Knowledge-Based Systems 241 (2022) 108250

Contents lists available at ScienceDirect

Knowledge-Based Systems

journal homepage: www.elsevier.com/locate/knosys

Consensus graph and spectral representation for one-step multi-view kernel based clustering

S. El Hajjar [b], F. Dornaika [a,b,c,*], F. Abdallah [d,e], N. Barrena [b]

[a] School of Computer and Information Engineering, Henan University, Kaifeng, China
[b] University of the Basque Country UPV/EHU, San Sebastian, Spain
[c] IKERBASQUE, Basque Foundation for Science, Bilbao, Spain
[d] Lebanese University, Beirut, Lebanon
[e] Luxembourg Institute of Socio-Economic Research (LISER), Esch-sur-Alzette, Luxembourg

ARTICLE INFO

Article history:
Received 29 April 2021
Received in revised form 17 January 2022
Accepted 19 January 2022
Available online 25 January 2022

Keywords:
Multi-view clustering
One-step clustering
Graph learning
Spectral representation
Nonnegative embedding
Automatic weighting
Clustering algorithms

ABSTRACT

Recently, multi-view clustering has received much attention in the fields of machine learning and pattern recognition. Spectral clustering for single and multiple views has been the common solution. Despite its good clustering performance, it has a major limitation: it requires an extra step of clustering. This extra step, which could be the famous k-means clustering, depends heavily on initialization, which may affect the quality of the clustering result. To overcome this problem, a new method called Multi-view Clustering via Consensus Graph Learning and Nonnegative Embedding (MVCGE) is presented in this paper. In the proposed approach, the consensus affinity matrix (graph matrix), consensus representation and cluster index matrix (nonnegative embedding) are learned simultaneously in a unified framework. Our proposed method takes as input the different kernel matrices corresponding to the different views. The proposed learning model integrates two interesting constraints: (i) the cluster indices should be as smooth as possible over the consensus graph and (ii) the cluster indices are set to be as close as possible to the graph convolution of the consensus representation. In this approach, no post-processing such as k-means or spectral rotation is required. Our approach is tested with real and synthetic datasets. The experiments performed show that the proposed method performs well compared to many state-of-the-art approaches.

1. Introduction

Clustering is one of the most important research topics in machine learning, which aims to group samples into different groups, called clusters, without knowing their labels [1,2]. In the last decades, many clustering approaches have been developed. In particular, multi-view clustering algorithms have been used and developed to obtain additional information to improve the final clustering [3–9]. Among these methods, Spectral Clustering (SC) [10–13] methods are the most popular approaches due to their well-defined mathematical framework and ease of implementation. After constructing the similarity matrix between the data points, these methods generate a nonlinear projection of the data by mapping the dataset into a space where clusters can be easily identified (spectral embedding). A major drawback of these methods is the use of a post-processing step such as k-means to obtain the final clustering result, which can be affected by the initialization or the presence of outliers. Matrix factorization methods [14,15] can be used for dimensionality reduction. For example, the method in [15] called Integration by Matrix Factorization (IMF). This method generates different representative clustering matrices computed independently for each view, generates an intermediate matrix for all views, and then performs a factorization process on this matrix to reconcile the different clustering matrices generated from the different views. Matrix factorization methods have a low computational cost compared to other methods. This is because factorization of a particular matrix decomposes it into its constituent parts, which can also simplify the matrix operations as they are applied to the obtained matrices and not to the original complex matrix. However, these methods cannot deal with the nonlinearity of the data. To deal with the problem of nonlinearity of data, several approaches have proposed a solution based on multiple kernels [7,16,17]. With these methods, the data is mapped into a space where it is linearly separable. However, it is noted that multi-view clustering still needs improvement. To address this issue, in this paper we present a novel approach that provides a consistent non-negative embedding matrix to determine the final cluster assignment. Our proposed method estimates the clustering of the data directly without any additional post-processing.

* Corresponding author at: University of the Basque Country UPV/EHU, San Sebastian, Spain.
E-mail address: fadi.dornaika@ehu.eus (F. Dornaika).

https://doi.org/10.1016/j.knosys.2022.108250

It enforces that the cluster index matrix is a kind of convolution of a unified spectral representation over a consistent graph. The method we propose is called multi-view clustering via consensus graph learning and nonnegative embedding (MVCGE). It can overcome some drawbacks of other approaches. The proposed method can simultaneously provide the consistent similarity graph, the non-negative cluster index matrix and the unified spectral projection matrix across all views. Moreover, this method automatically calculates the weight of each view without using any additional parameters. The proposed method combines the advantages of graph-based methods and multiple kernel methods. In other words, our method retains two interesting properties that the current methods NESE in [18] and MVCSK in [19] do not have simultaneously. The first property, inspired by NESE, is the non-dependence on a particular clustering algorithm such as k-means clustering. The other property (inspired by MVCSK) is the simultaneous estimation of a consistent unified graph and a unified spectral representation. Since our method combines the advantages of NESE and MVCSK, the main goal of our study is to outperform the previous two methods. Therefore, these two methods are used as the main competing methods in our study. The contributions of the paper are summarized below.

1. Unlike other approaches based on multilevel learning, our method can simultaneously provide the consensus similarity matrix, the nonnegative index cluster matrix, the spectral projection matrix, and the weight of each view automatically.
2. It generates the final clustering assignment directly without any post-processing step. Our method inherits the advantages of matrix factorization methods and graph based methods.
3. The proposed model successfully finds nonlinear interactions between different views. This method is able to compute the exact graph considering the underlying correlations from numerous views by using a kernel representation of each view.
4. The cluster index matrix, which is the consequence of the convolution of the coherent spectral projection matrix over the coherent graph, is learned as part of the proposed learning technique.
5. It has been validated on real and synthetic datasets. This validation shows that this approach can give better results compared to state-of-the-art clustering methods.

The rest of the paper is organized as follows. Section 2 introduces the main concepts and some related work. Section 3 describes our proposed method in detail. Section 4 reports experimental results obtained on real and synthetic datasets. Section 5 concludes the paper.

2. Preliminaries and related work

2.1. Notations

We consider a data matrix $\mathbf{X}^v$ with n data points as $(\mathbf{x}_1^v, \mathbf{x}_2^v, \ldots, \mathbf{x}_n^v) \in \mathbb{R}^{d^v \times n}$, where d^v is the number of features in the corresponding view v. We represent the matrices in bold uppercase letters, the vectors in bold lowercase letters, and the constants in non-bold letters. The trace of a matrix $\mathbf{M}$ is denoted by $Tr(\mathbf{M})$ and its transpose by $\mathbf{M}^T$. The Frobenius norm of $\mathbf{M}$ is given by $\|\mathbf{M}\|_F = \|\mathbf{M}\|_2 = \sqrt{\sum_{i=1}^{n}\sum_{j=1}^{d}|M_{ij}|^2}$. $\mathbf{I}$, $\mathbf{1}_n$, $\mathbf{D}$ and $\mathbf{L}$ are the identity matrix, the column vector with n elements equal to one, the diagonal matrix, and the Laplacian matrix of the graph, respectively. The number of clusters is denoted by K. The similarity matrix, the nonnegative embedding matrix, the spectral projection matrix, and the kernel matrix are denoted by $\mathbf{S}$, $\mathbf{H}$, $\mathbf{P}$ and $\mathbf{K}$, respectively.

2.2. Related work

Recently, several multiview clustering approaches have been proposed. The current approaches can be divided into several groups: Spectral clustering algorithms [12,13,20,21], Graph based clustering algorithms [22], Weighted multiview clustering approaches [10,21,23–25], Automatically weighted multiview clustering algorithms [19,26–28], Multiview subspace based clustering approaches [29,30], Kernel based Approaches [19], Matrix factorization approaches [15,18,31], Nonnegative matrix factorization methods [18,32], etc. In this section, we present several methods that belong to these categories. A popular category of approaches is Spectral clustering [12,13,20,21]. This method constructs a similarity graph between data points and then constructs a data representation matrix using the eigenvectors of the corresponding Laplacian of the graph. Therefore, a post-processing step is used to obtain the final clustering assignment. An example of a spectral clustering algorithm is the famous co-training clustering algorithm [12], which adjusts the similarity matrix of a given view based on the clustering result of another view so that the same instance is placed in the same cluster in different views.

Moreover, co-regulated spectral clustering [13] combines the similarity matrices obtained from different views with an adaptive scheme to obtain the result. weighted multiview clustering approaches [10,23–25] assign weights to each view to account for the contribution of each view to the final clustering result. Moreover, these approaches use a consistent scheme to merge the different views. A drawback of these approaches is the use of additional weighting parameters. To overcome this drawback, Automatically weighted multi-view clustering algorithms have been proposed [19,26–28]. Moreover, another approach related to the above categories is presented, which extends the spectral clustering algorithm with the idea of weighted views. This method is called Adaptive Weighted Procrustes (AWP) [21]. The final clustering assignment of this method is achieved by spectral rotation. Moreover, this method provides accurate clustering with low computational cost. Another famous category named Multi-view Subspace based Clustering approach (MVSC) was introduced in [29,30] to learn the best and consistent representation of the data. Moreover, a Multi-view Learning method with Adaptive Neighbors (MLAN) is proposed in [33] to jointly learn the similarity graph and perform the final clustering assignment.

The kernel-based approaches (e.g., [19]) are used to overcome the problem of nonlinearity of the data by mapping them to a space in which they are linearly separable, and then they solve the problem caused by the multiple shapes of the data. The matrix factorization approaches (e.g., [15,18,31]) are used because of their low computational cost, which makes them efficient for dimensionality reduction. They can provide high clustering performance compared to other methods. However, this category of clustering methods cannot handle nonlinear data.

In [32], the authors propose a Non-negative Matrix Factorization (NMF) approach that uses dual constraints. This approach exploits the labeling of some images and the sparsity of representations. In [34], the authors introduced a unified framework for Joint clustering and distance metric learning. These are solved via rank reduced regression. The approach provided some new insights for learning a clustering that adopts distance metric learning. In [35], the authors presented an Ensemble clustering method based on efficient propagation of clusterwise similarities via random walks. The work of [36] proposed a Locality Adaptive Latent MultiView Clustering (LALMVC) method. It simultaneously learns the latent consensus representation via linear transformations, the joint spectral representation and the consensus graph. The learned consensus graph matrix is then used in spectral clustering to obtain a cluster index matrix.

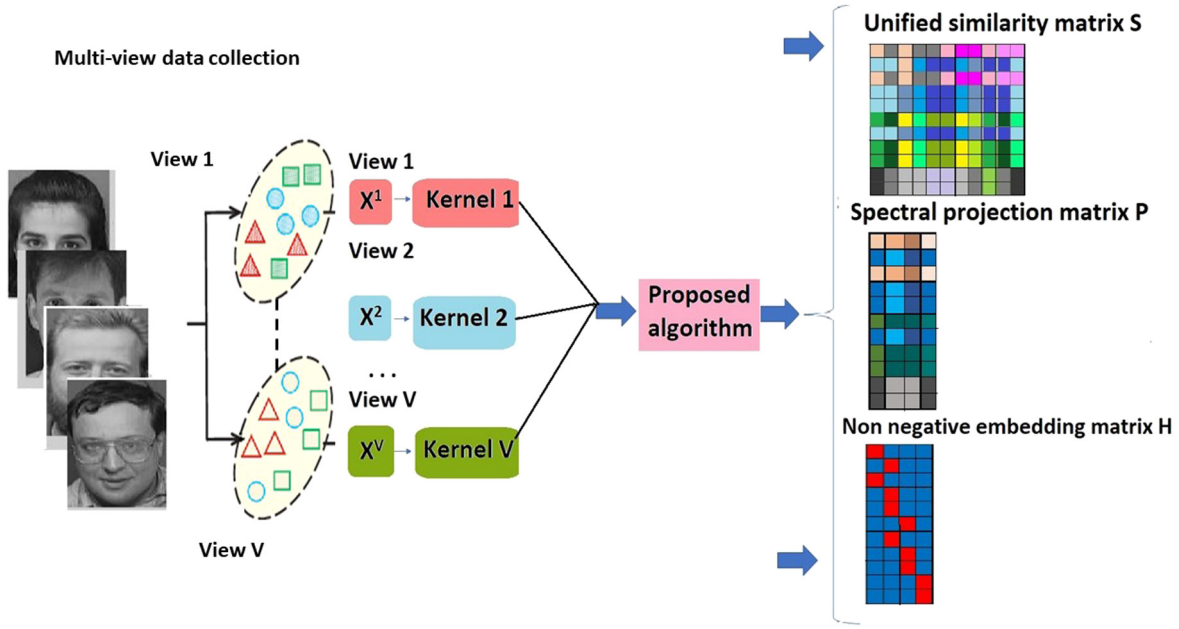

Fig. 1. Illustration of the proposed method.

In [37], the authors propose a model in which the individual graphs, a fused graph, and a spectral projection are estimated simultaneously. Self-representativeness of the data was used in estimating the individual graphs. In [38], a non-negative matrix factorization with multiple views is proposed. The model estimating the view-based two non-negative matrices integrates manifold regularization in the low-dimensional subspace and the pairwise consistencies of interview similarity in these low-dimensional subspaces. In [17], the authors jointly estimate an optimal graph and an adequate consensus kernel for clustering by forcing the global kernel matrix to be a convex combination of a set of basis kernels. Their proposed model enforces a regularization of the unified graph and the final kernel matrix. In [39], the authors use the correntropy-induced metric (CIM) to deal with the noise that exists in each view. They use view-specific embedding from an information theoretic perspective. In [40], the authors propose the algorithm Cross-view Matching Clustering (COMIC), which can cluster data with multiple views. The algorithm can also estimate the number of clusters. COMIC provides cross-view consensus on view-specific similarity graphs instead of view-specific data representations.

In [41], the authors provide an overview of multi-view clustering. This survey describes a wide range of multi-view clustering methods, including both generative and discriminative approaches. Furthermore, the authors of this survey divide these algorithms into many groups and give numerous examples of how they are used for multi-view clustering. In [42], the authors introduce a method called Multi-view cluster analysis with incomplete data to understand treatment effects. Indeed, sometimes data entries are missing in several of the views. Current multiview co-clustering approaches are not able to successfully deal with incomplete data, especially when there are many patterns of incomplete data. By using an indicator matrix whose entries indicate which data items are present, and measuring clustering performance based solely on the observed values, this method provides an improved approach for multi-view co-clustering algorithms to deal with the missing data problem. Moreover, this method is less prone to imputation uncertainty than standard methods that substitute missing data to perform regular multi-view data clustering.

3. Proposed approach

We introduce a new approach called Multi-View Clustering via Consensus Graph Learning and Nonnegative Embedding (MVCGE), which combines the advantages of graph learning methods and matrix factorization methods. MVCGE achieves the clustering results without any additional step. Fig. 1 shows an illustration of our proposed multi-view clustering method.

The proposed method can simultaneously estimate (1) the consensus similarity matrix, (2) the consensus data representation matrix, and (3) the nonnegative cluster index matrix. Moreover, the weight of each view is automatically updated without any additional parameters. Given n samples and V views (feature vectors), the data matrix of each view can be represented as $\mathbf{X}^v = [\mathbf{x}_1^v, \mathbf{x}_2^v, \ldots, \mathbf{x}_n^v] \in \mathbb{R}^{d^v \times n}$, where d^v represents the number of features in the corresponding view, where $v = 1, \ldots, V$. The corresponding kernel matrices are denoted by $\mathbf{K}^v$. The dataset is to be grouped into K clusters based on the V views. The unknown matrices are $\mathbf{S} \in \mathbb{R}^{n \times n}$, $\mathbf{P} \in \mathbb{R}^{n \times K}$, and $\mathbf{H} \in \mathbb{R}^{n \times K}$. Our proposed method estimates these matrices simultaneously by integrating several properties such as graph construction using self-representation of data, smoothness of cluster labels, and spectral data convolution. Thus, our proposed criterion has three main terms. To obtain the first term of our proposed criterion, we used the idea of MVCSK method in [19]. To estimate a consistent graph matrix, this method exploits the property of the data to express itself, where the data is mapped nonlinearly. Therefore, the consistent graph matrix $\mathbf{S}$ should satisfy the condition $\min \sum_{v=1}^{V} \|\Phi(\mathbf{X}^v) - \Phi(\mathbf{X}^v)\ \mathbf{S}\| = \sum_{v=1}^{V} \sqrt{Tr\,(\mathbf{K}^v - 2\,\mathbf{K}^v\,\mathbf{S} + \mathbf{S}^T\,\mathbf{K}^v\mathbf{S}\,)}$, where $\mathbf{K}^v = \Phi(\mathbf{X}^v)^T\Phi(\mathbf{X}^v)$ and $\Phi()$ is a given nonlinear mapping, which should not be explicitly stated, since only the knowledge of the kernel matrix $\mathbf{K}^v$ is needed. Moreover, to avoid the trivial solution of the consistent graph matrix, a regularization term is used to control the values in this matrix. The first term is as follows:

$$\min_{\mathbf{S}} \sum_{v=1}^{V} \sqrt{Tr\,(\mathbf{K}^v - 2\,\mathbf{K}^v\,\mathbf{S} + \mathbf{S}^T\,\mathbf{K}^v\mathbf{S}\,)} + \alpha\,\|\mathbf{S}\|_2^2 \;\; s.t. \;\; \mathbf{S} \geq 0. \qquad (1)$$

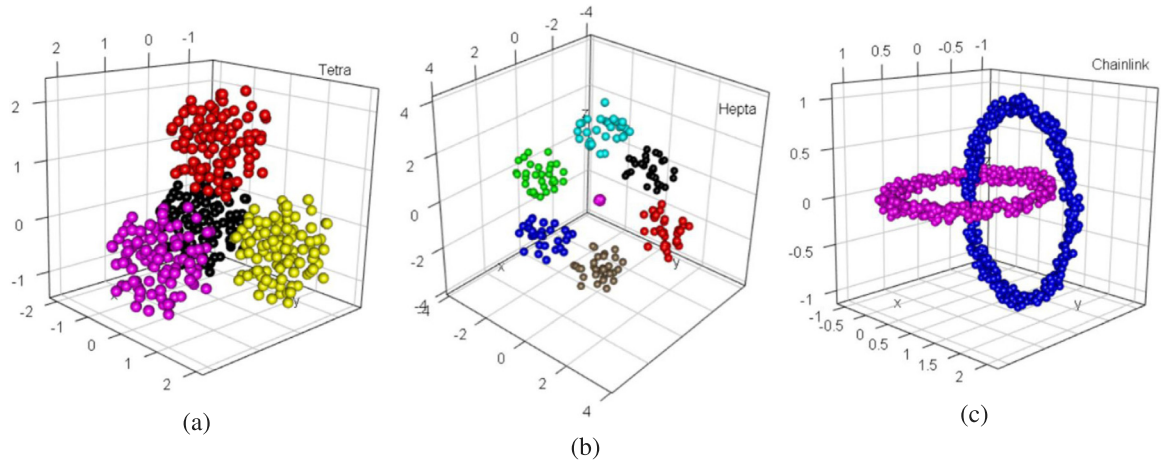

Fig. 2. Visualization of the original synthetic datasets: (a) Tetra, (b) Hepta, and (c) Chainlink.

It is also important to assign a weight parameter to each view to represent the contribution of each view to the clustering process. The square root in Eq. (1) is used to automatically update the weight of each view [19] automatically. The weight of the view w_v is given by:

$$w_v = \frac{1}{2\sqrt{Tr\left(\mathbf{K}^v - 2\,\mathbf{K}^v\,\mathbf{S} + \mathbf{S}^T\,\mathbf{K}^v\mathbf{S}\right)}}. \tag{2}$$

By using the weight expression in Eq. (2), it can be shown that problem (1) is equivalent to the following problem:

$$\min_{\mathbf{S}} \sum_{v=1}^{V} w_v\, Tr\left(\mathbf{K}^v - 2\,\mathbf{K}^v\,\mathbf{S} + \mathbf{S}^T\,\mathbf{K}^v\mathbf{S}\right) + \alpha\,\|\mathbf{S}\|_2^2 \;\; s.t. \;\; \mathbf{S} \geq 0. \tag{3}$$

The clustering result is obtained from the nonnegative embedding matrix $\mathbf{H}$, which provides the cluster indices by taking the index of the highest element in the row vector $\mathbf{H}_{i*} \in \mathbb{R}^K$. Since the matrix $\mathbf{H}$ is used for the final cluster assignment, it is important to use a smoothing term for this matrix so that it is more coherent with the graph entries. The smoothing term ensures that two data points $\mathbf{x}_i^v$ and $\mathbf{x}_j^v$ that are similar (i.e., the value of the corresponding value in the similarity matrix S_{ij} is large) are necessarily in the same cluster (i.e., the corresponding cluster index $\mathbf{H}_{i*}$ and $\mathbf{H}_{i*}$ are close). Therefore, the second term of our criterion is given by:

$$\min_{\mathbf{H}} \frac{1}{2} \sum_i \sum_j \|\mathbf{H}_{i*} - \mathbf{H}_{j*}\|^2 S_{ij} = \min_{\mathbf{H}} Tr\left(\mathbf{H}^T\,\mathbf{L}\,\mathbf{H}\right), \tag{4}$$

where $\mathbf{L} = \mathbf{D} - \mathbf{S} \in \mathbb{R}^{n\times n}$ is the Laplacian matrix of the consistent graph matrix, and $\mathbf{D}$ is a diagonal matrix whose elements are given by: $D_{ii} = \sum_{j=1}^{n} \frac{S_{ij}+S_{ji}}{2}$. The third term of our proposed method states that the cluster index of the ith instance (the row vector $\mathbf{H}_{i,*}$) is set to the convolution of the spectral representation $\mathbf{P}$ with the ith row of the graph matrix $\mathbf{S}_{i,*}$. This approach has two main advantages. First, the clustering is performed in a single step. Second, the clustering uses the consolidated spectral representation of the neighbors obtained in the consensus graph.

Moreover, inspired by the principle of data convolution, the nonnegative embedding matrix used to obtain the final clustering assignment will be equal to "$\mathbf{H} = \max(\mathbf{S}\,\mathbf{P}, 0)$". This means that the nonnegative matrix is the result of the convolution of the spectral data representation with the graph. The third term of our criterion binds the cluster index label to the consensus spectral representation. Therefore, the cluster index matrix should satisfy the following condition:

$$\min_{\mathbf{H}\geq 0} \|\mathbf{H} - \mathbf{S}\,\mathbf{P}\|_F^2 \tag{5}$$

where the matrix $\mathbf{P} \in \mathbb{R}^{n\times K}$ is a consensus data representation. In our work, it is initialized to a unified spectral representation of the data. The proposed method tunes this representation according to a global objective.

Since the matrix $\mathbf{P}$ is orthogonal, Eq. (5) can take another form to illustrate the factorization of the graph matrix $\mathbf{S}$ using the nonnegative embedding matrix $\mathbf{H}$ and the spectral projection matrix $\mathbf{P}$. It can be written as:

$$\min_{\mathbf{H},\,\mathbf{P}} \|\mathbf{S} - \mathbf{H}\,\mathbf{P}^T\|_F^2 \;\; s.t. \;\; \mathbf{H} \geq 0\,, \;\; \mathbf{P}^T\,\mathbf{P} = \mathbf{I}. \tag{6}$$

Our final objective function is obtained by adding the three terms from Eqs. (1), (4), and (6).

$$\begin{aligned} \min_{\mathbf{S},\,\mathbf{P},\,\mathbf{H}} & \sum_{v=1}^{V} w_v\, Tr\left(\mathbf{K}^v - 2\,\mathbf{K}^v\,\mathbf{S} + \mathbf{S}^T\,\mathbf{K}^v\mathbf{S}\right) + \alpha\,\|\mathbf{S}\|_2^2 \\ & + \lambda_1\, Tr\left(\mathbf{H}^T\,\mathbf{L}\,\mathbf{H}\right) + \lambda_2\,\|\mathbf{S} - \mathbf{H}\,\mathbf{P}^T\|_2^2 \\ s.t. \;\; & \mathbf{S} \geq 0\,, \;\; \mathbf{P}^T\,\mathbf{P} = \mathbf{I}\,, \;\; \mathbf{H}^T\,\mathbf{H} = \mathbf{I}\,, \;\; \mathbf{H} \geq 0, \end{aligned} \tag{7}$$

where α, λ_1 and λ_2 are three regularization parameters.

Optimization. We use an iterative update procedure to solve our objective function. In MVCGE, three matrices are unknown: $\mathbf{S}$, $\mathbf{H}$, and $\mathbf{P}$. An alternating optimization scheme is used for the optimization procedure. We proceed as follows:

Step 1: Fix all, estimate H: The problem (7) is:

$$\min_{\mathbf{H}} Tr\left(\mathbf{H}^T\,\mathbf{L}\,\mathbf{H}\right) + \frac{\lambda_2}{\lambda_1}\,\|\mathbf{S}\,\mathbf{P} - \mathbf{H}\|_2^2 \; s.t. \; \mathbf{H}^T\,\mathbf{H} = \mathbf{I}\,, \;\; \mathbf{H} \geq 0. \tag{8}$$

Vanishing the derivative of (8) w.r.t. $\mathbf{H}$ yields:

$$\mathbf{H} = \left(\mathbf{L} + \frac{\lambda_2}{\lambda_1}\,\mathbf{I}\right)^{-1} \frac{\lambda_2}{\lambda_1}\,\mathbf{S}\,\mathbf{P}. \tag{9}$$

To satisfy the orthogonality and non-negativity constraints, an orthogonalization step is first applied to the obtained $\mathbf{H}$, then the negative values of $\mathbf{H}$ are set to zero.

Step 2: Fix all, estimate P: The problem (7) becomes:

$$\min_{\mathbf{P}} \|\mathbf{S} - \mathbf{H}\,\mathbf{P}^T\|^{2_2}. \tag{10}$$

Since $\mathbf{P}$ is orthogonal, i.e., $\mathbf{P}^T\,\mathbf{P} = \mathbf{I}$, $\mathbf{P}$ is obtained by performing the singular value decomposition of $\mathbf{S}^T\,\mathbf{H}$. Let $\mathbf{U}\Sigma\mathbf{V}^T = \mathrm{SVD}\,(\mathbf{S}^T\,\mathbf{H})$, then the solution of (10) is given by:

$$\mathbf{P} = \mathbf{U}\,\mathbf{V}^T \;\; \text{with} \;\; \mathbf{U}\Sigma\mathbf{V}^T = SVD\,(\mathbf{S}^T\,\mathbf{H}). \tag{11}$$

Step 3: Fix all, estimate S:

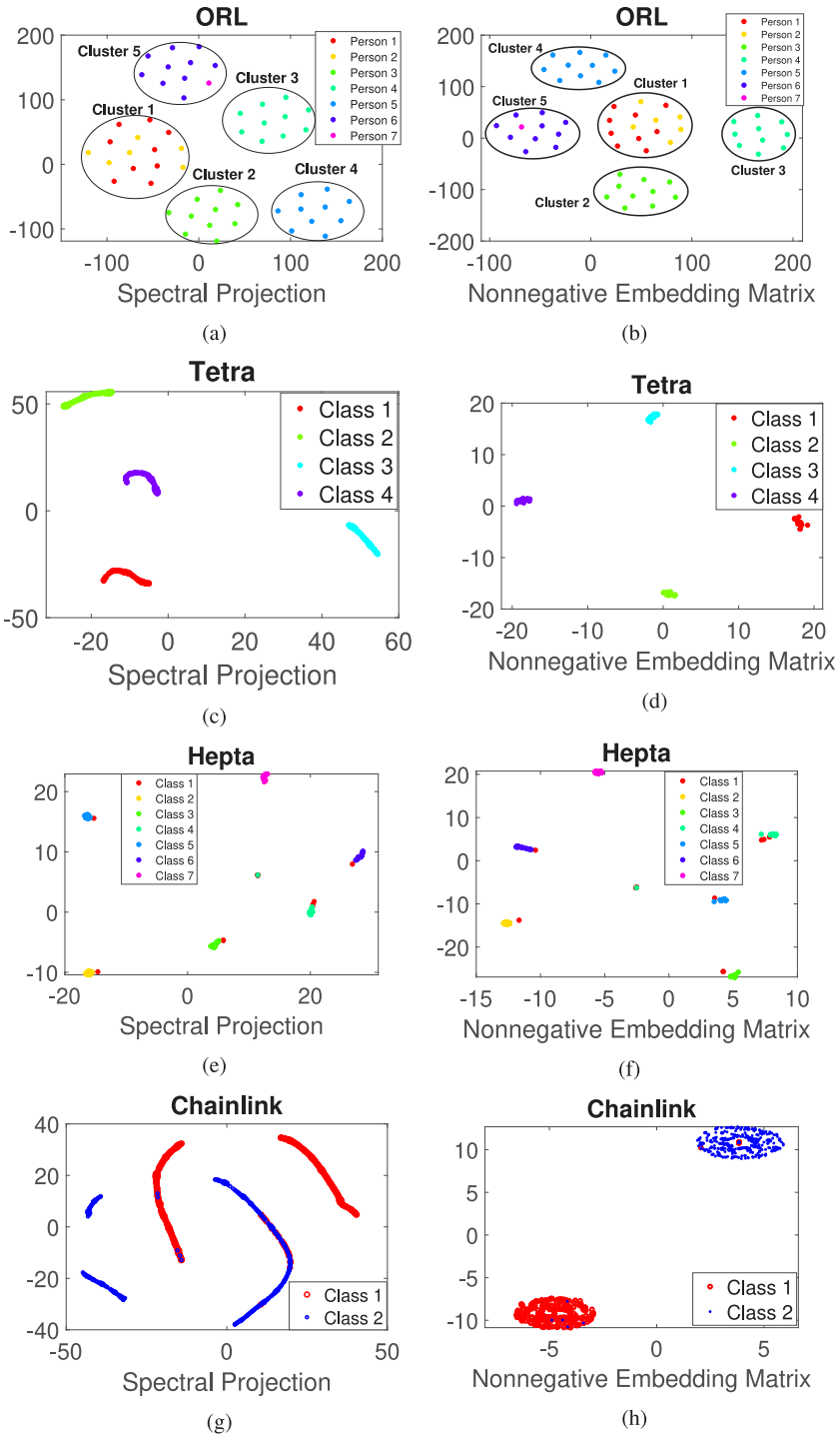

Fig. 3. t-SNE of the spectral projection and nonnegative embedding matrices obtained by the proposed clustering method MVCGE for different datasets.

If we fix $\mathbf{H}$ and $\mathbf{P}$, we need to solve the following problem:

$$\min_{\mathbf{S}} \sum_{v=1}^{V} w_v \, Tr\left(\mathbf{K}^v - 2\,\mathbf{K}^v\,\mathbf{S} + \mathbf{S}^T\,\mathbf{K}^v\,\mathbf{S}\right) + \alpha \left\|\mathbf{S}\right\|_2^2 + \lambda_1 \, Tr\left(\mathbf{H}^T\,\mathbf{L}\,\mathbf{H}\right) + \lambda_2 \left\|\mathbf{S} - \mathbf{H}\,\mathbf{P}^T\right\|_2^2 \;\; s.t. \;\; \mathbf{S} \geq 0. \tag{12}$$

After the spectral clustering analysis, we have the known identity:

$$Tr\left(\mathbf{H}^T\,\mathbf{L}\,\mathbf{H}\right) = \frac{1}{2}\sum_i \sum_j \left\|\mathbf{H}_{i*} - \mathbf{H}_{j*}\right\|^2 S_{ij} = Tr\left(\mathbf{Q}\mathbf{S}\right), \tag{13}$$

where $\mathbf{H}_{i*}$ is the ith row of $\mathbf{H}$. The symmetric matrix $\mathbf{Q}$ denotes the pairwise distance associated with the rows of the matrix $\mathbf{H}$. It is given by $Q_{ij} = \frac{1}{2}\left\|\mathbf{H}_{i*} - \mathbf{H}_{j*}\right\|^2$. Substituting Eq. (13) into Eq. (12), the latter becomes:

$$\min_{\mathbf{S}} \sum_{v=1}^{V} w_v \, Tr\left(\mathbf{K}^v - 2\,\mathbf{K}^v\,\mathbf{S} + \mathbf{S}^T\,\mathbf{K}^v\,\mathbf{S}\right) + \alpha \left\|\mathbf{S}\right\|_2^2 + \lambda_1 \, Tr\left(\mathbf{Q}\mathbf{S}\right) + \lambda_2 \left\|\mathbf{S} - \mathbf{H}\,\mathbf{P}^T\right\|_2^2 \;\; s.t. \;\; \mathbf{S} \geq 0. \tag{14}$$

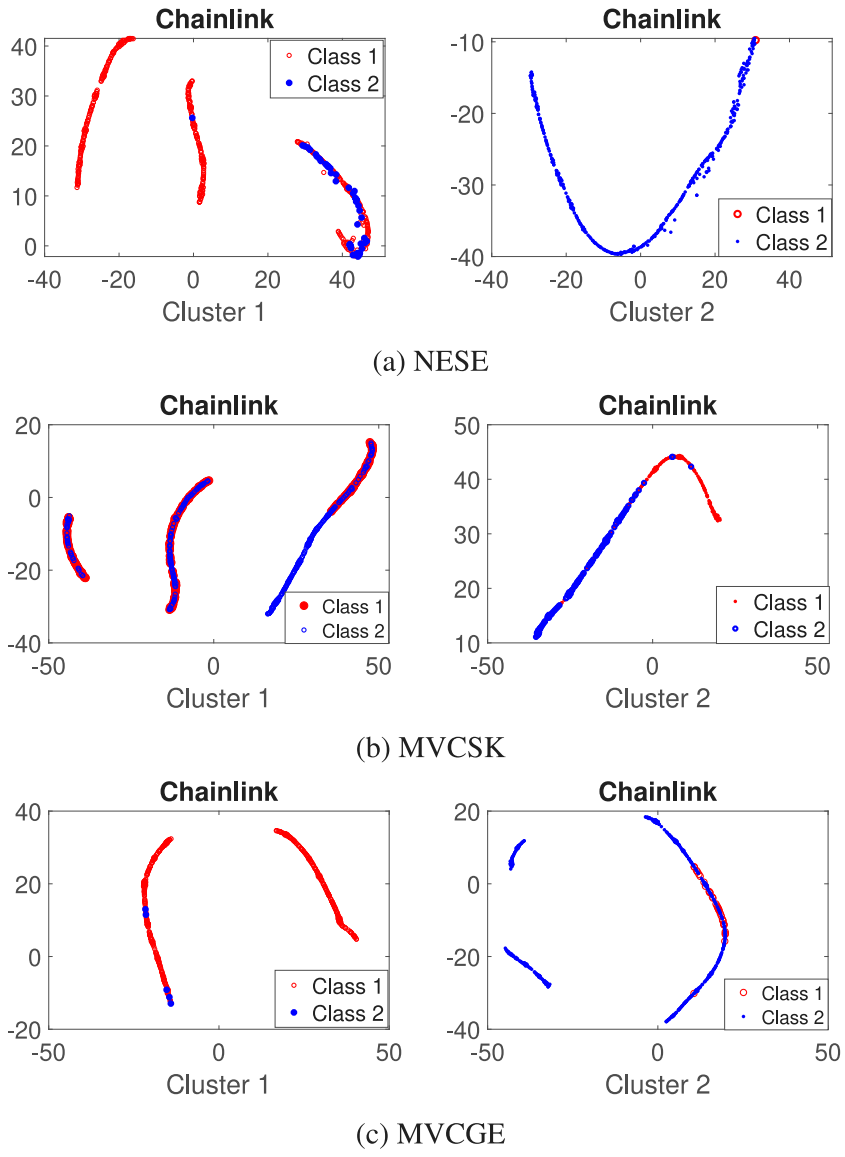

Fig. 4. Visualization of the two clusters obtained by three different methods for the Chainlink dataset.

By making the derivative of Eq. (14) w.r.t. $\mathbf{S}$ vanish, we obtain $\mathbf{S}$ as ($ReLU()$ is the Rectified Linear Unit function):

$$\mathbf{S} = ReLU\left\{\left(\sum_{v=1}^{V} w_v \mathbf{K}^v + (\alpha+\lambda_2)\mathbf{I}\right)^{-1} \times \left(\sum_{v=1}^{V} w_v \mathbf{K}^v + \lambda_2 \mathbf{H}\mathbf{P}^T - \frac{1}{2}\lambda_1 \mathbf{Q}\right)\right\}. \quad (15)$$

Step 4: Fix $\mathbf{H}$, $\mathbf{P}$, and $\mathbf{S}$, and update w_v $(v = 1, \ldots, V)$ using Eq. (2).

The main steps of the proposed approach "Multi-view Clustering via Consensus Graph Learning and Nonnegative Embedding" (MVCGE) are summarized in Algorithm 1.

Algorithm 1 MVCGE

Input: Data samples in V views $\mathbf{X}^v \in \mathbb{R}^{n\times d^v}$, $v = 1, \ldots, V$.
The graph matrices $\mathbf{S}^v$, $v = 1, \ldots, V$.
The spectral embedding matrices $\mathbf{P}^v$, $v = 1, \ldots, V$.
Parameters α, λ_1, λ_2.

Output: The consensus graph matrix $\mathbf{S}$.
The consensus spectral representation matrix $\mathbf{P}$.
The cluster index matrix (nonnegative embedding matrix) $\mathbf{H}$.

Initialization:
The weight of each view $w_v = \frac{1}{V}$.
Compute the kernel matrix $\mathbf{K}^v$ for each view.
Initialize $\mathbf{S}$ and $\mathbf{P}$ by taking the average of the matrices $\mathbf{S}^v$ and $\mathbf{P}^v$.

Repeat
Update $\mathbf{H}$ using Eq. (9).
Update $\mathbf{P}$ using Eq. (11).
Update $\mathbf{S}$ using Eq. (15).
Update w_v using Eq. (2).
Until convergence

To initialize the two matrices $\mathbf{S}$ and $\mathbf{P}$, the efficient method used in [43] is used. This method finds the similarity matrix and the corresponding spectral projection matrix for each view. To

Table 1
Description of the real datasets used in the paper.

View	COIL20	ORL	Out-Scene	BBCSport	MSRCv1
1	Intensity-1024	GIST-512	GIST-512	Intensity-3183	GIST-512
2	LBP-3304	LBP-59	LBP-48	LBP-3203	LBP-256
3	Gabor-6750	HOG-864	HOG-256	–	Color moment-24
4	–	Centrist-254	Color mom.-432	–	Centrist-254
5	–	–	–	–	Sift-512
# Samples	1440	400	2688	544	210
# Classes	20	40	8	5	7

View	Extended-Yale	MNIST	MNIST-1000
1	Covariance ch9 gray-45	Resnet50-2048	Resnet50 Pooling-2048
2	LBP-900	VGG16-4096	VGG16 FC1-4096
3	–	–	–
4	–	–	–
5	–	–	–
# Samples	1774	10000	1000
# Classes	28	10	10

obtain the initial unified matrix, the average of all the individual matrices is used.

4. Performance evaluation

4.1. Datasets

The effectiveness of the proposed approach is evaluated using eight real image datasets and three synthetic datasets. The MNIST dataset is relatively large. Table 1 describes the real datasets.

We also used three synthetic datasets: Tetra, Hepta, and Chainlink. They were selected from the Fundamental Clustering Problem Suite (FCPS). For these datasets, only one view is considered.

Tetra contains 400 3D points divided into four groups. Hepta contains 212 3D points grouped into seven well-defined clusters with different variances. Chainlink is formed by two clusters that are not linearly separable. It consists of 1000 3D points. These datasets are visualized in Fig. 2 [44]. All these synthetic datasets use 3D data points $\mathbf{p}_i \in \mathbb{R}^3$. The 3-dimensional datasets are transformed into high-dimensional datasets $\mathbf{x}_i \in \mathbb{R}^{100}$ using the following linear and nonlinear mappings $\mathbf{x}_i = \sigma(\mathbf{U}\sigma(\mathbf{W}\mathbf{p}_i))$ where the sigmoid function σ is used to introduce nonlinearity, $\mathbf{W} \in \mathbb{R}^{10\times 3}$ and $\mathbf{U} \in \mathbb{R}^{100\times 10}$ are two matrices whose entries follow the Gaussian distribution with zero-mean unit variance i.i.d.

4.2. Experimental setup

Several competing methods are used for comparison: (1) Co-training approach for multi-view Spectral Clustering (CotSC) [12], (2) Co-regularized approach for multi-view Spectral Clustering (CorSC) [13], (3) Multi-view Learning Clustering with Adaptive Neighbors (MLAN) [33], (4) Self-weighted Multi-view Clustering with multiple graphs (SwMC) [45], (5) Affinity Aggregation for Spectral Clustering (AASC) [46], (6) Graph Learning for Multi-View clustering (MVGL) [22], (7) Parameter-free Auto-weighted Multiple Graph Learning (AMGL) [28], (8) Multi-view clustering via Adaptively Weighted Procrustes (AWP) [21], (9) Auto-weighted Multi-View Clustering via Kernelized graph learning (MVCSK) [19], (10) Multi-view spectral clustering via integrating Non-negative Embedding and Spectral Embedding (NESE) [18], (11) Sparse Multi-view Spectral Clustering (S-MVSC) [47], (12) Consistency-aware and Inconsistency-aware Graph-based Multi-View Clustering (CI-GMVC) [48], (13) Multi-View Clustering in Latent Embedding Space (MCLES) [49] and (14) multi-view spectral clustering via Constrained Nonnegative Embedding (CNESE) [9]. We also report the Spectral Clustering best view result (SC) [20].

The clustering performance of the proposed approach is compared with other methods by using the authors' source codes with the default or proposal parameter settings,[1] or by directly reporting the best experimental results from the corresponding published papers.[2] A Gaussian kernel function is used to construct the kernel matrix of each view. To initialize our algorithm, we use the same method as in [18], which constructs the similarity matrix of each view according to a smoothing constraint, an ℓ_2 regularization term, and a non-negativity constraint. Then, the corresponding spectral projection matrix of each view is computed, and the final unified similarity matrix and spectral projection matrix is the average of the corresponding matrix of all views. In this way, we obtain the initial values of the matrices **S** and **P**.

In our method, three parameters are used: α, λ_1 and λ_2. The values of α are in the range [0.005 0.9], the values of the parameter λ_1 vary over the set $\{10^{-10}, 10^{-9}, 10^{-8}, 10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}, 10^{-3}\}$ and the values of the parameter λ_2 vary over the set $\{10^{-7}, 10^{-6}, 10^{-5}, 10^{-4}\ 10^{-3}, 10^{-2}, 10^{-1}\}$. In our experiments, the range for each parameter is chosen to encompass a wide range. This ensures that the optimal values for these parameters are within this range. In selecting the values in these ranges, we used a grid search method. Grid search is a method for exhaustively searching a manually defined subset of the parameter space of a given algorithm. The number of parameters of the algorithm is the spatial dimension of the grid. So in our case, the grid is in a 3D space. This method starts by creating the grid, sampling the predefined regions. Then, for each parameter combination (represented by the nodes of the grid), a model is created to find the best parameter combination that provides the best clustering performance. The best clustering result is indicated by a cluster performance metric. To compare our method with other methods, we use four clustering performance metrics: Accuracy (ACC), Normalized Mutual Information (NMI), Purity, and Adjusted Rand Index (ARI). Their definition can be found in [50].

4.3. Experimental results

Our algorithm is tested on real and synthetic datasets. Table 2 shows the results obtained by MVCGE and some other methods on the datasets: ORL, Out-Scene, and Coil20. In this table, the highest scores are marked in bold. The proposed method MVCGE was superior on these datasets. For some competing methods listed in Table 2, the corresponding method is repeated in multiple trials, and then a standard deviation for each indicator is given in parentheses. From this table, we can see that our method

[1] This concerns SC, MVCSK, NESE, S-MVSC, CI-GMVC, MCLES and CNESE.

[2] This concerns the following methods: CotSC, CorSC, MLAN, SwMC, AASC, MVGL, AMGL and AWP.

Table 2
Clustering performance on the ORL, Outdoor-Scene and Coil20 datasets.

Dataset	Method	ACC	NMI	Purity	ARI
ORL	SC-Best [20]	0.66 (±0.02)	0.76 (±0.02)	0.71 (±0.02)	0.67 (±0.01)
	AWP [21]	0.80 (±0.00)	0.91 (±0.00)	0.83 (±0.00)	0.76 (±0.00)
	MLAN [33]	0.78 (±0.00)	0.88 (±0.00)	0.82 (±0.00)	0.67 (±0.00)
	SwMC [45]	0.77 (±0.00)	0.90 (±0.00)	0.83 (±0.00)	0.62 (±0.00)
	AMGL [28]	0.75 (±0.02)	0.90 (±0.02)	0.82 (±0.02)	0.63 (±0.09)
	AASC [46]	0.82 (±0.02)	0.91 (±0.01)	0.85 (±0.01)	0.76 (±0.02)
	MVGL [22]	0.75 (±0.00)	0.88 (±0.00)	0.80 (±0.00)	0.55 (±0.00)
	CorSC [13]	0.77 (±0.03)	0.90 (±0.01)	0.82 (±0.03)	0.72 (±0.04)
	CotSC [12]	0.75 (±0.04)	0.87 (±0.01)	0.78 (±0.03)	0.67 (±0.03)
	NESE [18]	0.82 (±0.00)	0.91 (±0.00)	0.85 (±0.00)	0.75 (±0.00)
	MVCSK [19]	0.85 (±0.02)	0.94 (±0.01)	0.88 (±0.02)	0.81 (±0.02)
	S-MVSC [47]	0.80 (±0.02)	0.93 (±0.01)	0.82 (±0.02)	0.89 (±0.01)
	CI-GMVC [48]	0.81 (±0.00)	0.92 (±0.00)	0.85 (±0.00)	0.74 (±0.00)
	MCLES [49]	0.84 (±0.00)	0.94 (±0.00)	0.88 (±0.00)	0.79 (±0.00)
	CNESE [9]	0.87 (± 0.00)	0.95 (± 0.00)	0.89 (± 0.00)	0.84 (± 0.00)
	MVCGE	**0.93** (±**0.00**)	**0.97** (±**0.00**)	**0.95** (±**0.00**)	**0.92** (±**0.00**)
Out-Scene	SC-Best [20]	0.47 (±0.01)	0.39 (±0.01)	0.57 (±0.01)	0.34 (±0.01)
	AWP [21]	0.65 (±0.00)	0.51 (±0.00)	0.65 (±0.00)	0.42 (±0.00)
	MLAN [33]	0.55 (±0.02)	0.47 (±0.01)	0.55 (±0.02)	0.33 (±0.03)
	SwMC [45]	0.50 (±0.00)	0.47 (±0.00)	0.50 (±0.00)	0.38 (±0.00)
	AMGL [28]	0.51 (±0.05)	0.45 (±0.03)	0.52 (±0.04)	0.34 (±0.05)
	AASC [46]	0.60 (±0.00)	0.48 (±0.00)	0.60 (±0.00)	0.35 (±0.00)
	MVGL [22]	0.42 (±0.00)	0.31 (±0.00)	0.43 (±0.00)	0.16 (±0.00)
	CorSC [13]	0.51 (±0.04)	0.39 (±0.03)	0.52 (±0.03)	0.31 (±0.02)
	CotSC [12]	0.38 (±0.02)	0.22 (±0.01)	0.39 (±0.02)	0.16 (±0.01)
	NESE [18]	0.63 (±0.00)	0.53 (±0.00)	0.66 (±0.00)	0.46 (±0.00)
	MVCSK [19]	0.65 (±0.01)	0.52 (±0.00)	0.65 (±0.01)	0.42 (±0.00)
	S-MVSC [47]	0.48 (±0.01)	0.54 (±0.02)	0.65 (±0.01)	0.46 (±0.04)
	CI-GMVC [48]	0.35 (±0.01)	0.31 (±0.00)	0.35 (±0.01)	0.19 (±0.00)
	MCLES [49]	0.65 (±0.00)	0.53 (±0.00)	0.67 (±0.00)	0.46 (±0.00)
	CNESE [9]	0.66 (± 0.00)	**0.55** (± **0.00**)	0.67 (± 0.00)	**0.47** (± **0.00**)
	MVCGE	**0.70** (±**0.00**)	**0.55** (±**0.00**)	**0.70** (±**0.00**)	**0.47** (±**0.00**)
COIL20	SC-Best [20]	0.73 (±0.01)	0.82 (±0.01)	0.75 (±0.01)	0.68 (±0.02)
	AWP [21]	0.68 (±0.00)	0.87 (±0.00)	0.75 (±0.00)	0.71 (±0.00)
	MLAN [33]	0.84 (±0.00)	0.92 (±0.00)	0.88 (±0.00)	0.81 (±0.00)
	SwMC [45]	0.86 (±0.00)	0.94 (±0.00)	0.90 (±0.00)	0.84 (±0.00)
	AMGL [28]	0.80 (±0.04)	0.91 (±0.02)	0.85 (±0.03)	0.74 (±0.07)
	AASC [46]	0.79 (±0.00)	0.89 (±0.00)	0.83 (±0.00)	0.76 (±0.00)
	MVGL [22]	0.78 (±0.00)	0.88 (±0.00)	0.81 (±0.00)	0.75 (±0.00)
	CorSC [13]	0.68 (±0.04)	0.78 (±0.02)	0.70 (±0.03)	0.62 (±0.03)
	CotSC [12]	0.70 (±0.03)	0.80 (±0.02)	0.72 (±0.03)	0.65 (±0.03)
	NESE [18]	0.77 (±0.00)	0.88 (±0.00)	0.82 (±0.00)	0.69 (±0.00)
	MVCSK [19]	0.65 (±0.04)	0.80 (±0.02)	0.70 (±0.03)	0.61 (±0.05)
	S-MVSC [47]	0.62 (±0.01)	0.86 (±0.02)	0.77 (±0.02)	0.97 (±0.02)
	CI-GMVC [48]	0.86 (±0.00)	0.94 (±0.00)	0.90 (±0.00)	0.83 (±0.00)
	MCLES [49]	0.79 (±0.00)	0.88 (±0.00)	0.83 (±0.00)	0.75 (±0.00)
	CNESE [9]	0.82 (± 0.00)	0.88 (± 0.00)	0.82 (± 0.00)	0.78 (± 0.00)
	MVCGE	**1.00** (±**0.00**)	**1.00** (±**0.00**)	**1.00** (±**0.00**)	**1.00** (±**0.00**)

and methods MVCSK, NESE S-MVSC, CI-GMVC, MCLES and CNESE perform best, so we can adopt them to test the other datasets.

Table 3 shows a comparison between our method and the aforementioned methods for the BBCSport, MSRCv1, Extended-Yale, MNIST and MNIST-1000 datasets. For the MNIST dataset, which is a large image dataset (i.e., the number of samples is equal to 10000), each image has two deep descriptors, which means that the data already has some nonlinearity. Then, the use of the large kernel matrices can be skipped. Therefore, the criterion of our method reduces to the last two terms, where we only update the spectral projection matrix and the non-negative embedding matrix.

Our method is applied to the synthetic datasets: Tetra, Chainlink, and Hepta. The results are presented in Table 4.

4.4. Ablation study

Our proposed criterion (7) contains three main terms: the graph construction and its regularization, the smoothness term and the convolution term. To illustrate the relevance of the proposed criterion and its terms, we generate four different models with different combinations. These four different variants of MVCGE are: MVCGE-G, MVCGE-S, MVCGE-C and MVCGE-SC. (1) No graph regularization term in the global objective function (7) (i.e., α is set to zero), and we call the obtained method MVCGE-G, which means that only the smoothness and convolution terms in MVCGE are used, (2) No smoothness constraint (λ_1 is set to zero), and we call the obtained method MVCGE-S, (3) No convolution term (λ_2 is set to zero), and we call the obtained method MVCGE-C, and (4) No smoothness and no convolution terms (λ_1 and λ_2 are set to zero). This method is called MVCGE-SC because it is reduced to a consistent graph construction followed by a spectral clustering step. The results obtained with MVCGE-G, MVCGE-S, MVCGE-C, MVCGE-SC and MVCGE are summarized in Table 5. We used three datasets: ORL, MSRCv1 and Tetra. From the results in Table 5, we can see that the regularization of the graph is indeed crucial, since the last two terms depend on this graph. For the ORL and Tetra datasets, it can be seen from the table that the

Table 3
Clustering performance on the BBCSport, MSRCv1, Extended-Yale, MNIST and MNIST-1000 datasets.

Dataset	Method	ACC	NMI	Purity	ARI
BBCSport	MVCSK [19]	0.90 (±0.07)	0.82 (±0.02)	0.90 (±0.02)	0.85 (±0.07)
	NESE [18]	0.72 (±0.00)	0.69 (±0.00)	0.75 (±0.00)	0.60 (±0.00)
	S-MVSC [47]	0.58 (±0.07)	0.67 (±0.01)	0.73 (±0.02)	0.83 (±0.04)
	CI-GMVC [48]	0.61 (±0.00)	0.46 (±0.00)	0.63 (±0.00)	0.36 (±0.00)
	MCLES [49]	0.88 (±0.00)	0.80 (±0.00)	0.88 (±0.00)	0.83 (±0.00)
	CNESE [9]	0.72 (± 0.00)	0.68 (± 0.00)	0.76 (± 0.00)	0.60 (± 0.00)
	MVCGE	**0.98** (±**0.00**)	**0.94** (±**0.00**)	**0.98** (±**0.00**)	**0.95** (±**0.00**)
MSRCv1	MVCSK [19]	0.70 (±0.02)	0.59 (±0.03)	0.70 (±0.02)	0.50 (±0.04)
	NESE [18]	0.77 (±0.00)	0.72 (±0.00)	0.80 (±0.00)	0.64 (±0.00)
	S-MVSC [47]	0.60 (±0.00)	0.69 (±0.02)	0.74 (±0.02)	0.79 (±0.01)
	CI-GMVC [48]	0.74 (±0.00)	0.72 (±0.00)	0.77 (±0.00)	0.59 (±0.00)
	MCLES [49]	0.90 (±0.01)	0.83 (±0.02)	0.90 (±0.01)	0.77 (±0.00)
	CNESE [9]	0.86 (±0.00)	0.76 (±0.00)	0.86 (±0.00)	0.72 (±0.00)
	MVCGE	**0.93** (±**0.00**)	**0.87** (±**0.00**)	**0.93** (±**0.00**)	**0.85** (±**0.00**)
Extended-Yale	MVCSK [19]	0.33 (±0.00)	0.42 (±0.00)	0.34 (±0.00)	0.18 (±0.00)
	NESE [18]	0.43 (±0.00)	0.58 (±0.00)	0.47 (±0.00)	0.25 (±0.00)
	S-MVSC [47]	0.48 (± 0.03)	0.61 (± 0.01)	0.60 (± 0.01)	0.36 (± 0.05)
	CI-GMVC [48]	0.32 (± 0.00)	0.34 (± 0.00)	0.35 (± 0.00)	0.02 (± 0.00)
	MCLES [49]	0.48 (± 0.03)	0.48 (± 0.00)	0.48 (± 0.01)	0.10 (± 0.05)
	CNESE [9]	0.60 (± 0.00)	0.75 (± 0.00)	0.60 (± 0.00)	0.51 (± 0.00)
	MVCGE	**0.88** (±**0.00**)	**0.86** (±**0.00**)	**0.88** (±**0.00**)	**0.77** (±**0.00**)
MNIST	MVCSK [19]	0.49 (± 0.00)	0.41 (± 0.00)	0.50 (± 0.00)	0.29 (± 0.00)
	NESE [18]	**0.81** (±**0.00**)	**0.83** (±**0.00**)	0.85 (±0.00)	0.76 (± 0.00)
	S-MVSC [47]	0.77 (± 0.01)	0.81 (± 0.01)	0.81 (± 0.02)	0.76 (± 0.07)
	CI-GMVC [48]	0.66 (± 0.00)	0.71 (± 0.00)	0.71 (± 0.00)	0.51 (± 0.00)
	MCLES [49]	0.80 (±0.00)	**0.83** (±**0.00**)	0.85 (±0.00)	0.77 (±0.00)
	CNESE [9]	**0.81** (±**0.00**)	0.83 (± 0.00)	**0.86** (±**0.00**)	**0.78** (±**0.00**)
	MVCGE	**0.81** (±**0.00**)	**0.83** (±**0.00**)	0.85 (±0.00)	0.77 (±0.00)
MNIST-1000	MVCSK [19]	0.70 (± 0.00)	0.61 (± 0.00)	0.70 (± 0.00)	0.52 (± 0.00)
	NESE [18]	0.78 (± 0.00)	0.79 (± 0.00)	0.83 (±0.00)	0.71 (±0.00)
	S-MVSC [47]	0.66 (± 0.02)	0.76 (± 0.01)	0.76 (± 0.00)	0.77 (± 0.05)
	CI-GMVC [48]	0.65 (± 0.00)	0.71 (± 0.00)	0.73 (±0.00)	0.50 (±0.00)
	MCLES [49]	0.73 (± 0.02)	0.72 (± 0.01)	0.77 (±0.02)	0.58 (±0.04)
	CNESE [9]	0.77 (± 0.00)	0.77 (± 0.00)	0.81 (±0.00)	0.68 (±0.00)
	MVCGE	**0.86** (±**0.00**)	**0.83** (±**0.00**)	**0.86** (±**0.00**)	**0.78** (±**0.00**)

Table 4
Clustering performance on the three synthetic datasets.

Dataset	Method	ACC	NMI	Purity	ARI
Tetra	NESE [18]	0.64	0.75	0.75	0.63
	MVCSK [19]	0.97	0.93	0.97	0.92
	S-MVSC [47]	0.70	0.50	0.44	0.70
	CI-GMVC [48]	0.63	0.52	0.67	0.43
	MCLES [49]	0.85	0.88	0.89	0.80
	CNESE [9]	0.66	0.62	0.75	0.54
	MVCGE	**1.00**	**1.00**	**1.00**	**1.00**
Hepta	NESE [18]	0.81	0.79	0.85	0.73
	MVCSK [19]	0.89	0.85	0.89	0.80
	S-MVSC [47]	0.66	0.63	0.47	0.70
	CI-GMVC [48]	0.77	0.76	0.81	0.68
	MCLES [49]	0.87	0.82	0.84	0.80
	CNESE [9]	0.78	0.70	0.79	0.63
	MVCGE	**0.92**	**0.85**	**0.92**	**0.83**
Chainlink	NESE [18]	0.93	0.69	0.93	0.73
	MVCSK [19]	0.63	0.05	0.63	0.07
	S-MVSC [47]	0.67	0.14	0.78	0.12
	CI-GMVC [48]	0.55	0.01	0.55	0.01
	MCLES [49]	0.90	0.72	0.86	0.76
	CNESE [9]	0.95	0.70	0.95	0.78
	MVCGE	**0.96**	**0.78**	**0.96**	**0.85**

Table 5
Ablation study with different models. The best performance for each indicator is in bold.

Dataset	Variant	ACC	NMI	Purity	ARI
ORL	MVCGE-G	0.46	0.66	0.48	0.27
	MVCGE-S	0.75	0.88	0.76	0.72
	MVCGE-C	0.86	0.94	0.88	0.81
	MVCGE-SC	0.69	0.86	0.75	0.57
	MVCGE	**0.93**	**0.97**	**0.95**	**0.92**
MSRCv1	MVCGE-G	0.68	0.57	0.68	0.45
	MVCGE-S	0.72	0.63	0.74	0.54
	MVCGE-C	0.70	0.60	0.70	0.50
	MVCGE-SC	0.59	0.54	0.61	0.37
	MVCGE	**0.93**	**0.87**	**0.93**	**0.85**
Tetra	MVCGE-G	0.70	0.59	0.72	0.55
	MVCGE-S	0.91	0.79	0.91	0.77
	MVCGE-C	0.97	0.93	0.97	0.92
	MVCGE-SC	0.56	0.35	0.57	0.33
	MVCGE	**1.00**	**1.00**	**1.00**	**1.00**

smoothness term has a larger impact on the clustering results. However, for the MSRCv1 dataset, the convolution term is more important than the smoothness term. This is normal and is due to the different types of datasets used in this work. The results obtained with MVCGE-SC show the importance of the last two terms in the objective function for all datasets. All these results indicate that the inclusion of all terms in the objective function contributed to the good clustering performance of our proposed method.

4.5. Analysis of results and method comparison

According to Table 2, the performance of all multi-view clustering methods is better than that of SC-Best, which corresponds to the spectral clustering method applied to the best single view. In fact, the presence of multiple views brings additional information to the clustering method so that it can process the datasets better. The proposed method gives the best performance followed by NESE, MVCSK, S-MVSC, CI-GMVC, MCLES and CNESE

methods. With respect to the large MNIST dataset shown in Table 3, MVCGE shows similar results to CNESE for most cluster indicators. Moreover, the performance of our method is better than most competing methods for the same dataset, which shows that we are able to handle large datasets with this new approach and achieve good results. The results we obtained on the MNIST-1000 dataset (see Table 3) demonstrate the superiority of the proposed method. From Table 4, MVCGE achieves the best results for the synthetic datasets even when applied to the single view datasets.

4.6. Clustering visualization

In this section, we visualize the clustering obtained by the proposed MVCGE method on four datasets using the t-SNE technique [51]. In all subfigures of Fig. 3, the spectral projection matrix **P** and the nonnegative embedding matrix **H** are shown for ORL, Tetra, Hepta and Chainlink. In these subfigures, each point corresponds to an image (ORL) or a 3D point (synthetic datasets). We emphasize that the color corresponds to the ground-truth classes.

For ORL we present five clusters. From Figs. 3(a) and 3(b), it can be seen that Cluster 1 and Cluster 5 are not pure, as they each contain images associated with two different individuals, which explains the result obtained in Table 2. The clustering of the synthetic datasets of Tetra and Chainlink is shown in Figs. 3(c), 3(d), 3(g) and 3(h). Some clustering errors are observed, which explain the results obtained in Table 4. Moreover, the visualization of the spectral representation and cluster index matrices (nonnegative embedding) associated with the Tetra dataset shows well-separated clusters in Figs. 3(e) and 3(f). This confirms the perfect performance of 100% in Table 4. Fig. 4 shows the estimated two clusters obtained by MVCSK, NESE and MVCGE methods for the Chainlink dataset. According to this figure, the worst result is that of MVCSK and the best is that of our method. It is clear that the two clusters are well separated by using MVCGE and Fig. 4(c) has few clustering errors.

5. Conclusion

A novel approach for multi-view clustering is proposed. Unlike existing methods, it simultaneously learns the unified similarity matrix, the uniform spectral projection matrix, the non-negative embedding matrix (cluster index matrix) and the weight of each view. Thus, the final clustering result can be obtained directly from the nonnegative embedding matrix, which is a convolution of the consensus data representation over the graph. The proposed method combines the advantages of graph-based approaches and matrix factorization-based methods. Experimental results on real and synthetic datasets have shown that MVCGE outperforms many state-of-the-art methods. As an outlook, we envision the development of a scalable variant of the proposed approach capable of handling large datasets with reasonable computational cost. Moreover, following the CNESE method in [9], our approach could be tested by adding multiple constraints to the non-negative embedding matrix to make it more precise. By applying multiple constraints to the non-negative embedding matrix generated by our approach, we can improve the results. We also propose to apply the presented method to other types of machine learning, such as semi-supervised learning and classification tasks. Finally, our method can be extended to cases where some views have missing data, i.e., the corresponding value in the similarity matrix is missing.

CRediT authorship contribution statement

S. El Hajjar: Software, Validation, Data curation, Writing – original draft, Writing – review & editing. **F. Dornaika:** Conceptualization, Formal analysis, Investigation, Writing – original draft, Writing – review & editing, Supervision. **F. Abdallah:** Formal analysis. **N. Barrena:** Validation, Data curation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

[1] X. Xu, J. Li, M. Zhou, J. Xu, J. Cao, Accelerated two-stage particle swarm optimization for clustering not-well-separated data, IEEE Trans. Syst. Man Cybern. Syst. 51 (1) (2020) 4212–4223.
[2] G. Zhong, C.-M. Pun, Subspace clustering by simultaneously feature selection and similarity learning, Knowl.-Based Syst. 193 (2020) 105512.
[3] Z. Ren, H. Lei, Q. Sun, C. Yang, Simultaneous learning coefficient matrix and affinity graph for multiple kernel clustering, Inform. Sci. 547 (2021) 289–306.
[4] Q. Zhao, L. Zong, X. Zhang, X. Liu, H. Yu, Multi-view clustering via clusterwise weights learning, Knowl.-Based Syst. 193 (2020) 105459.
[5] Y. Zheng, X. Zhang, Y. Xu, M. Qin, Z. Ren, X. Xue, Robust multi-view subspace clustering via weighted multi-kernel learning and co-regularization, IEEE Access 8 (2020) 113030–113041.
[6] T. Zhou, C. Zhang, X. Peng, H. Bhaskar, J. Yang, Dual shared-specific multiview subspace clustering, IEEE Trans. Cybern. 50 (2020) 3517–3530.
[7] S. Zhou, X. Liu, M. Li, E. Zhu, L. Liu, C. Zhang, J. Yin, Multiple kernel clustering with neighbor-kernel subspace segmentation, IEEE Trans. Neural Netw. Learn. Syst. 31 (4) (2020) 1351–1362.
[8] X. Zhu, Y. Zhu, W. Zheng, Spectral rotation for deep one-step clustering, Pattern Recognit. 105 (2020) 107175.
[9] S. El Hajjar, F. Dornaika, F. Abdallah, Multi-view spectral clustering via constrained nonnegative embedding, Inf. Fusion 78 (2022) 209–217.
[10] T. Xia, D. Tao, T. Mei, Y. Zhang, Multiview spectral embedding, IEEE Trans. Syst. Man Cybern. B 40 (6) (2010) 1438–1446.
[11] C.-G. Li, R. Vidal, Structured sparse subspace clustering: A unified optimization framework, in: Proceedings Of The IEEE Conference On Computer Vision And Pattern Recognition, 2015, pp. 277–286.
[12] A. Kumar, H. Daumé, A co-training approach for multi-view spectral clustering, in: Proceedings Of The 28th International Conference On International Conference On Machine Learning, 2011, pp. 393–400.
[13] A. Kumar, P. Rai, H. Daumé, Co-regularized multi-view spectral clustering, in: Proceedings Of The 24th International Conference On Neural Information Processing Systems, 2011, pp. 1413–1421.
[14] J. Li, J.Z. Wang, Real-time computerized annotation of pictures, IEEE Trans. Pattern Anal. Mach. Intell. 30 (6) (2008) 985–1002.
[15] D. Greene, P. Cunningham, A matrix factorization approach for integrating multiple data views, in: Joint European Conference On Machine Learning And Knowledge Discovery In Databases, 2009, pp. 423–438.
[16] Z. Kang, C. Peng, Q. Cheng, Kernel-driven similarity learning, Neurocomputing 267 (2017) 210–219.
[17] Z. Ren, H. Li, C. Yang, Q. Sun, Multiple kernel subspace clustering with local structural graph and low-rank consensus kernel learning, Knowl.-Based Syst. 188 (2020) 105040.
[18] Z. Hu, F. Nie, R. Wang, X. Li, Multi-view spectral clustering via integrating nonnegative embedding and spectral embedding, Inf. Fusion 55 (2020) 251–259.
[19] S. Huang, Z. Kang, I.W. Tsang, Z. Xu, Auto-weighted multi-view clustering via kernelized graph learning, Pattern Recognit. 88 (2019) 174–184.
[20] W. Zhu, F. Nie, X. Li, Fast spectral clustering with efficient large graph construction, in: 2017 IEEE International Conference On Acoustics, Speech And Signal Processing, 2017, pp. 2492–2496.
[21] F. Nie, L. Tian, X. Li, Multiview clustering via adaptively weighted procrustes, in: Proceedings Of The 24th ACM SIGKDD International Conference On Knowledge Discovery & Data Mining, 2018, pp. 2022–2030.
[22] K. Zhan, C. Zhang, J. Guan, J. Wang, Graph learning for multiview clustering, IEEE Trans. Cybern. 48 (10) (2017) 2887–2895.
[23] Y.-M. Xu, C.-D. Wang, J.-H. Lai, Weighted multi-view clustering with feature selection, Pattern Recognit. 53 (2016) 25–35.
[24] L. Zong, X. Zhang, X. Liu, H. Yu, Weighted multi-view spectral clustering based on spectral perturbation, in: Proceedings Of The AAAI Conference On Artificial Intelligence, 2018.

[25] C. Xu, D. Tao, C. Xu, Multi-view self-paced learning for clustering, in: Twenty-Fourth International Joint Conference On Artificial Intelligence, 2015, pp. 3974–3980.
[26] S. Huang, Z. Kang, Z. Xu, Auto-weighted multi-view clustering via deep matrix decomposition, Pattern Recognit. 97 (2020) 107015.
[27] P. Ren, Y. Xiao, P. Xu, J. Guo, X. Chen, X. Wang, D. Fang, Robust auto-weighted multi-view clustering, in: Proceedings Of The Twenty-Seventh International Joint Conference On Artificial Intelligence, 2018, pp. 2644–2650.
[28] F. Nie, J. Li, X. Li, et al., Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification, in: Proceedings Of The Twenty-Fifth International Joint Conference On Artificial Intelligence, 2016, pp. 1881–1887.
[29] Z. Wu, S. Liu, C. Ding, Z. Ren, S. Xie, Learning graph similarity with large spectral gap, IEEE Trans. Syst. Man Cybern. Syst. (2019).
[30] M. White, Y. Yu, X. Zhang, D. Schuurmans, Convex multi-view subspace learning, in: NIPS, 2012, pp. 1682–1690.
[31] J. Liu, C. Wang, J. Gao, J. Han, Multi-view clustering via joint nonnegative matrix factorization, in: Proceedings Of The 2013 SIAM International Conference On Data Mining, 2013, pp. 252–260.
[32] Z. Yang, Y. Zhang, Y. Xiang, W. Yan, S. Xie, Non-negative matrix factorization with dual constraints for image clustering, IEEE Trans. Syst. Man Cybern. Syst. 50 (7) (2020) 2524–2533.
[33] F. Nie, G. Cai, J. Li, X. Li, Auto-weighted multi-view learning for image clustering and semi-supervised classification, IEEE Trans. Image Process. 27 (3) (2017) 1501–1511.
[34] W. Guo, Y. Shi, S. Wang, A unified scheme for distance metric learning and clustering via rank-reduced regression, IEEE Trans. Syst. Man Cybern. Syst. (2019) 1–12.
[35] Z. Huang, Y. Ren, X. Pu, L. Pan, D. Yao, G. Yu, Dual self-paced multi-view clustering, Neural Netw. 140 (2021) 184–192.
[36] D. Xie, X. Zhang, Q. Gao, J. Han, S. Xiao, X. Gao, Multiview clustering by joint latent representation and similarity learning, IEEE Trans. Cybern. 50 (11) (2020) 4848–4854.
[37] Z. Kang, G. Shi, S. Huang, W. Chen, X. Pu, J.T. Zhou, Z. Xu, Multi-graph fusion for multi-view spectral clustering, Knowl.-Based Syst. 189 (2020) 105102.
[38] X. Wang, T. Zhang, X. Gao, Multiview clustering based on non-negative matrix factorization and pairwise measurements, IEEE Trans. Cybern. 49 (9) (2019) 3333–3346.
[39] L. Xing, B. Chen, S. Du, Y. Gu, N. Zheng, Correntropy-based multiview subspace clustering, IEEE Trans. Cybern. (2019) 1–14.
[40] X. Peng, Z. Huang, J. Lv, H. Zhu, J.T. Zhou, COMIC: Multi-view clustering without parameter selection, in: Proceedings Of The 36th International Conference On Machine Learning, 2019, pp. 5092–5101.
[41] G. Chao, S. Sun, J. Bi, A survey on multi-view clustering, IEEE Trans. Artif. Intell. (2021).
[42] G. Chao, J. Sun, J. Lu, A.-L. Wang, D.D. Langleben, C.-S. Li, J. Bi, Multi-view cluster analysis with incomplete data to understand treatment effects, Inform. Sci. 494 (2019) 278–293.
[43] F. Nie, X. Wang, M. Jordan, H. Huang, The constrained Laplacian rank algorithm for graph-based clustering, in: Proceedings Of The Thirtieth AAAI Conference On Artificial Intelligence, 2016, pp. 1969–1976.
[44] M.C. Thrun, A. Ultsch, Clustering benchmark datasets exploiting the fundamental clustering problems, Data Brief 30 (2020) 105501.
[45] F. Nie, J. Li, X. Li, et al., Self-weighted multiview clustering with multiple graphs, in: Proceedings Of The Twenty-Sixth International Joint Conference On Artificial Intelligence, 2017, pp. 2564–2570.
[46] H.-C. Huang, Y.-Y. Chuang, C.-S. Chen, Affinity aggregation for spectral clustering, in: 2012 IEEE Conference On Computer Vision And Pattern Recognition, 2012, pp. 773–780.
[47] Z. Hu, F. Nie, W. Chang, S. Hao, R. Wang, X. Li, Multi-view spectral clustering via sparse graph learning, Neurocomputing 384 (2020) 1–10.
[48] M. Horie, H. Kasai, Consistency-aware and inconsistency-aware graph-based multi-view clustering, in: 2020 28th European Signal Processing Conference, 2021, pp. 1472–1476.
[49] M.-S. Chen, L. Huang, C.-D. Wang, D. Huang, Multi-view clustering in latent embedding space, in: Proceedings Of The AAAI Conference On Artificial Intelligence, Vol. 34, 2020, pp. 3513–3520.
[50] K. Zhan, F. Nie, J. Wang, Y. Yang, Multiview consensus graph clustering, IEEE Trans. Image Process. 28 (3) (2018) 1261–1270.
[51] L. Van der Maaten, G. Hinton, Visualizing data using t-SNE, J. Mach. Learn. Res. 9 (11) (2008).