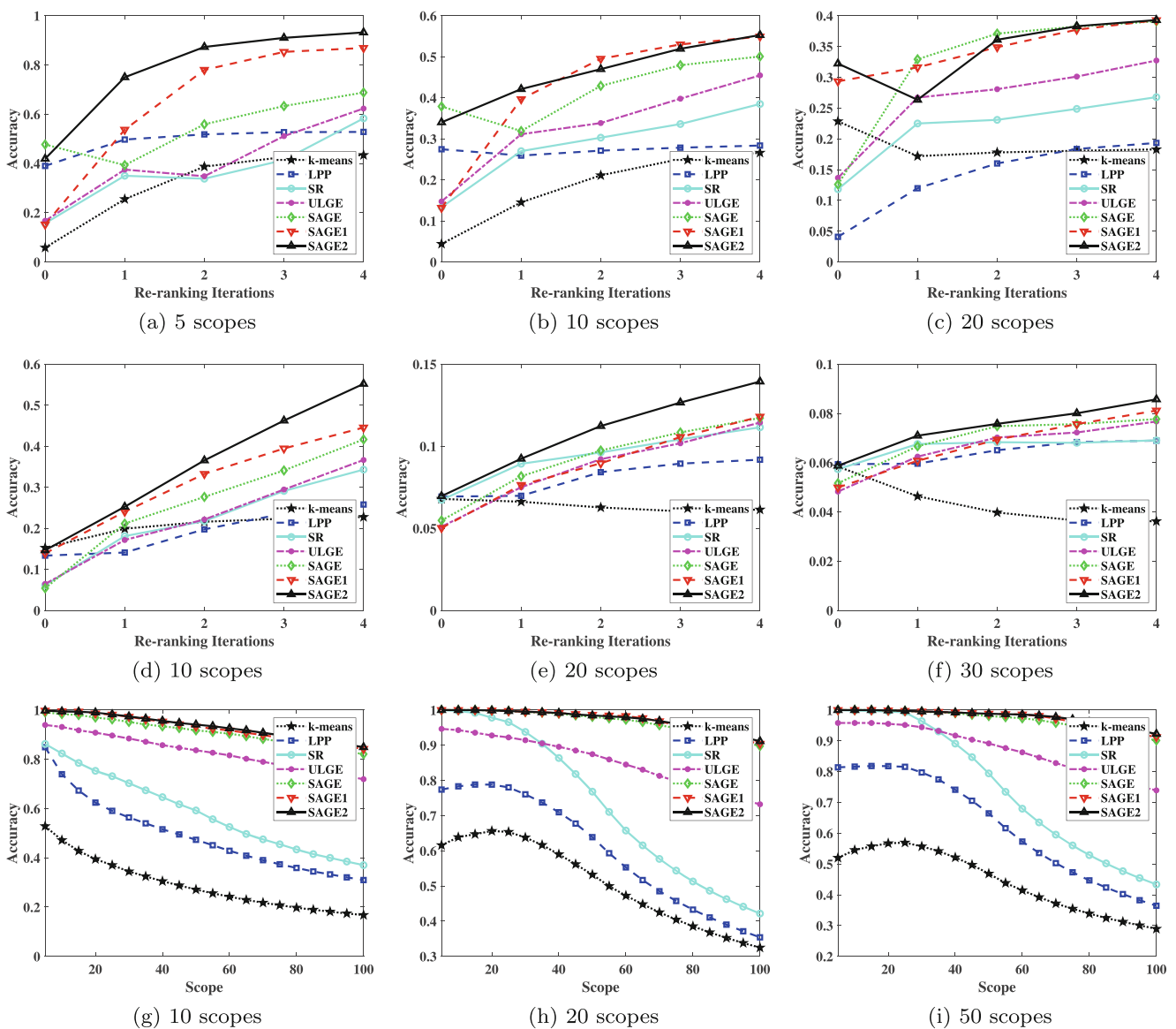

Fig. 9 MAP-scope curves based on all test set at Scope 5, 10 and 20 for PKU [50] dataset; Scope 10, 20 and 30 for WARD [51] and Scope 10, 20 and 50 for RAID [52] datasets

those of other state-of-the-art methods. In detail, the scope is the number of top-ranked images given to a certain query, while **M**ean **A**verage **P**recision (MAP) represents the average ratios of the numbers of relevant data to certain scope. Therefore, it evaluates the accuracy under different scopes for certain method.

Figure 8 gives the mean of MAP-scope under the fixed iteration number 1, 2 and 4, respectively. In addition, Fig. 9 gives MAP with increased iterations but with certain scope of 4, 7, 10 for PKU dataset, and 10, 20, and 50 for WARD and RAiD datasets respectively. From results in Figs. 8 and 9, we can see that the MAPs for all methods will become better given the user feedback is increased. This indicates that user feedback is indeed helpful to increase the discriminative information for handling CBIR problem. For example, the proposed method in Fig. 9 can achieve 15% enhancement with the fourth iteration than those without iteration in 20 scope of RAiD dataset. In addition, the proposed method can obtain better retrieval results over almost all scope and iterations than other compared methods.

5 Conclusion

In this paper, we in this paper develop a new scalable manifold ranking method for Person ReID by incorporating both graph weight construction and manifold regularized term in the same framework. The graph we developed is discriminative and doubly-stochastic. This can make the graph highly robust and insensitive to the dataset and parameters. In addition, by adopting the side information into consideration, the proposed graph can guarantee the data points with different clusters will not be mis-connected so that it can enhance the classification performances. Meriting from such a graph, we then incorporate the graph construction, the subspace learning method in the unified loss term. Therefore, the subspace results can be used to enhance the geometrical preserving ability of graph construction, and the modified graph can in turn incorporate discriminative information for graph embedding. Simulation indicates that the proposed method can achieve superor Person Re-ID performance.

While the proposed work can achieve satisfied results, our future work will focus on several issues: first, graph neural network (GNN) is one of the most popular topics during the past few years, which has been proved to be effective for dealing with structured data. We can consider to extend the proposed work to form a graph convolutional layer so that it can merit from the strong ability of CNN for grasp both the geometrical and data attribute features;

secondly, we can also consider to connect with CNN network to form an end-to-end deep learning framework, which can adopt the powerful feature extraction ability of CNN as well as maintain the geometrical information via GCN. This is also of great significance to improve the performance of image retrieval.

Appendix

We in this Appendix derive (16) from (15). We first rewrite (16) as follows:

$$
\begin{cases}
n = 1 W_t^0 1^T + |T| \, t + n\mu 1^T + n1\mu^T \\
|T| = 1 W_t^0 T 1^T + 1 T^2 1^T t + 1\mu^T |T| + n\mu T 1^T \\
1^T = W_t^0 1^T + t \, T 1^T + n\mu^T + 1^T \mu 1^T
\end{cases}
\tag{18}
$$

Noted that in (14), $t = \left(-Tr\left(W_t^0 T\right) - 2\mu T 1^T\right)/|T|$. By replacing it in (15), we have

$$
\begin{cases}
n = 1 W_t^0 1^T - Tr\left(W_t^0 T\right) - 2\mu T 1^T + 2n\mu 1^T \\
|T| = 1 W_t^0 T 1^T - \dfrac{Tr\left(W_t^0 T\right)\left(1 T^2 1^T\right)}{|T|} \\
\qquad - \dfrac{2\mu T 1^T \left(1 T^2 1^T\right)}{|T|} + 1\mu^T |T| + n\mu T 1^T \\
1^T = W_t^0 1^T - \dfrac{Tr\left(W_t^0 T\right) T 1^T}{|T|} - \dfrac{2 T 1^T \mu T 1^T}{|T|} \\
\qquad + n\mu^T + 1^T \mu 1^T
\end{cases}
\tag{19}
$$

Then, we some math derivation, we have:

$$
\begin{cases}
\mu 1^T = \dfrac{n - 1 W_t^0 1^T + Tr\left(W_t^0 T\right)}{2n} + \dfrac{\mu T 1^T}{n} \\
\mu T 1^T = \dfrac{n|T| - 2n\left(1 W_t^0 T 1^T\right) + |T|\left(1 W_t^0 1^T\right)}{2n^2 + 2|T| - 4n 1 T^2 1^T / |T|} \\
\qquad + \dfrac{\left(\dfrac{2n 1 T^2 1^T}{|T|} - |T|\right) Tr\left(W_t^0 T\right)}{2n^2 + 2|T| - 4n 1 T^2 1^T / |T|} \\
\mu^T = \dfrac{I}{n}\left(1^T - W_t^0 1^T + \dfrac{Tr\left(W_t^0 T\right)}{|T|} T 1^T\right) \\
\qquad + \dfrac{2 T 1^T \mu T 1^T}{|T| n} - \dfrac{1^T \mu 1^T}{n}
\end{cases}
\tag{20}
$$

By multiply 1^T into both sides of the first and second equations, and replacing the derivation results $1^T \mu 1^T$ and $1^T \mu T 1^T$ into the third equation, we have:

$$
\begin{cases}
1^T \mu 1^T = \dfrac{1^T 1}{2n} \left(1^T - W_t^0 1^T + \dfrac{Tr\left(W_t^0 T\right)}{|T|} T 1^T \right) \\
\qquad\quad + \dfrac{1^T \mu T 1^T}{n} \\
1^T \mu T 1^T = \dfrac{1^T 1 \left(2nT - |T| I \right)}{2n^2 + 2|T| - 4n1T^2 1^T / |T|} \\
\qquad\quad \left(1 - 1 W_t^0 + 1 T \dfrac{Tr\left(W_t^0 T\right)}{|T|} \right) \\
\mu = \left(1 - 1 W_t^0 + 1 T \dfrac{Tr\left(W_t^0 T\right)}{|T|} \right) \left(\dfrac{I}{n} - \dfrac{1^T 1}{2n^2} \right) \\
\qquad + 1 T \mu 1 \dfrac{2}{n|T|} \left(T - \dfrac{|T|}{2n} I \right)
\end{cases}
\tag{21}
$$

Finally, we prove:

$$
\mu = \left(1 - 1 W_t^0 + 1 T \frac{Tr\left(W_t^0 T\right)}{|T|} \right)
$$
$$
\left(\frac{I}{n} - \frac{1^T 1}{2n^2} - \frac{\left(T - \frac{|T|}{2n} I \right) 1^T 1 \left(T - \frac{|T|}{2n} I \right)}{n 1 T^2 1^T - n^2 |T| / 2 - |T|^2 / 2} \right)
\tag{22}
$$

where the final equations hold as in (23) and (24):

$$
1^T \mu 1^T
$$
$$
= 1^T \frac{n - 1 W_t^0 1^T + Tr\left(W_t^0\right)}{2n} + \frac{1^T \mu T 1^T}{n}
$$
$$
= \frac{1^T 1 1^T}{2n} - \frac{1^T 1}{2n} W_t^0 1^T + \frac{1^T 1 T 1^T}{2n|T|} Tr\left(W_t^0\right) + \frac{1^T \mu T 1^T}{n}
$$
$$
= \frac{1^T 1}{2n} \left(1^T - W_t^0 1^T + \frac{Tr\left(W_t^0\right) T 1^T}{|T|} \right) + \frac{1^T \mu T 1^T}{n}
\tag{23}
$$

$$
1^T \mu T 1^T
$$
$$
= \frac{1^T \left(2n|T| - 2n \left(1 W_t^0 T 1^T \right) + \frac{2n 1 T^2 1^T}{|T|} Tr\left(W_t^0 T\right) \right)}{2n^2 + 2|T| - 4n 1 T^2 1^T / |T|}
$$
$$
- \frac{1^T \left(n|T| - |T| \left(1 W_t^0 1^T \right) + |T| Tr\left(W_t^0 T\right) \right)}{2n^2 + 2|T| - 4n 1 T^2 1^T / |T|}
$$
$$
= \frac{2n \left(1^T 1 T 1^T - 1^T 1 T W_t^0 1^T + 1^T 1 T \frac{Tr\left(W_t^0 T\right) T 1^T}{|T|} \right)}{2n^2 + 2|T| - 4n 1 T^2 1^T / |T|}
$$
$$
- \frac{|T| \left(1^T 1 1^T - 1^T 1 W_t^0 1^T + 1^T 1 \frac{Tr\left(W_t^0 T\right) T 1^T}{|T|} \right)}{2n^2 + 2|T| - 4n 1 T^2 1^T / |T|}
$$

$$
= \frac{2n 1^T 1 T \left(1^T - W_t^0 1^T + \frac{Tr\left(W_t^0 T\right)}{|T|} T 1^T \right)}{2n^2 + 2|T| - 4n 1 T^2 1^T / |T|}
$$
$$
- \frac{1^T 1 |T| \left(1^T - W_t^0 1^T + \frac{Tr\left(W_t^0 T\right)}{|T|} T 1^T \right)}{2n^2 + 2|T| - 4n 1 T^2 1^T / |T|}
$$
$$
= \frac{2n 1^T 1 \left(T - \frac{|T|}{2n} I \right) \left(1^T - W_t^0 1^T + \frac{Tr\left(W_t^0 T\right)}{|T|} T 1^T \right)}{2n^2 + 2|T| - 4n 1 T^2 1^T / |T|}
\tag{24}
$$

Declarations

Conflict of Interests The authors declared that they have no conflicts of interest to this work.

References

1. Song W, Zheng J, Wu Y, Chen C, Liu F (2021) Discriminative feature extraction for video person re-identification via multi-task network. Appl Intell 51(2):788–803
2. Pang Z, Guo J, Sun W, Xiao Y, Yu M (2021) Cross-domain person re-identification by hybrid supervised and unsupervised learning. Applied Intelligence, pp 1–15
3. Su J, He X, Qing L, Cheng Y, Peng Y (2021) An enhanced siamese angular softmax network with dual joint-attention for person re-identification. Applied Intelligence, pp 1–19
4. Luo H, Jiang W, Gu Y, Liu F, Liao X, Lai S, Gu J (2019) A strong baseline and batch normalization neck for deep person re-identification. IEEE Trans Multimed 22(10):2597–2609

5. Cheng Z, Dong Q, Gong S, Zhu X (2020) Inter-task association critic for cross-resolution person re-identification. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 2605–2615

6. He L, Liao X, Liu W, Liu X, Cheng P, Mei T (2020) Fastreid: A pytorch toolbox for general instance re-identification. arXiv preprint arXiv:2006.02631

7. Zhou D, Weston J, Gretton A, Bousquet O, Schölkopf B. (2003) Ranking on data manifolds. In: NIPS, vol 3

8. Xu B, Bu J, Chen C, Wang C, Cai D, He X (2015) Emr: A scalable graph-based ranking model for content-based image retrieval. IEEE Trans Knowl Data Eng 27(1):102–114

9. Wang F, Zhang C, Shen HC, Wang J (2006) Semi-supervised classification using linear neighborhood propagation. In: 2006 IEEE computer society conference on Computer vision and pattern recognition, vol 1, IEEE, pp 160–167

10. Yang Y, Nie F, Xu D, Luo J, Zhuang Y, Pan Y (2012) A multimedia retrieval framework based on semi-supervised ranking and relevance feedback. IEEE Trans Pattern Anal Mach Intell 34(4):723–742

11. Zhao M, Zhang Z, Chow TW (2012) Trace ratio criterion based generalized discriminative learning for semi-supervised dimensionality reduction. Pattern Recogn 45(4):1482–1499

12. Zhao M, Chow TW, Wu Z, Zhang Z, Li B (2015) Learning from normalized local and global discriminative information for semi-supervised regression and dimensionality reduction. Inf Sci 324:286–309

13. Zhao M, Zhang Z, Chow TW, Li B (2014) A general soft label based linear discriminant analysis for semi-supervised dimensionality reduction. Neural Netw 55:83–97

14. Zhang Z, Li F, Zhao M, Zhang L, Yan S (2017) Robust neighborhood preserving projection by nuclear/l2, 1-norm regularization for image feature extraction. IEEE Trans Image Process 26(4):1607–1622

15. Gao Y, Ma J, Zhao M, Liu W, Yuille AL (2019) Nddr-cnn: Layerwise feature fusing in multi-task cnns by neural discriminative dimensionality reduction. In: Proceedings of the IEEE/CVF conference on computer vision and pattern recognition, pp 3205–3214

16. Zhang H, Zhang Z, Zhao M, Ye Q, Zhang M, Wang M (2020) Robust triple-matrix-recovery-based auto-weighted label propagation for classification. IEEE Trans Neural Netw Learn Syst 31(11):4538–4552

17. Zhang Z, Jia L, Zhao M, Liu G, Wang M, Yan S (2018) Kernel-induced label propagation by mapping for semi-supervised classification. IEEE Trans Big Data 5(2):148–165

18. Zhao M, Lin M, Chiu B, Zhang Z, Tang XS (2018) Trace ratio criterion based discriminative feature selection via l2, p-norm regularization for supervised learning. Neurocomputing 321:1–16

19. Zhang Z, Li F, Zhao M, Zhang L, Yan S (2016) Joint low-rank and sparse principal feature coding for enhanced robust representation and visual classification. IEEE Trans Image Process 25(6):2429–2443

20. Tenenbaum JB, De Silva V, Langford JC (2000) A global geometric framework for nonlinear dimensionality reduction. Science 290(5500):2319–2323

21. Roweis ST, Saul LK (2000) Nonlinear dimensionality reduction by locally linear embedding. Science 290(5500):2323–2326

22. Belkin M, Niyogi P (2002) Laplacian eigenmaps and spectral techniques for embedding and clustering. Adv Neural Inf Process Syst 14(6):585–591

23. Yu W, Teng X, Liu C (2006) Face recognition using discriminant locality preserving projections. Image Vis Comput 24(3):239–248

24. Deng C, He X, Han J (2007) Spectral regression: a unified subspace learning framework for content-based image retrieval. In: Proceedings of the 15th international conference on multimedia 2007, Augsburg, Germany, September 24-29, 2007

25. Cai D (2015) Compressed spectral regression for efficient nonlinear dimensionality reduction. In: Twenty-fourth international joint conference on artificial intelligence

26. Nie F, Zhu W, Li X (2017) Unsupervised large graph embedding. In: Thirty-first AAAI conference on artificial intelligence

27. Chen X, Cai D (2011) Large scale spectral clustering with landmark-based representation. In: Twenty-fifth AAAI conference on artificial intelligence

28. Li Y, Nie F, Huang H, Huang J (2015) Large-scale multi-view spectral clustering via bipartite graph. In: Twenty-ninth AAAI conference on artificial intelligence

29. Defferrard M, Bresson X, Vandergheynst P (2016) Convolutional neural networks on graphs with fast localized spectral filtering. arXiv preprint arXiv:1606.09375

30. Kipf TN, Welling M (2016) Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907

31. Hamilton WL, Ying R, Leskovec J (2017) Inductive representation learning on large graphs. arXiv preprint arXiv:1706.02216

32. Veličković P, Cucurull G, Casanova A, Romero A, Lio P (2017) Bengio, Y.: Graph attention networks. arXiv preprint arXiv:1710.10903

33. Lu S, Zhu Z, Gorriz JM, Wang SH, Zhang YD (2022) Nagnn: Classification of covid-19 based on neighboring aware representation from deep graph neural network. Int J Intell Syst 37(2):1572–1598

34. Wang SH, Govindaraj V, Gorriz JM, Zhang X, Zhang YD (2021) Explainable diagnosis of secondary pulmonary tuberculosis by graph rank-based average pooling neural network. Journal of Ambient Intelligence and Humanized Computing, pp 1–14

35. Zhao M, Zhang Y, Zhang Z, Liu J, Kong W (2019) Alg: Adaptive low-rank graph regularization for scalable semi-supervised and unsupervised learning. Neurocomputing 370:16–27

36. Liu W, Chang SF (2009) Robust multi-class transductive learning with graphs. In: 2009 IEEE Conference on computer vision and pattern recognition, IEEE, pp 381–388

37. Liu W, He J, Chang SF (2010) Large graph construction for scalable semi-supervised learning. In: Proceedings of the 27th international conference on machine learning (ICML-10), pp 679–686

38. Cai D, Chen X (2014) Large scale spectral clustering via landmark-based sparse representation. IEEE Trans Cybern 45(8):1669–1680

39. Zhou D, Bousquet O, Lal TN, Weston J, Schölkopf B (2004) Learning with local and global consistency. In: Advances in neural information processing systems, pp 321–328

40. Von Neumann J (1950) Functional operators: Measures and integrals, vol 1, Princeton University Press

41. Horn RA, Johnson CR (2012) Matrix analysis. Cambridge University Press

42. Nie F, Zhu W, Li X (2020) Unsupervised large graph embedding based on balanced and hierarchical k-means IEEE Transactions on Knowledge and Data Engineering

43. Nene SA, Nayar SK, Murase H (1996) Columbia object image library (coil-100). Tech. Rep. Technical Report CUCS-005-96 Columbia University

44. Liu CL, Yin F, Wang DH, Wang QF (2011) Casia online and offline chinese handwriting databases. In: 2011 International conference on document analysis and recognition, IEEE, pp 37–41

45. Lee K, Ho J, Kriegman D (2005) Acquiring linear subspaces for face recognition under variable lighting. IEEE Trans Pattern Anal Mach Intell 27(5):947–963

46. Xiao H, Rasul K, Vollgraf R (2017) Fashion-mnist: A novel image dataset for benchmarking machine learning algorithms

47. Fan J, Tian Z, Zhao M, Chow TW (2018) Accelerated low-rank representation for subspace clustering and semi-supervised classification on large-scale data. Neural Netw 100:39–48

48. Zhao M, Chow TW, Zhang Z, Li B (2015) Automatic image annotation via compact graph based semi-supervised learning. Knowl-Based Syst 76:148–165

49. Zhao M, Liu J, Zhang Z, Fan J (2021) A scalable sub-graph regularization for efficient content based image retrieval with long-term relevance feedback enhancement. Knowledge-Based Systems 212:106505

50. Ma L, Liu H, Hu L, Wang C, Sun Q (2016) Orientation driven bag of appearances for person re-identification. arXiv preprint arXiv:1605.02464

51. Martinel N, Micheloni C (2012) Re-identify people in wide area camera network. In: 2012 IEEE Computer society conference on computer vision and pattern recognition workshops. IEEE, Providence, RI, pp 31–36, https://doi.org/10.1109/CVPRW.2012.6239203

52. Das A, Chakraborty A, Roy-Chowdhury AK (2014) Consistent re-identification in a camera network. In: European conference on computer vision, lecture notes in computer science, vol 8690, Springer, pp 330–345

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Affiliations

Jiao Liu[1,2] · Mingquan Lin[2] · Mingbo Zhao[2,3] · Choujun Zhan[2,3] · Bing Li[4] · John Kwok Tai Chui[5]

Jiao Liu
liujiaolndx@gmail.com

Mingquan Lin
mqlin5-c@my.cityu.edu.hk

Bing Li
lib675@whut.edu.cn

John Kwok Tai Chui
jktchui@hkmu.edu.hk

[1] Shanghai University of Engineering Science, Shanghai, China

[2] City University of Hong Kong, Hong Kong, Hong Kong

[3] Nanfang College Guangzhou, Guangzhou, China

[4] Wuhan University of Technology, Wuhan, China

[5] Hong Kong Metropolitan University, Hong Kong, Hong Kong