Consensus One-step Multi-view Subspace Clustering

Pei Zhang, Xinwang Liu, *Senior Member, IEEE,* Sihang Zhou, Wentao Zhao, En Zhu and Zhiping Cai

Abstract—Multi-view clustering has attracted increasing attention in multimedia, machine learning and data mining communities. As one kind of the essential multi-view clustering algorithm, multi-view subspace clustering (MVSC) becomes more and more popular due to its strong ability to reveal the intrinsic low dimensional clustering structure hidden across views. Despite superior clustering performance in various applications, we observe that existing MVSC methods *directly fuse multi-view information in the similarity level by merging noisy affinity matrices*; and *isolate the processes of affinity learning, multi-view information fusion and clustering*. Both factors may cause insufficient utilization of multi-view information, leading to unsatisfying clustering performance. This paper proposes a novel consensus one-step multi-view subspace clustering (COMVSC) method to address these issues. Instead of directly fusing multiple affinity matrices, COMVSC optimally integrates discriminative partition-level information, which is helpful to eliminate noise among data. Moreover, the affinity matrices, consensus representation and final clustering labels matrix are learned simultaneously in a unified framework. By doing so, the three steps can negotiate with each other to best serve the clustering task, leading to improved performance. Accordingly, we propose an iterative algorithm to solve the resulting optimization problem. Extensive experiment results on benchmark datasets demonstrate the superiority of our method against other state-of-the-art approaches.

Index Terms—Multi-view Clustering, Subspace Clustering, Data Fusion.

1 Introduction

Traditional clustering methods usually use single kinds of features to measure the similarity of samples. As well known, the individual feature is insufficient for depicting data points. In contrast, different features often contain complementary information, which could be of benefit to exploring the underlying structure of data [1], [2], [3]. Therefore, multi-view clustering (MVC) has attracted attention in feature selection, data mining and many other machine learning communities [4], [5], [6], [7], [8]. MVC aims to categorize similar data points into the same cluster and dissimilar points into different groups by combining the available multiple feature information and searching for consistent clustering results across different views [9], [10], [11], [12], [13], [14], [15].

- *P. Zhang, X. Liu, S. Wang, X. Guo, E. Zhu and Z. Cai are with the School of Computer, National University of Defense Technology, Changsha 410073, P.R. China (E-mail: xinwangliu@nudt.edu.cn).*
- *S. Zhou is with the College of Intelligence Science and Technology, National University of Defense Technology, Changsha 410073, P.R. China.*

Researchers have proposed numerous multi-view clustering methods in recent years. Yang *et al.* [16] summarize multi-view clustering methods into four categories in terms of the mechanisms they base. The first category is the co-training strategy. This kind of algorithm reaches a consensus by iteratively exploiting prior knowledge or information learned from other views. There are two representative method: co-training multi-view spectral clustering [17] and co-regularized multi-view spectral clustering [18]. The second category refers to multi-kernel clustering [12], [19], [20], [21], [22]. Unlike the other multi-view clustering methods, whose inputs are original data matrices, the multi-kernel inputs are commonly the base kernel matrices. The critical point of multi-kernel clustering is reaching a consistent result by extracting a common structure from multiple kernels. For example, Zhou *et al.* [23] designs a novel multi-kernel clustering method based on the neighbor kernel and subspace segmentation, which not only reveals the inherent structure among multiple kernels but removes the noise and outliers.

The third perspective is the multi-view graph clustering. This kind of method finds a unified graph and then performs spectral clustering or other graph-cut algorithms on the consensus graph [24], [25], [26], [27]. As one of the most successful extensions, multi-view subspace clustering methods recover the underlying subspace structure of data under the assumption that high-dimensional data can be well characterized within low-dimensional subspaces.

Existing multi-view subspace clustering methods have shown their effectiveness and robustness in many applications. Most existing multi-view subspace clustering methods integrate multi-view information in similarity or representation level by merging multiple graphs or representation matrices into a shared one. For example, Guo *et al.* [28] learn a shared sparse subspace representation by performing matrix factorization. Similarly, the centroid-based multi-view low-rank sparse subspace clustering method [29] induces low-rank and sparsity constraints on the shared affinity matrix across different views. Li *et al.* [29] construct a latent representation by maximizing the dependence between pairwise views, which essentially encodes the complementary information among views. Different from obtaining a shared representation or graph directly, [30] and [31] induce Hilbert-Schmidt Independence Criterion(HSIC) and Markov chain to learn complementary subspace representations and then add them together directly

or adaptively to get a unified representation. Although these subspace methods mentioned above have achieved significant improvements, they can still be improved from the following two points: i) The majority of existing multi-view subspace clustering methods learn a shared affinity matrix or graph and then apply spectral clustering to obtain the final clustering result. However, directly learning a common affinity matrix or graph from the original data may affect the clustering structure because the original data often consists of noise and redundancy [32]. Few of them make full use of more informative multi-view partition information for improving clustering results. ii) Previous approaches are usually conducted in a two-step fashion, which implies that the learned common representation may not be suitable for the clustering task. They cannot obtain optimal clustering performance since the similarity learning step is separated from the subsequent clustering step.

To address the above-mentioned issues, in this paper, we propose a novel Consensus One-step Multi-view Subspace Clustering method (COMVSC) to integrate the representation learning and clustering process into a unified framework. In this framework, as shown in Fig. 1, we jointly optimize individual similarity matrices, partition matrices and clustering labels. To be specific, COMVSC firstly establishes similarity learning based on the self-representation manner in each view. Based on the assumption that the individual view's clustering structure should be similar, we propose to fuse clustering indicator matrices of different views into a consensus one. Different from the previous similarity-fusion manner, our method adopts partition-level fusion, which avoids the noise and redundancy in the original data. Furthermore, spectral rotation is introduced to the consensus clustering indicator matrix to directly obtain clustering labels, avoiding the additional k-means or spectral clustering step in previous methods. By taking the interior interactions between the three sub-processes (similarity learning, partition fusion and spectral rotation), each of them can be boosted by others. Moreover, we develop an efficient algorithm to solve the resulting optimization problem. Extensive experiments on several multi-view datasets are conducted to evaluate the effectiveness of our method. As demonstrated, the proposed approach enjoys superior clustering performances in comparison with several state-of-the-art multi-view subspace clustering methods.

The main contributions of this work are summarized as follows:

- We propose a unified multi-view subspace clustering framework which jointly optimizes similarity learning, clustering partition and final clustering labels. Hence our method directly outputs the discrete clustering labels, which avoids the sub-optimal solution of existing two-step approaches.
- Our COMVSC method incorporates multiple source information in partition level from each individual view, which not only preserves the view specific local clustering structure but also guarantees the consistency among multiple views. In addition, partition-level fusion method avoids the noise and redundancy occurring in fusing information in similarity level.
- An iterative algorithm is proposed to solve the resulting optimization problem. Extensive experiments on several multi-view benchmark datasets demonstrate the effectiveness of our method comparing to other state-of-the-art approaches.

The rest of this paper is organized as follows. The related works are introduced in Section 2. Then we formulate our proposed COMVSC method in Section 3 . The corresponding optimization method and its analysis are given in Section 4 and Section 5, respectively. In Section 6, we report the experiment results and conduct experimental and statistical analysis. At the end of the paper, we make a conclusion and prospect in Section 7.

2 Related Work

2.1 Notation

In this paper, matrices are represented with bold capital symbols. For a matrix $\mathbf{A}$, $\mathbf{A}_{i,:}$ and $A_{i,j}$ represent its i-th row and the ij-th element. The Frobenius norm of matrix $\mathbf{A}$ is denoted as $\|\mathbf{A}\|_F$. The ℓ_2 norm of vector $\mathbf{A}_{i,:}$ is $\|\mathbf{A}_{i,:}\|_2$. The transpose, the trace of matrix $\mathbf{A}$ are denoted by $\mathbf{A}^{\mathrm{T}}$, $\mathrm{Tr}(\mathbf{A})$, respectively.

Given a dataset with n samples from m views, points in the v-th view are denoted as $\mathbf{X}^v = [\mathbf{x}_1^v, \mathbf{x}_2^v, \cdots, \mathbf{x}_n^v] \in \mathbb{R}^{d_v \times n}$, where $\mathbf{x}_n^v$ is a d_v-dimension column vector. Then the multi-view dataset can be expressed as $\mathbf{X} = [\mathbf{X}^1, \mathbf{X}^2, \cdots, \mathbf{X}^m]^{\mathrm{T}} \in \mathbb{R}^{d \times n}$, where $d = \sum_{v=1}^{m} d_v$ and d_v is the feature dimension of the v-th view.

In addition, $\mathbf{1}$ denotes a column vector whose elements are all one. $\mathbf{I}_k$ refers to k-dimension identity matrix.

2.2 Subspace clustering

Data points can be represented by underlying low-dimensional subspace. Given n data point $\mathbf{A} \in \mathbb{R}^{d \times n}$, self-representation method [33] is utilized to express each data points with a linear combination of the data themselves. It can be formulated as:

$$\mathbf{A} = \mathbf{A}\mathbf{Z} + \mathbf{E}, \tag{1}$$

where $\mathbf{Z}$ is the subspace representation matrix with each column being the representation of corresponding data point. $\mathbf{E}$ is the noise matrix.

By minimizing the reconstruction loss between $\mathbf{A}$ and $\mathbf{A}\mathbf{Z}$, the general formulation of subspace clustering can be expressed as:

$$\begin{aligned} &\min_{\mathbf{Z}} \mathbf{L}(\mathbf{A}, \mathbf{A}\mathbf{Z}) + \lambda\, \boldsymbol{\Omega}(\mathbf{Z}) \\ &\text{s.t. } 0 \le Z_{i,j} \le 1, \mathbf{Z}^{\mathrm{T}}\mathbf{1} = \mathbf{1}, \end{aligned} \tag{2}$$

where $\mathbf{L}(\cdot)$ and $\boldsymbol{\Omega}(\cdot)$ denote the reconstruction loss function and regularization term respectively. $\lambda > 0$ is a balance parameter. $\mathbf{Z}$ is also called the self-representation matrix, reflecting the similarity among data points. Based on it, constraint $0 \le Z_{i,j} \le 1$ is applied to keep $\mathbf{Z}$ non-negative. Meanwhile, the diagonal elements of $\mathbf{Z}$ is unequal to zero avoiding the trivial solution. It means that each sample can only be represented with the combination of other samples. The constraint $\mathbf{Z}^{\mathrm{T}}\mathbf{1} = \mathbf{1}$ expresses that samples lie in a union of affine subspaces, rather than linear subspaces[34].

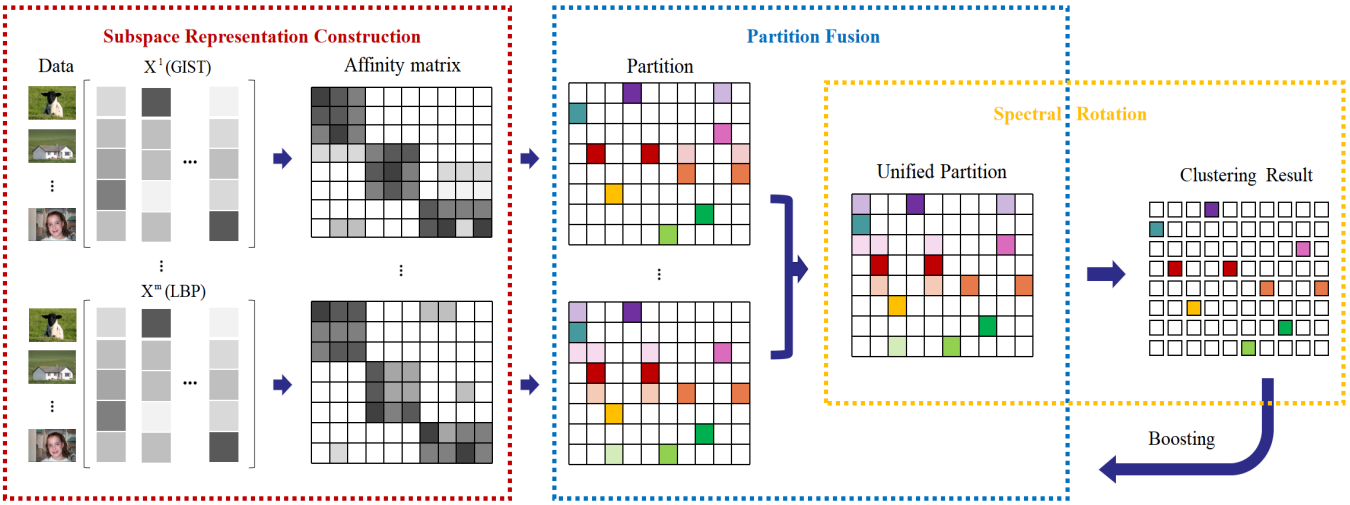


Fig. 1: Framework of proposed COMVSC. With the multi-view data input, COMVSC learns the partition information from corresponding affinity matrix. Then a partion-level fusion method is applied to integrate the complementary information across the multiple views. The final clustering results can directly output by performing spectral rotation on the unified partition. The three processes are integrated into a framework and boosted by each oher.

After that, operations like normalization and symmetrization are imposed on subspace representation $\mathbf{Z}$ in subsequent task to obtain the affinity matrix. Thus, we can perform spectral clustering on the affinity matrix to get the clustering indicator matrix and then the clustering result.

2.3 Multi-view subspace clustering

When datasets have multiple features $\mathbf{X} = [\mathbf{X}^1, \mathbf{X}^2, \cdots, \mathbf{X}^m]^{\mathrm{T}} \in \mathbb{R}^{d \times n}$, problem can be extended into the following mulit-view subspace clustering accordingly.

$$\begin{aligned} &\min_{\mathbf{Z}^v} \mathbf{L}(\mathbf{X}^v, \mathbf{X}^v\mathbf{Z}^v) + \lambda\, \boldsymbol{\Omega}(\mathbf{Z}^v) + \mathbf{Cons}(\mathbf{Z}^v, \mathbf{Z}^*) \\ &\text{s.t. } 0 \le Z^v_{i,j} \le 1, (\mathbf{Z}^v)^{\mathrm{T}}\mathbf{1} = \mathbf{1}, \mathrm{diag}(\mathbf{Z}^v) = 0\ , \end{aligned} \tag{3}$$

$\mathbf{Z}^v \in \mathbb{R}^{n \times n}$ is regarded as the subspace representation matrix of v-th view while $\mathbf{Z}^* \in \mathbb{R}^{n \times n}$ is the consensus subspace representation across multiple views. $\mathbf{Cons}(\cdot)$ are some strategies to reach consensus from several view-specific subspace representations. This is so called similarity-fusion methods.

There are several drawbacks shared by the similarity-fusion methods. Firstly, the subspace representations directly learned from data points are usually full of noise and redundancy. Secondly, although the clustering structure of the different views is similar theoretically, the magnitude of element values in $\mathbf{Z}^v$ are remarkably different[34]. Thirdly, most existing methods usually adopt a two-step strategy, separating the representation learning and clustering process. Commonly, the first step is to utilize subspace learning to get a unified representation. Then traditional clustering methods (e.g., k-means or spectral clustering) are applied to the consensus representation to get the clustering result.

Consequently, the key issue in multi-view subspace clustering is how to utilize multiple representations from different views into a consensus one. There are two common strategies in existing methods to address this problem. One representative strategy, Diversity-induced Multi-view Subspace Clustering(DiMSC) [31] explores the complementary information among multi-view features using Hilbert-Schmidt Independence Criterion(HSIC). Then diverse subspace representations are simply added together as spectral clustering input to generate the final result. Another strategy, such as Multi-view Subspace Clustering (MVSC) [34] strengthens the consistency between different views by performing spectral embedding on them to get a unified clustering indicator matrix.

These multi-view subspace clustering methods have achieved promising performance in real applications. However, rare of the existing approaches make full use of multi-view partition information for improving clustering results. Many real-world data contains noise and outliers, which results in a poor similarity matrix. Nevertheless, the information in the partition level reflects the intrinsic clustering structure. Therefore, considering to fuse multi-view information in partition level is a natural and novel idea. Furthermore, most of the previous approaches are a two-step strategy. They are not able to obtain optimal clustering performance since the similarity learning step is separated from the subsequent clustering step [32]. In order to solve these issues, we propose a novel Consensus One-step Multi-view Subspace Clustering (COMVSC) method to integrate representation learning, partition fusion and clustering into a unified framework. By doing so, the subspace representations we learned are designed for the clustering goal and more informative partition-level information is fused to obtain the final clustering result.

In addition to the multi-view clustering presented in this paper, related literature utilizing multiple information includes multi-task clustering and multi-level clustering.

Multi-task clustering belonging to the field of multi-task learning, refers to strategies that utilize multiple related tasks to promote clustering performance and generalization. A latent representation is shared by different tasks in the learning process, containing the complementary do-

main information learned from each task. A representative work [35] combines multi-task and multi-view learning. The proposed framework contains three parts: within-view-task clustering, multi-view relationship learning, and multi-task relationship learning. Under this framework, a common view that incorporates task-shared features and task-specific features is attained and therefore boosts the clustering performance. Work in [36] is an application of single-view multi-task clustering in the field of environmental science and the examination of air pollution in Turkey.

Multi-level clustering coarsens large-scale or complicated datasets to smaller problems recursively, making issues easier than the original datasets. Then, the partition of the coarsest level successively reverts to the original problem. As an application of multi-level clustering methods on analysis of human intracranial electroencephalogram, Wulsin *et al.* [37] proposes to cluster over the different levels of seizures data to boost the performance of clustering. The experimental results present to approach the manual results by experts.

3 Proposed Approach

In this section, we introduce our novel Consensus One-Step Multi-View Subspace Clustering method and give a unified objective function.

As in the aforementioned multi-view subspace clustering, we can get the subspace representation $\mathbf{Z}^v$ for each view.

$$\begin{aligned}\min_{\mathbf{Z}^v}\sum_{v=1}^{m}\|\mathbf{X}^v-\mathbf{X}^v\mathbf{Z}^v\|_F^2+\lambda\|\mathbf{Z}^v\|_F^2\\ \text{s.t. } 0\le Z_{i,j}^v\le 1,(\mathbf{Z}^v)^{\mathrm{T}}\mathbf{1}=\mathbf{1},\mathrm{diag}(\mathbf{Z}^v)=0\ .\end{aligned}\tag{4}$$

$\mathbf{Z}^v$ can be regarded as affinity matrix, indicating the similarities between data points. The elements in $\mathbf{Z}^v$ should be non-negative and vertically added up to one. As mentioned in Theory 1, an ideal similarity structure for clustering should hold the property that the number of connected components in affinity matrix is equal to the number of clusters. Similarity structure with this property could contribute to the subsequent clustering.

Theorem 1. *The number of connected components in similarity matrix is equal to the multiplicity of eigenvalue 0 of corresponding Laplacian matrix [38].*

Theorem 1 means that the samples can be divided into k clusters if the number of the components in the affinity matrix is exactly equal to k. However, the solution $\mathbf{Z}^v$ learned from Eq. (4) may not satisfy the desirable property. Ideally, if the affinity matrix has k connected components, we can get the rank of corresponding laplacian matrix $\mathbf{L}^v$ is $n-k$. Naturally, we add a rank constraint in Eq. (4) to achieve this condition. The optimization problem becomes:

$$\begin{aligned}\min_{\mathbf{Z}^v}\sum_{v=1}^{m}\|\mathbf{X}^v-\mathbf{X}^v\mathbf{Z}^v\|_F^2+\lambda\|\mathbf{Z}^v\|_F^2\\ \text{s.t. } 0\le Z_{i,j}^v\le 1,(\mathbf{Z}^v)^{\mathrm{T}}\mathbf{1}=\mathbf{1},\\ \mathrm{diag}(\mathbf{Z}^v)=0,\mathrm{rank}(\mathbf{L}^v)=n-k\ ,\end{aligned}\tag{5}$$

where $\mathbf{L}^v=\mathbf{D}^v-\frac{\mathbf{Z}^v+(\mathbf{Z}^v)^{\mathrm{T}}}{2}$, $\mathbf{D}^v$ is the degree matrix of $\mathbf{Z}^v$ whose i-th diagonal element $D_{i,i}^v=\sum_{j=1}^{n}Z_{i,j}^v$.

However, directly applying rank constraint $\mathrm{rank}(\mathbf{L}^v)=n-k$ to Eq. (4) will make the optimization problem hard to tackle. We can transform the rank constrained problem into minimizing $\mathrm{Tr}((\mathbf{F}^v)^{\mathrm{T}}\mathbf{L}^v\mathbf{F}^v)$ thanks to Ky Fan's Theorem [39] $\sum_{i=1}^{k}\sigma_i(\mathbf{L}^v)=\min_{\mathbf{F}^v\mathbf{F}^v=\mathbf{I}_k}\mathrm{Tr}((\mathbf{F}^v)^{\mathrm{T}}\mathbf{L}^v\mathbf{F}^v)$. $\sigma_i(\mathbf{L}^v)$ is the i-th smallest eigenvalues of $\mathbf{L}^v$ and $\mathbf{F}^v\in\mathbb{R}^{n\times k}$ is the clustering indicator matrix of v-th view. It is obvious that the constraint $\mathrm{rank}(\mathbf{L}^v)=n-k$ holds when $\sum_{i=1}^{k}\sigma_i(\mathbf{L}^v)=0$. Therefore, the problem in Eq. (5) can be transformed into the following trace form, which is much easier to solve.

$$\begin{aligned}\min_{\mathbf{F}^v,\mathbf{Z}^v}\sum_{v=1}^{m}\|\mathbf{X}^v-\mathbf{X}^v\mathbf{Z}^v\|_F^2+\lambda\|\mathbf{Z}^v\|_F^2+\mathrm{Tr}((\mathbf{F}^v)^{\mathrm{T}}\mathbf{L}^v\mathbf{F}^v)\\ \text{s.t. } 0\le Z_{i,j}^v\le 1,\ (\mathbf{Z}^v)^{\mathrm{T}}\mathbf{1}=\mathbf{1},\mathrm{diag}(\mathbf{Z}^v)=0,(\mathbf{F}^v)^{\mathrm{T}}\mathbf{F}^v=\mathbf{I}_k\end{aligned}\tag{6}$$

As shown in Eq. (6), each view can get its individual spectral representation $\mathbf{F}^v$. Multi-view clustering holds the assumption that various clustering structures in different views should be analogous to each other [34]. Namely, similar samples should be divided into same cluster no matter from which view. Therefore we enforce each $\mathbf{F}^v$ to align with a consensus $\mathbf{F}^*$ to integrate multiple information across views. Since $\mathbf{F}^v$ provides more discriminative information and less redundancy and noise than the similarity matrix, fusing the partition representation $\mathbf{F}^v$ could achieve superior performance over similarity fusion. Mathematically, the partition fusion term can be formulated as:

$$\begin{aligned}\min_{\mathbf{F}^*,\mathbf{F}^v}\sum_{v=1}^{m}\|\mathbf{F}^v-\mathbf{F}^*\|_F^2,\\ \text{s.t. } (\mathbf{F}^v)^{\mathrm{T}}\mathbf{F}^v=\mathbf{I}_k,\ (\mathbf{F}^*)^{\mathrm{T}}\mathbf{F}^*=\mathbf{I}_k,\end{aligned}\tag{7}$$

After obtaining the consensus partition representation $\mathbf{F}^*$, it is conventional to feed $\mathbf{F}^*$ into k-means to get the final clustering result. However, such a strategy separates the learning of the representation from the final clustering process, which leads to a suboptimal solution. Therefore we introduce a rotation matrix $\mathbf{R}\in\mathbb{R}^{k\times k}$ to jointly optimize the representation and the clustering result. This term can be written as:

$$\begin{aligned}\min_{\mathbf{Y},\mathbf{R},\mathbf{F}^*}\sum_{i=1}^{n}\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\|\mathbf{t}_c-\mathbf{F}_{i,:}^*\mathbf{R}\|_2^2\\ \text{s.t. } Y_{i,c}\ge 0,\ \mathbf{Y}_{i,:}\mathbf{1}_k=1,\mathbf{R}^{\mathrm{T}}\mathbf{R}=\mathbf{I}_k,\end{aligned}\tag{8}$$

where γ is considered as fuzzy coefficient. Similar to one-hot encoding, $\mathbf{t}_c$ is a $1\times k$ dimensional vector used to distinguish each cluster. Specifically, $\mathbf{t}_c$ terms the vector that the c-th element equals to 1 and others are 0, where $c\in\{1,2,\cdots,k\}$. $\mathbf{F}_{i,:}^*$ is the i-th row of $\mathbf{F}^*$, indicating the representation corresponding to the i-th sample. Inspired by [40], $Y_{i,c}$ signify the probability of the i-th sample belonging to the c-th cluster. $\mathbf{R}$ establishes the rational interactions between $\mathbf{Y}$ and $\mathbf{F}^*$. To be specific, matrix $\mathbf{R}$ extracts the distinguished clustering structure of $\mathbf{F}^*$. If the i-th sample representation $\mathbf{F}_{i,:}^*$ shows prominent structure at the c-th position after rotation, the label matrix $\mathbf{Y}$ will has a relatively large probability value at its position of i-th row c-th column.

Through the spectral rotation in Eq. (8), we can avoid performing a subsequent discretization on the consensus

representation and directly output the clustering results in an end-to-end manner. Combining Eq. (6)(7)(8), we can fulfill our COMSC framework as follows:

$$\begin{aligned}\min_{\substack{\mathbf{F}^*,\mathbf{R},\mathbf{Y},\\ \mathbf{F}^v,\mathbf{Z}^v}} & \underbrace{\sum_{v=1}^{m}\|\mathbf{X}^v-\mathbf{X}^v\mathbf{Z}^v\|_F^2+\lambda\|\mathbf{Z}^v\|_F^2+\mathrm{Tr}((\mathbf{F}^v)^{\mathrm{T}}\mathbf{L}^v\mathbf{F}^v)}_{Subspace\ Representation\ Construction}\\ &+\underbrace{\sum_{v=1}^{m}\|\mathbf{F}^v-\mathbf{F}^*\|_F^2}_{Partition\ Fusion}+\underbrace{\sum_{i=1}^{n}\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\|\mathbf{t}_c-\mathbf{F}^*_{i,:}\mathbf{R}\|_2^2}_{Spectral\ Rotation}\\ \text{s.t. } & 0\le Z^v_{i,j}\le 1,(\mathbf{Z}^v)^{\mathrm{T}}\mathbf{1}=\mathbf{1},\mathrm{diag}(\mathbf{Z}^v)=0,\mathbf{R}^{\mathrm{T}}\mathbf{R}=\mathbf{I}_k,\\ & (\mathbf{F}^v)^{\mathrm{T}}\mathbf{F}^v=\mathbf{I}_k,(\mathbf{F}^*)^{\mathrm{T}}\mathbf{F}^*=\mathbf{I}_k,Y_{i,c}\ge 0,\ \mathbf{Y}_{i,:}\mathbf{1}_k=1\ .\end{aligned} \tag{9}$$

In this way, the affinity matrices, consensus partition, and final clustering labels matrix are learned simultaneously in a unified framework. The three steps negotiate with each other to better serve clustering, leading to promising clustering performance.

4 Optimization

The constrained problem in Eq. (9) is difficult to solve directly. In this section, we propose an iterative algorithm to solve this optimization problem efficiently.

4.1 Update subspace representation matrix $\mathbf{Z}^v$

Fixing variables $\mathbf{F}^v$, $\mathbf{F}^*$, $\mathbf{R}$, $\mathbf{Y}$, the optimization for $\mathbf{Z}^v$ can be transformed into solving

$$\begin{aligned}\min_{\mathbf{Z}^v}\ & \|\mathbf{X}^v-\mathbf{X}^v\mathbf{Z}^v\|_F^2+\lambda\|\mathbf{Z}^v\|_F^2+\mathrm{Tr}((\mathbf{F}^v)^{\mathrm{T}}\mathbf{L}^v\mathbf{F}^v)\\ \text{s.t. } & \mathbf{Z}^v\ge 0,(\mathbf{Z}^v)^{\mathrm{T}}\mathbf{1}=\mathbf{1},\mathrm{diag}(\mathbf{Z}^v)=0\ .\end{aligned} \tag{10}$$

In this paper, we adopt a two-step approximation methods to optimize $\mathbf{Z}^v$. We first obtain the closed-form solution without any constraints. Then we apply the constraints on $\mathbf{Z}^v$ to approximate the optimal solution.

Each of $\mathbf{Z}^v$ can be solved separately since views are independent of each other. In the first step, we can rewrite Eq. (10) into Eq. (11) by ignoring the superscript and constraints.

$$\min_{\mathbf{Z}}\|\mathbf{X}-\mathbf{X}\mathbf{Z}\|_F^2+\lambda\|\mathbf{Z}\|_F^2+\mathrm{Tr}(\mathbf{F}^{\mathrm{T}}\mathbf{L}\mathbf{F})\ . \tag{11}$$

Note that $\mathrm{Tr}((\mathbf{F}^{\mathrm{T}}\mathbf{L}\mathbf{F})=\sum_{i,j=1}^{n}\frac{1}{2}Z_{i,j}\|\mathbf{f}_{i,:}-\mathbf{f}_{j,:}\|^2$. We denote $Q_{i,j}=\|\mathbf{f}_{i,:}-\mathbf{f}_{j,:}\|$, therefore $\mathbf{Q}_{i,:}\in\mathbb{R}^{1\times n}$. By setting the derivative of Eq. (11) with respect to $\mathbf{Z}$ to zero, we can get the following closed-form solution:

$$\hat{\mathbf{Z}}=(\mathbf{X}^{\mathrm{T}}\mathbf{X}+\lambda\mathbf{I})^{-1}(\mathbf{X}^{\mathrm{T}}\mathbf{X}-\frac{1}{4}\mathbf{Q}^{\mathrm{T}})\ . \tag{12}$$

In the second step, we can projecting $\hat{\mathbf{Z}}$ into a constrained space. Then the approximate solution of $\mathbf{Z}$ can be derived through the following problem: For each row, we can get

$$\min_{\mathbf{Z}^v_{i,:}\ge 0,(\mathbf{Z}^v_{i,:})^{\mathrm{T}}\mathbf{1}=1,Z_{i,i}=0}\left\|\mathbf{Z}^v_{i,:}-\hat{\mathbf{Z}}^v_{i,:}\right\|_F^2 \tag{13}$$

Then we can obtain the lagrange function of Eq. (13) as

$$\mathcal{L}\left(\mathbf{Z}^v_{i,:},\alpha,\beta\right)=\left\|\mathbf{Z}^v_{i,:}-\hat{\mathbf{Z}}^v_{i,:}\right\|_F^2-\alpha_i\left(\mathbf{Z}^{v\ \mathrm{T}}_{i,:}\mathbf{1}-1\right)-\beta_i^{\mathrm{T}}\mathbf{Z}^v_{i,:}\ , \tag{14}$$

where α_i and $\beta_i\ge 0$ is the lagrangian multipliers. And then according to KKT condition and some calculation [41], [42], we can easily obtain

$$\mathbf{Z}^v_{i,:}=\max\left(\hat{\mathbf{Z}}^v_{i,:}+\alpha_i\mathbf{1},0\right),\mathbf{Z}^v_{i,i}=0,\alpha_i=\frac{1+(\hat{\mathbf{Z}}^v_{i,:})^{\mathrm{T}}\mathbf{1}}{n-1}\ . \tag{15}$$

4.2 Update consensus representation $\mathbf{F}^*$

With $\mathbf{Z}^v$, $\mathbf{F}^v$, $\mathbf{R}$, $\mathbf{Y}$ being fixed, the optimization problem for $\mathbf{F}^*$ can be simplified as

$$\begin{aligned}\min_{\mathbf{F}^*}\ & \sum_{v=1}^{m}\|\mathbf{F}^v-\mathbf{F}^*\|_F^2+\sum_{i=1}^{n}\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\|\mathbf{t}_c-\mathbf{F}^*_{i,:}\mathbf{R}\|_2^2\\ \text{s.t. } & (\mathbf{F}^*)^{\mathrm{T}}\mathbf{F}^*=\mathbf{I}_k\end{aligned} \tag{16}$$

Proposition 1. *The minimum problem of* $\min\sum_{i=1}^{n}\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\|\mathbf{t}_c-\mathbf{F}^*_{i,:}\mathbf{R}\|_2^2$ *is equivalent to* $\max\ \mathrm{Tr}(\mathbf{R}^{\mathrm{T}}(\mathbf{F}^*)^{\mathrm{T}}\mathbf{G})$, *where* $\mathbf{G}\in\mathbb{R}^{n\times k}$, *and its i-th row* $\mathbf{G}_{i,:}=\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\mathbf{t}_c$.

Proof. Equation $\min\sum_{i=1}^{n}\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\|\mathbf{t}_c-\mathbf{F}^*_{i,:}\mathbf{R}\|_2^2$ equals to $\max\sum_{i=1}^{n}\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\ \mathrm{Tr}(\mathbf{R}^{\mathrm{T}}(\mathbf{F}^*_{i,:})^{\mathrm{T}}\mathbf{t}_c)$

It is easy to obtain that $\mathrm{Tr}(\mathbf{R}^{\mathrm{T}}(\mathbf{F}^*_{i,:})^{\mathrm{T}}t_c)=\mathbf{R}^{\mathrm{T}}_{c,:}(\mathbf{F}^*)^{\mathrm{T}}_{:,i}$. Therefore the equation can be rewritten as:

$$\max\sum_{i=1}^{n}\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\mathbf{R}^{\mathrm{T}}_{c,:}(\mathbf{F}^*)^{\mathrm{T}}_{:,i} \tag{17}$$

By expanding Eq. (17) element-wise, we can get: $\sum_{i=1}^{n}\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\mathbf{R}^{\mathrm{T}}_{c,:}(\mathbf{F}^*)^{\mathrm{T}}_{:,i}=\mathrm{Tr}(\mathbf{R}^{\mathrm{T}}(\mathbf{F}^*)^{\mathrm{T}}\mathbf{G})$, where $\mathbf{G}_{i,:}=\sum_{c=1}^{k}(Y^{\gamma}_{i,c})\mathbf{t}_c$.

Consequently, $\min\sum_{i=1}^{n}\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\|\mathbf{t}_c-\mathbf{F}^*_{i,:}\mathbf{R}\|_2^2$ is equivalent to $\max\ \mathrm{Tr}(\mathbf{R}^{\mathrm{T}}(\mathbf{F}^*)^{\mathrm{T}}\mathbf{G})$, where $\mathbf{G}\in\mathbb{R}^{n\times k}$, and its i-th row $\mathbf{G}_{i,:}=\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\mathbf{t}_c$. □

Proposition 2. *We can easily get the the economic rank-k SVD of* $\mathbf{B}$ *is* $\mathbf{B}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$. *Accordingly, the constrained problem*

$$\max_{\mathbf{A}}\ \mathrm{Tr}(\mathbf{A}^{\mathrm{T}}\mathbf{B})\ s.t.\ \ \mathbf{A}^{\mathrm{T}}\mathbf{A}=\mathbf{I} \tag{18}$$

has closed form solution:

$$\mathbf{A}=\mathbf{U}\mathbf{V}^{\mathrm{T}}$$

Proof. By taking the the normal singular value decomposition $\mathbf{B}=\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}}$, we can get :

$$\mathrm{Tr}(\mathbf{A}^{\mathrm{T}}\mathbf{B})=\mathrm{Tr}(\mathbf{A}^{\mathrm{T}}\mathbf{U}\mathbf{\Sigma}\mathbf{V}^{\mathrm{T}})=\mathrm{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{U}\mathbf{\Sigma})$$

Setting $\mathbf{Q}=\mathbf{V}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{U}$, we have $\mathbf{V}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{U}\mathbf{U}^{\mathrm{T}}\mathbf{A}\mathbf{V}=\mathbf{I}$. Therefore we get $\mathrm{Tr}(\mathbf{V}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{U}\mathbf{\Sigma})=\mathrm{Tr}(\mathbf{Q}\mathbf{\Sigma})\le\mathrm{Tr}(\mathbf{I}\mathbf{\Sigma})=\Sigma_{i=1}^{k}\sigma_i$, where σ_i is the i-th diagonal element of Σ.

The solution of maximize Eq. (18) can be obtained when $\mathbf{Q}\mathbf{\Sigma}=\mathbf{I}\mathbf{\Sigma}$, that is $\mathbf{V}^{\mathrm{T}}\mathbf{A}^{\mathrm{T}}\mathbf{U}=\mathbf{I}$, so we can get the closed solution $\mathbf{A}=\mathbf{U}\mathbf{V}^{\mathrm{T}}$. □

According to proposition 1, the problem of Eq. (16) can be transformed into the follow matrix form:

$$\begin{aligned}\min_{\mathbf{F}^*}\ & \sum_{v=1}^{m}-2\mathrm{Tr}((\mathbf{F}^v)^{\mathrm{T}}\mathbf{F}^*)-\mathrm{Tr}(\mathbf{R}^{\mathrm{T}}(\mathbf{F}^*)^{\mathrm{T}}\mathbf{G})\\ \text{s.t. } & (\mathbf{F}^*)^{\mathrm{T}}\mathbf{F}^*=\mathbf{I}_k.\end{aligned} \tag{19}$$

Eq. (19) can be transformed into Eq. (20) by simple transformation.

$$\max_{\mathbf{F}^*} \mathrm{Tr}((\mathbf{F}^*)^{\mathrm{T}}\mathbf{N}) \quad \text{s.t. } (\mathbf{F}^*)^{\mathrm{T}}\mathbf{F}^* = \mathbf{I}_k, \tag{20}$$

where $\mathbf{N} = 2\sum_{v=1}^{m}\mathbf{F}^v + \mathbf{G}\mathbf{R}^{\mathrm{T}}$. According to proposition 2 and its proof, we can get the closed-form optimal solution $\mathbf{F}^* = \mathbf{U}_1\mathbf{V}_1^{\mathrm{T}}$, where $\mathbf{U}_1$ and $\mathbf{V}_1$ are left singular matrix and right singular matrix of matrix $\mathbf{N}$ respectively.

4.3 Update spectral representations $\mathbf{F}^v$

By fixing the other variables and removing terms that are irrelevant to $\mathbf{F}^v$, the optimization for $\mathbf{F}^v$ can be transformed into solving the following problem:

$$\begin{aligned}\min_{\mathbf{F}^v} \ & \mathrm{Tr}((\mathbf{F}^v)^{\mathrm{T}}\mathbf{L}^v\mathbf{F}^v) - 2\mathrm{Tr}((\mathbf{F}^v)^{\mathrm{T}}\mathbf{F}^*) \\ & \text{s.t. } (\mathbf{F}^v)^{\mathrm{T}}\mathbf{F}^v = \mathbf{I}_k\end{aligned} \tag{21}$$

Equation Eq. (21) can be relaxed into the following form, where λ_{max} is the largest eigenvalue of $\mathbf{L}^v$.

$$\begin{aligned}\max_{\mathbf{F}^v} \ & \mathrm{Tr}((\mathbf{F}^v)^{\mathrm{T}}(\lambda_{max}\mathbf{I} - \mathbf{L}^v)\mathbf{F}^v) + 2(\mathrm{Tr}(\mathbf{F}^v)^{\mathrm{T}}\mathbf{F}^*) \\ & \text{s.t. } (\mathbf{F}^v)^{\mathrm{T}}\mathbf{F}^v = \mathbf{I}_k\end{aligned} \tag{22}$$

Proposition 3. *If $\mathbf{A}$,$\mathbf{B}$ are positive semidefinite matrices, $f(\mathbf{W}) = \mathrm{Tr}(\mathbf{W}^{\mathrm{T}}\mathbf{A}\mathbf{W}\mathbf{B}) + \mathrm{Tr}(\mathbf{W}^{\mathrm{T}}\mathbf{C})$ is a convex function. Problem can be solved by optimizing $\max_{\mathbf{W}^{\mathrm{T}}\mathbf{W}=\mathbf{I}} \mathrm{Tr}(\mathbf{W}^{\mathrm{T}}\mathbf{M})$ iteratively, where $\mathbf{M} = f'(\mathbf{W}) = 2\mathbf{A}\mathbf{W}\mathbf{B} + \mathbf{C}$. [43]*

According to proposition 3 and proposition 2, convex function $f(\mathbf{F}^v) = \mathrm{Tr}((\mathbf{F}^v)^{\mathrm{T}}(\lambda_{max}\mathbf{I} - \mathbf{L}^v)\mathbf{F}^v) + 2(\mathrm{Tr}(\mathbf{F}^v)^{\mathrm{T}}\mathbf{F}^*)$ can be solved by the following algorithm.

Algorithm 1 Update $\mathbf{F}^v$

1: **while** not converged **do**
2: $\quad\mathbf{M} = 2(\lambda_{max}\mathbf{I} - \mathbf{L}^v)\mathbf{F}^v + 2\mathbf{F}^*$.
3: $\quad$Solve $\max_{(\mathbf{F}^v)^{\mathrm{T}}\mathbf{F}^v=\mathbf{I}} \mathrm{Tr}((\mathbf{F}^v)^{\mathrm{T}}\mathbf{M})$ to update $\mathbf{F}^v$.
4: $\quad\quad$Perform SVD on $\mathbf{M}$, $\mathbf{M} = \mathbf{U}_2\Sigma\mathbf{V}_2^{\mathrm{T}}$
5: $\quad\quad\mathbf{F}^v = \mathbf{U}_2\mathbf{V}_2^{\mathrm{T}}$
6: **end while**

4.4 Update spectral rotation matrix $\mathbf{R}$

Fixing variables like $\mathbf{Z}^v$, $\mathbf{F}^v$, $\mathbf{F}^*$, $\mathbf{Y}$ and removing terms that are irrelevant to $\mathbf{R}$, the optimization for $\mathbf{R}$ can be transformed into solving the following problem:

$$\begin{aligned}\min_{\mathbf{R}} \ & \sum_{i=1}^{n}\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\|\mathbf{t}_c - \mathbf{F}^*_{i,:}\mathbf{R}\|_2^2 \\ & \text{s.t. } \mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}_k.\end{aligned} \tag{23}$$

As showed in proposition 1, this problem is equivalent to :

$$\max_{\mathbf{R}} \mathrm{Tr}(\mathbf{R}^{\mathrm{T}}(\mathbf{F}^*)^{\mathrm{T}}\mathbf{G}) \quad \text{s.t. } \mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}_k, \tag{24}$$

where $\mathbf{G} \in \mathbb{R}^{n\times k}$ and its i-th row is $\mathbf{g}_{i,:} = \sum_{c=1}^{k}(Y_{i,c})^{\gamma}\mathbf{t}_c$. Denoting $(\mathbf{F}^*)^{\mathrm{T}}\mathbf{G}$ as $\mathbf{H}$, we can get

$$\max_{\mathbf{R}} \mathrm{Tr}(\mathbf{R}^{\mathrm{T}}\mathbf{H}) \quad \text{s.t. } \mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}_k \tag{25}$$

The optimal solution $\mathbf{R} = \mathbf{U}_3\mathbf{V}_3^{\mathrm{T}}$ can be derived from proposition 2, where $\mathbf{U}_3$ and $\mathbf{V}_3$ are left singular matrix and right singular matrix of $\mathbf{H}$ respectively.

4.5 Update probability labels matrix $\mathbf{Y}$

Fixing variables like $\mathbf{Z}^v$, $\mathbf{F}^v$, $\mathbf{F}^*$, $\mathbf{R}$ and removing terms that are irrelevant to $\mathbf{Y}$, the optimization for each $\mathbf{Y}_{i,:}$, the i-th row of the matrix $\mathbf{Y}$, can be transformed into solving the following problem:

$$\begin{aligned}\min & \sum_{c=1}^{k}(Y_{i,c})^{\gamma}\|\mathbf{t}_c - \mathbf{F}^*_{i,:}\mathbf{R}\|_2^2 \\ & \text{s.t. } Y_{i,c} \geq 0, \mathbf{Y}_{i,:}\mathbf{1}_k = 1\end{aligned} \tag{26}$$

We denote $P_{i,c} = \|\mathbf{t}_c - \mathbf{F}^*_{i,:}\mathbf{R}\|_2^2$ as the element of the i-th row and c-th column of matrix $\mathbf{P}$. Then the optimization function can be rewritten as $\min_{Y_{i,c}\geq 0, \mathbf{Y}_{i,:}\mathbf{1}_k=\mathbf{1}} \sum_{c=1}^{k}(Y_{i,c})^{\gamma}P_{i,c}$.

When $\gamma = 1$, the optimal solution of Eq. (26) can be formulated as: $Y_{i,c} =< c = \arg_j \min P_{i,j} >$, where $< \cdot >$ is 1 if the argument is true or 0 otherwise.

When $\gamma > 1$, we can get the following closed-form solution by setting the derivative of its Lagrangian function with respect to $Y_{i,c}$ to zero.

$$Y_{i,c} = \frac{(P_{i,c})^{\frac{1}{1-\gamma}}}{\sum_{c=1}^{k}(P_{i,c})^{\frac{1}{1-\gamma}}} \tag{27}$$

The entire optimization is summarized in Algorithm 2. The objective of Algorithm 2 is monotonically decreased when optimizing one variable with the others fixed at each iteration. At the same time, the whole optimization problem is lower-bounded. As a result, the proposed algorithm can be verified to be convergent.

Algorithm 2 COMVSC

Input: Data points in v views $\{\mathbf{X}^v\}_{v=1}^m$, the number of cluster k, hyper-parameters λ and γ.
Output: Probability clustering labels $\mathbf{Y}$
Initialize:Initialize $\mathbf{F}^v$ with the eigenvectors of corresponding laplacian matrix. Randomly initialize the orthogonal matrix $\mathbf{F}^*$. Initialize rotation matrix $\mathbf{R}$ with $k \times k$ dimensional identity matrix. Initialize label matrix $\mathbf{Y}$ with only one 1 in each row.

1: **while** not converged **do**
2: $\quad$Update $\mathbf{Z}^v$ by Eq. (12).
3: $\quad$Update $\mathbf{F}^*$ by solving Eq. (20).
4: $\quad$Update $\mathbf{F}^v$ by solving Eq. (21).
5: $\quad$Update $\mathbf{R}$ by solving Eq. (24).
6: $\quad$Update $\mathbf{Y}$ by Eq. (27).
7: **end while**
8: **return** clustering labels $\mathbf{Y}$. In each row, the column number corresponding to the largest element is exactly the cluster to which the data point belongs.

5 Analysis and Discussions

In this section, we analyze the computational complexity and give some discussions of the proposed method.

Computational Complexity: With the optimization process outlined in Algorithm 2, the computational complexity of COMVSC consists of three sub-processes. In first step (similarity learning), the main computational complexity of matrix inverse operation $(\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I})^{-1}$ is $\mathcal{O}(mn^3)$. Then

the updating process of $\mathbf{F}^*$ costs $\mathcal{O}(n^3)$ while updating $\mathbf{F}^v$ is $\mathcal{O}(t_1 mnk^2)$, where t_1 is the number of iterations of Algorithm 1 . In order to update $\mathbf{R}$ and $\mathbf{Y}$, we need $\mathcal{O}(mk^3)$ and $\mathcal{O}(mnk^2 + nk)$. Overall, the complexity of our Algorithm 2 is $\mathcal{O}(Tmn^3)$, where T is the number of total iterations.

Discussion: Our proposed COMVSC enjoys several advantages. Firstly, COMVSC fuses multiple subspace information in partition level since every individual partition captures the local clustering structure in its respective view. Furthermore, comparing to similarity-fusion methods, it is much easier and more reasonable to reach an agreement in partition matrices. Secondly, our method is considered as an end-to-end framework which can directly output the clustering labels. Therefore, the joint manner ensures that the learned representation best serves the clustering task.

Our method is limited by the following two points. The first one is the cubic computational complexity. It may limit the scalability of the proposed algorithm. There are some ways to accelerate our calculations, such as parallelization, a fast solution for $\mathbf{F}^v$, fast SVD and automatic convergence condition. All of the three factors could help reduce the running time to some degree. The second point is the capacity to deal with the data with imbalanced categories needs to be improved. We will later consider balancing the categories by performing data augmentation on small clusters or sampling from large clusters, or utilizing the strategy of weighted samples.

6 Experiments

In this section, we extensively evaluate the clustering property of the proposed method on seven widely used multi-view benchmark datasets. The performance of COMVSC is compared with a single-view clustering algorithm and six state-of-the-art multi-view methods in terms of three clustering evaluation metrics.

6.1 Datasets Description

Seven public multi-view benchmark datasets including BBCSport [44], yaleA [45], Cornell [46], MSRC-v1 (Microsoft Research Cambridge Volume 1) [47], Wikipedia Articles [48], Webkb [49] and Handwritten are used in our experiments. Specifically, the key information of them is summarized in Table 1 .

BBCSport is a subset dataset with 116 samples chosen from the original sports news database in five topical areas. Each document is split into four related segments as features, each consisting of successive paragraphs of text.

yaleA is a widely used dataset of face images, consisting of 165 images of fifteen different people. Each person has 11 grayscale images, and these images reflect different facial expressions of the object or configurations. Each of image has three features: 9-dimensional color moment, 512-dimensional GIST [50] and 50-dimensional LBP (Local Binary Pattern) [51].

Cornell is a web page dataset collected from Cornell University. There are 195 pages in five categories. Content features and cites features are utilized to describe each page. To be specific, the content view contains 1073 words to indicate the absence and presence of a word on a page. The cites view reflects the number of citation links between this page and others.

MSRC-v1 is a scene recognition dataset containing 240 images with each category 30 samples. We select 7 classes (tree, car, face, cow, bicycle, building and airplane) totally 210 images from them and extract 1302-dimensional CENT (Census Transform) [52], 512-dimensional GIST, 256-dimensional LBP, 210-dimensional SIFT (Scale Invariant Feature Transform) [52], 100 dimensional HOG (Histogram of Oriented Gradient) [53], 48-dimensional CMT (Color Moment) [54] features from each image.

Wikipedia Articles is a widely used dataset for cross-modal retrieval, which consists of 693 samples in 10 categories. 128-dimensional SIFT features for images and 10-dimensional features for text deriving from an LDA (Latent Dirichlet Allocation) model are extracted from the data.

Webkb is also a dataset of web pages. They were collected by the World Wide Knowledge Base project. There are 1051 pages divided into two categorize. Each instant can be interpreted by two views, the same as the Cornell dataset.

Handwritten contains 2000 images of '0'-'9' handwritten digits. Each of these classes has 200 images, described by six features, listed in the following Table 1 .

6.2 Compared Methods

We compare our proposed COMVSC with the following methods, including a baseline and 11 state-of-the-art multi-view subspace clustering algorithms.

FeatureConcate (**FeaCon**) is regarded as a baseline method. It concatenates the features from all views directly and then performs k-means to get the final result. Co-regularized multi-view spectral clustering [55] utilizes a co-regularization term to make the partitions in different views agree with each other. Two schemes, i.e., centroid-based method (**Co-reg_c**) and pairwise method (**Co-reg_p**), are proposed to accomplish this goal. Multi-view Low-Rank Sparse Subspace Clustering (**MLRSSC**) [29] learns a joint subspace representation across all the views by conducting sparsity and low-rank constraint on each affinity matrix. Latent Multi-view Subspace Clustering (**LMSC**) [56] conducts subspace clustering on latent representation learned from multi-view features to generate a common subspace representation. Robust Multi-view K-Means clustering (**RMKMC**) [9] integrates multiple representations adaptively and induces structured sparsity-inducing norm to make it more robust to outliers. Multiple Partition Aligned Clustering (**mPAC**) [13] learns the affinity matrix and obtains the clustering result by assigning each partition with a respective rotation matrix. Flexible Multi-view Representation Learning for Subspace Clustering (**FMR**) [1] utilizes HSIC to flexibly enforce different views to be close to a latent representation. With other additional constraints, FMR learns a more comprehensive and suitable for subspace clustering. Graph-based multi-view clustering (**GMC**) [57] adaptively fuses multiple graph matrices to generate a unified graph matrix. The unified graph guides the optimization of the graph matrices and also gives the clustering indicator matrix. Large-scale Multi-view Subspace Clustering in Linear Time (**LMVSC**) [58] integrates multiple anchor

TABLE 1: Statistic of seven benchmark multi-view datasets.

view	BBCSport	yaleA	Cornell	MSRC-v1	Wikipedia	Webkb	Handwritten
1	Segment1 (1991)	Color moment (9)	Cites (195)	CENT (1320)	SIFT (128)	Cites (1840)	Profile correlations (216)
2	Segment2 (2063)	GIST (512)	Content (1703)	GIST (512)	LDA (10)	Content (3000)	Fourier coefficients (76)
3	Segment3 (2113)	LBP (50)	-	LBP (256)	-	-	Karhunen coefficients (64)
4	Segment4 (2158)	-	-	SIFT (210)	-	-	Morphological (6)
5	-	-	-	HOG (100)	-	-	Pixel averages (240)
6	-	-	-	CMT (48)	-	-	Zernike moments (47)
Views	4	3	2	6	2	2	6
Clusters	5	15	5	7	10	2	10
Points	116	165	195	210	693	1051	2000

graphs and performs spectral clustering on the small graph, which makes the runtime linear. Partition level multiview subspace clustering (**PMSC**) [59] implements the graph construction, the generation of basic partitions, and fusion of consensus clustering in an interactive way.

For the above methods, the parameters are tuned as suggested in their papers to generate the best results. To evaluate the clustering performance, three metrics: ACC (Accuracy), NMI (Normalized Mutual Information), and F-score, are reported in this paper. Notably, higher values indicate better performance.

Denoting q_i as the clustering result and p_i as the true label of data point x_i, ACC is defined as follows:

$$\text{ACC} = \frac{\sum_{i=1}^{n} \delta(p_i, map(q_i))}{n} \tag{28}$$

where $\delta(x, y) = 1$ if $x = y$, otherwise $\delta(x, y) = 0$. $map(q_i)$ is the best mapping function that permutes clustering labels to match the true labels using the KuhnMunkres algorithm.

Given two variables P and Q, NMI is defined as

$$\text{NMI}(P, Q) = \frac{I(P, Q)}{\sqrt{(H(P)H(Q))}}, \tag{29}$$

where $H(P)$ and $H(Q)$ are the entropies of P and Q, respectively, and $I(P, Q)$ is the mutual information between P and Q. For clustering, P and Q are the clustering results and the true labels, respectively. NMI reflects the consistency between clustering results and ground truth labels.

F-score is denotd by precision and recall. We take F_1-score in this paper:

$$\text{F}_1\text{-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}, \tag{30}$$

where $\text{Precision} = \frac{\text{TP}}{\text{TP+FP}}$ and $\text{Recall} = \frac{\text{TP}}{\text{TP+FN}}$. Note that TP is true positive, TN is true negative, FP is false positive and FN is false negative.

6.3 Experiment Results and Analysis

The three evaluation metrics (ACC, NMI, and F-score) of the compared algorithms on the seven real-world datasets are displayed in Table 2. The best result is highlighted in red, while the second-best is reported in blue.

Experiment analysis. From the overall results of the seven datasets, our method achieves the best results for all the other datasets and all evaluation criteria, except for the suboptimal NMI on the Handwritten dataset. In terms of accuracy, our method exceeded the second-best results on the datasets BBCSport, yaleA,Cornell, MSRC-v1, Wikipedia, Webkb, and Handwritten by 8.63%, 16.37%, 2.57%, 6.19%, 3.18%, 1.8% and 5.6%, respectively. The F-score exceeds the next best algorithm by 2.15%, 20.69%, 2.84%, 11.74%, 2.53%, 1.24% and 2.81% for the corresponding dataset, respectively. As for NMI, our method is only 1.16% below the best method on Handwritten but surpasses the third algorithm by 5.64%. The above results demonstrate the validity and sophistication of our proposed method and clearly show that our COMVSC method is a valuable method for multi-view subspace clustering.

In most cases, baseline FeaCon performs the worst in comparison with the other methods. This indicates the effectiveness of multi-view clustering methods on exploiting the complementary information across views. It is also claimed that although individual view could be used for finding clustering pattern, the clustering performance will be more accurate by exploring information across multiple views.

Compared with similarity-fused methods like GMC and LMVSC and the multi-view k-means methods RMKMC, our approach and mPAC could achieve more impressive performance. This mainly because the partition-level information is more informative and less noisy than the similarity-level information. More informative multiple partition information is utilized to serve the clustering task and therefore gain better performance. This proves that fusing partition information is a practical approach in dealing with multi-view clustering problems.

Multi-view subspace clustering methods like LMSC and FMR learn a new representation in latent space, which avoid fusing information in noisy original space. However, there is a large gap between these methods and ours and mPAC method, probably because they separate the representation learning and clustering processes, resulting in a sub-optimal representation. Compared with mPAC and PMSC, we are able to achieve higher clustering performance with fewer hyper-parameters. As for ACC, our performance is 8.62%, 6.37%, 2.57%, 4.28%, 6.79%, 1.32% and 5.60% superior to mPAC on each dataset, respectively. This is attributed to our soft assignment strategy and rational constraints for $\mathbf{Z}^v$.

The above experimental results have well demonstrated the effectiveness of our proposed COMVSC in comparison with other state-of-the-art methods. We summarize the superiority of the proposed approach in two aspects: 1) COMVSC employs a joint fusion to optimize self-representation, partition matrices, and clustering labels. To be specific, when more accurate clustering labels are obtained during one iteration, we could further use high-quality labels to guide the generation of representations in

TABLE 2: The clustering performance of compared methods on seven datasets

Datasets	Metrics	FeaCon	Co-reg_c [55]	Co-reg_p [55]	MLRSSC [29]	LMSC [56]	RMKMC [9]	mPAC [13]	FMR [1]	GMC [57]	LMVSC [58]	PMSC [59]	Ours
BBCSport	ACC	0.3448	0.5069	0.5414	0.4086	0.4741	0.3190	0.6121	0.3793	0.5603	0.4828	0.3138	**0.6983**
	NMI	0.2775	0.3098	0.3241	0.2113	0.3093	0.0333	0.4894	0.0995	0.4771	0.2720	0.0653	**0.5346**
	F-score	0.3864	0.4385	0.4403	0.3901	0.3890	0.2239	0.5107	0.3689	0.4439	0.3934	0.3830	**0.5322**
YALE	ACC	0.5091	0.6806	0.7230	0.2539	0.7394	0.6121	0.7636	0.6303	0.6788	0.4849	0.1176	**0.9152**
	NMI	0.5730	0.7506	0.7840	0.3738	0.7831	0.6584	0.7812	0.7367	0.7324	0.5535	0.1760	**0.9036**
	F-score	0.4222	0.5901	0.6278	0.2271	0.6482	0.4590	0.6371	0.5437	0.4550	0.3328	0.1212	**0.8271**
Cornell	ACC	0.3564	0.4256	0.4031	0.3864	0.3487	0.4308	0.5692	0.4000	0.3692	0.4410	0.4451	**0.5949**
	NMI	0.0753	0.1587	0.1148	0.1189	0.0810	0.1491	0.2898	0.1747	0.1389	0.1677	0.0589	**0.3008**
	F-score	0.3288	0.3416	0.3389	0.3031	0.2859	0.3376	0.4741	0.3237	0.3549	0.3896	0.4311	**0.5025**
MSRC-v1	ACC	0.4541	0.7991	0.6933	0.4852	0.7476	0.7095	0.8143	0.8524	0.8952	0.8810	0.6052	**0.9571**
	NMI	0.4217	0.6991	0.6032	0.3642	0.6450	0.6167	0.7508	0.7490	0.8200	0.8111	0.5695	**0.9120**
	F-score	0.3856	0.6827	0.5866	0.4021	0.6384	0.5894	0.7320	0.7381	0.7997	0.7665	0.5016	**0.9171**
Wikipedia	ACC	0.2338	0.4760	0.2245	0.3033	0.4141	0.5743	0.5382	0.5556	0.3939	0.5137	0.2003	**0.6061**
	NMI	0.1000	0.3793	0.0983	0.1773	0.3297	0.5345	0.4766	0.5249	0.3838	0.4746	0.0700	**0.5361**
	F-score	0.1492	0.3661	0.1452	0.1945	0.2964	0.4810	0.4450	0.4741	0.2561	0.4114	0.1934	**0.5063**
Webkb	ACC	0.5821	0.7830	0.8979	0.9049	0.9163	0.7812	0.8211	0.5423	0.7869	0.8820	0.6179	**0.9343**
	NMI	0.0054	0.4584	0.4118	0.5334	0.5277	0.0435	0.3964	0.0069	0.0312	0.4128	0.0802	**0.5958**
	F-score	0.5940	0.7942	0.8790	0.8608	0.8917	0.7936	0.7774	0.5707	0.7961	0.8589	0.6582	**0.9041**
Handwritten	ACC	0.5830	0.8188	0.7912	0.3702	0.6135	0.6710	0.8890	0.6660	0.8820	0.8540	0.8381	**0.9450**
	NMI	0.6151	0.7785	0.7574	0.5141	0.5996	0.6533	0.8361	0.6572	**0.9041**	0.8081	0.8226	0.8925
	F-score	0.5649	0.7450	0.7209	0.4163	0.5284	0.5922	0.8215	0.5890	0.8653	0.7757	0.8150	**0.8934**

the next iteration and further improve the performance. 2) Comparing with the existing similarity-fusion methods, the proposed COMVSC fuses multiple subspace information in partition level and learns a consensus partition, verifying the advantages of combing high-level information more informative and less noise and redundancy. These two factors contribute to the significant improvements in clustering performance.

Statistical Analysis. To demonstrate the statistical property of our proposed method, we conduct the Friedman test and Nemenyi post-hoc test.

The Friedman test assumes that all of the h compared algorithms hold the same performance on D datasets. Specifically, this test includes two main step. The first step is calculating $\tau_{\mathcal{X}^2}$ and τ_F according Eq. (31),

$$\tau_F = \frac{(D-1)\tau_{\mathcal{X}^2}}{D(h-1)-\tau_{\mathcal{X}^2}}, \tag{31}$$

where $\tau_{\mathcal{X}^2} = \frac{h-1}{h}\frac{12D}{h^2-1}(\sum_{i=1}^{h} r_i - \frac{h+1}{2})^2$, r_i represents the average rank of the i-th algorithm over all datasets. Besides, τ_F obeys the F-distribution with the degree of freedom $h-1$ and $(h-1)(D-1)$.

The second step is eliminating whether the assumption is true or not by comparing the τ_F and its corresponding threshold. If the assumption is denied, it indicates that the performance of the compared methods is significantly different. Then a post-hoc test is required to further distinguish the algorithms. The Nemenyi test is a common post-hoc test.

The Nemenyi calculates the critical distance by Eq. (32) to reflect the difference between the average ordinal results of various methods.

$$CD = q_\alpha \sqrt{\frac{h(h+1)}{6D}}, \tag{32}$$

where q_α can be calculate by $qtukey(1-\alpha, h, \text{Inf})/sqrt(2)$ in R programming language.

In our case, the number of compared methods h equals to 12 and N equals to 7. We sort the ACC, NMI and F-score of the compared methods from high to low and obtain the average ranking of each algorithm in terms of all datasets. Especially, the equal performance of the two algorithms would result in equal ordinal values.

According to Eq. (31), the τ_F value of our proposed algorithm is 6.5424, which is larger than the threshold 1.9370 when $\alpha = 0.05$. This rejects the assumption that all of the compared algorithms hold the same performance. Then, we perform the Nemenyi post-hoc test to further distinguish multiple methods. After obtaining $CD = 6.2982$ according to Eq. (32), we can draw the Friedman test chart as Fig. 2.

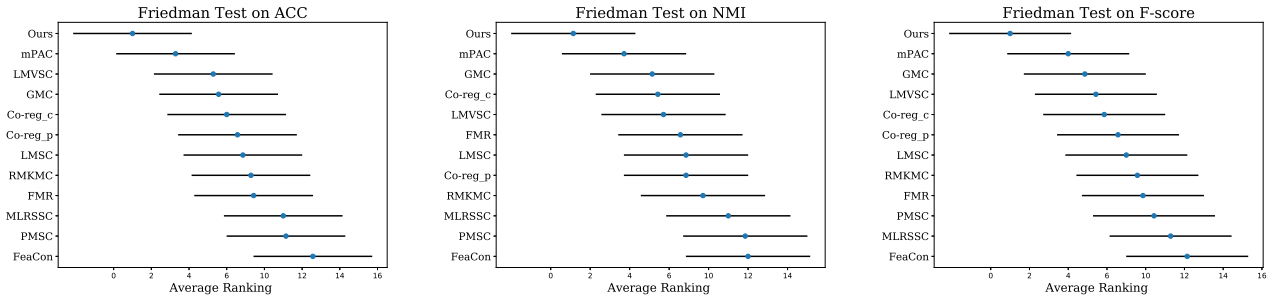
From this figure, we come to the conclusion that the proposed method is significantly different from RMKMC, FMR, MLRSSC, PMSC and FeaCon in terms of ACC, NMI and F-score. mPAC only has a significant difference with the baseline FeaCon. And the other methods do not differ from any methods significantly. As also can be seen in Fig. 2, our method holds the best average ranking in comparison with other methods no matter under which metric. In summary, our proposed method holds well statistical superior.

Running time Analysis. We record the running time of compared algorithms on the benchmark datasets and report them in Fig. 3. Our running times are comparable on other datasets, except on yaleA, which is slightly higher than mPAC by 6.2 seconds. The fastest and most suitable for large-scale datasets are RMKMC, LMVSC and GMC. These methods are designed to solve large-scale problems. They are more concerned with efficiency than effectiveness and thus suffer relatively poor clustering performance.

Although the proposed algorithm is slower than algorithms specifically designed for large-scale scenarios, our COMVSC could achieve significant improvements in clustering performance compared to them. Besides, the running time of our method could be further reduced by using parallelization, a fast solution for updating $\mathbf{F}^v$ [60], fast SVD, and automatic convergence condition.

6.4 Evolution

We use t-Distributed Stochastic Neighbor Embedding(t-SNE) to visualize the structure of the probability labels matrix we learned. For example, on Handwritten, as can be seen in Fig. 4, as the algorithm is iterated, the clustering



(a) Friedman Test on ACC (b) Friedman Test on NMI (c) Friedman Test on F-score

Fig. 2: Friedman Test Charts . For each algorithm, the blue dot marks its average rank. The horizontal lines with the dot as the center indicate the critical distance. No overlapping areas of the lines indicating a significant difference.

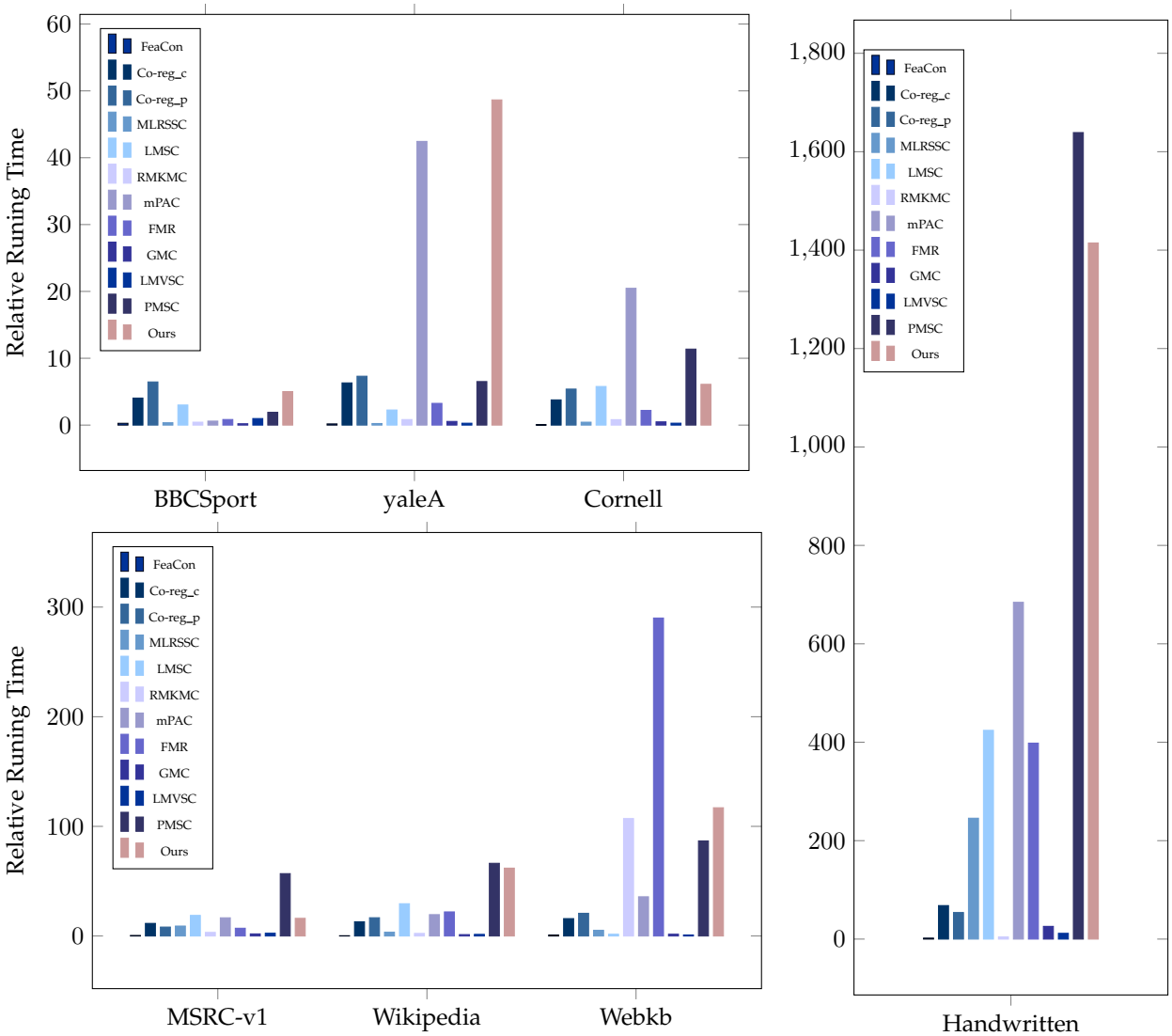
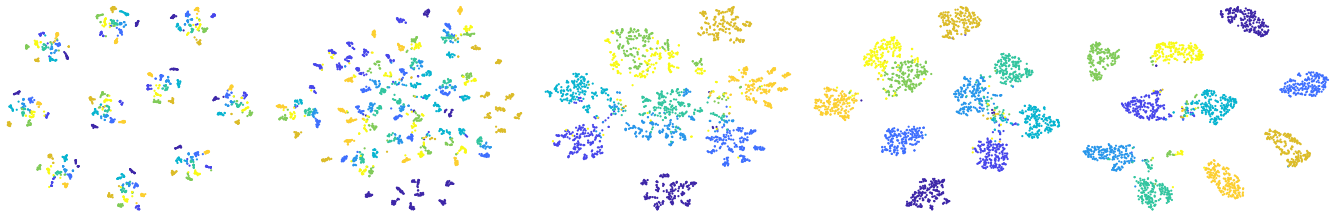


Fig. 3: The running time comparison of different algorithms on seven benchmark datasets.

structure becomes clearer, which visually demonstrates the feasibility and validity of the proposed method.

Fig. **??** shows the increasing performance of COMVSC within iterations on yaleA in terms of ACC NMI and F-score, which also verifies that the labels matrix derived from the consensus representation further guides the update of the consensus representation. The better consensus representation leads to better representations. They facilitate each other at each iteration to generate promising performance.



(a) 1 iter (b) 5 iter (c) 25 iter (d) 50 iter (e) 100 iter

Fig. 4: The cluster structure evolution illustration on Handwritten. In each figure, the low dimensional clustering reflected by the learned cluster labels matrix Y is illustrated by the t-SNE algorithm. (a)-(d) indicate the t-SNE results of the corresponding iterations.

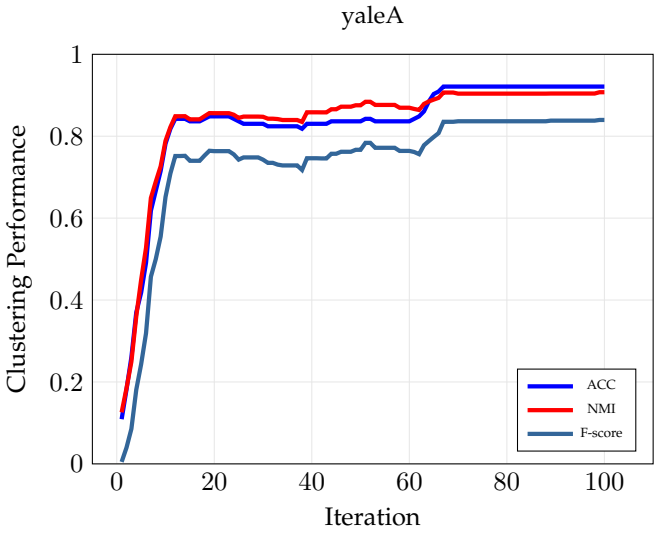


Fig. 5: The performance variation curve of the proposed algorithms on the yaleA.

TABLE 3: Comparative results for two-step and one-step strategies.

Dataset	Metric	w/o rotation	COMVSC
BBCSport	ACC	0.5603	**0.6983**
	NMI	0.2142	**0.5346**
	F-score	0.3967	**0.5322**
yaleA	ACC	0.8000	**0.9273**
	NMI	0.6504	**0.9252**
	F-score	0.6117	**0.8551**
Wikipedia	ACC	0.5498	**0.6061**
	NMI	0.3928	**0.5361**
	F-score	0.4726	**0.5063**
Handwritten	ACC	0.9200	**0.9450**
	NMI	0.8310	**0.8925**
	F-score	0.8394	**0.8934**

6.5 Parameter analysis

We also conduct the sensitivity analysis of the hyper-parameters. The proposed COMVSC method has two hyper-parameters $\{\lambda, \gamma\}$. λ is tuned within the set of $\{2^3, 2^5, \cdots, 2^{13}\}$, while γ is tuned in the range of $\{1.3, 1.5, \cdots, 2.7\}$ with incremental step 0.2.

We visualize the effect of the hyper-parameters λ and γ on purity for different datasets in Fig. 6. When $\gamma = 1$, the strategy degenerates into a hard partition pattern, which is often not applicable in practice. It can be observed that our approach maintains relatively stable clustering performance over a wide range under most datasets. In particular, the datasets BBCSport, Webkb, and yaleA perform better when λ takes the median value of its content, the datasets MSRC-v1 and Wikipedia perform better when λ takes the smaller value of its range. Handwritten maintains good performance under most parameters.

6.6 Ablation Study

To further illustrate the effectiveness of the one-step strategy, we conduct an ablation study. We remove the last spectral rotation term and feed the consensus $\mathbf{F}^*$ into k-means to get the clustering results as contrast experiment (w/o rotation). Without this term, the algorithm degrades to a two-step strategy, which requires additional discretization to obtain the final results. Compared with the two-step strategy, our algorithm outperforms it on all datasets in terms of all metrics. Part of the results are shown in Table 3.

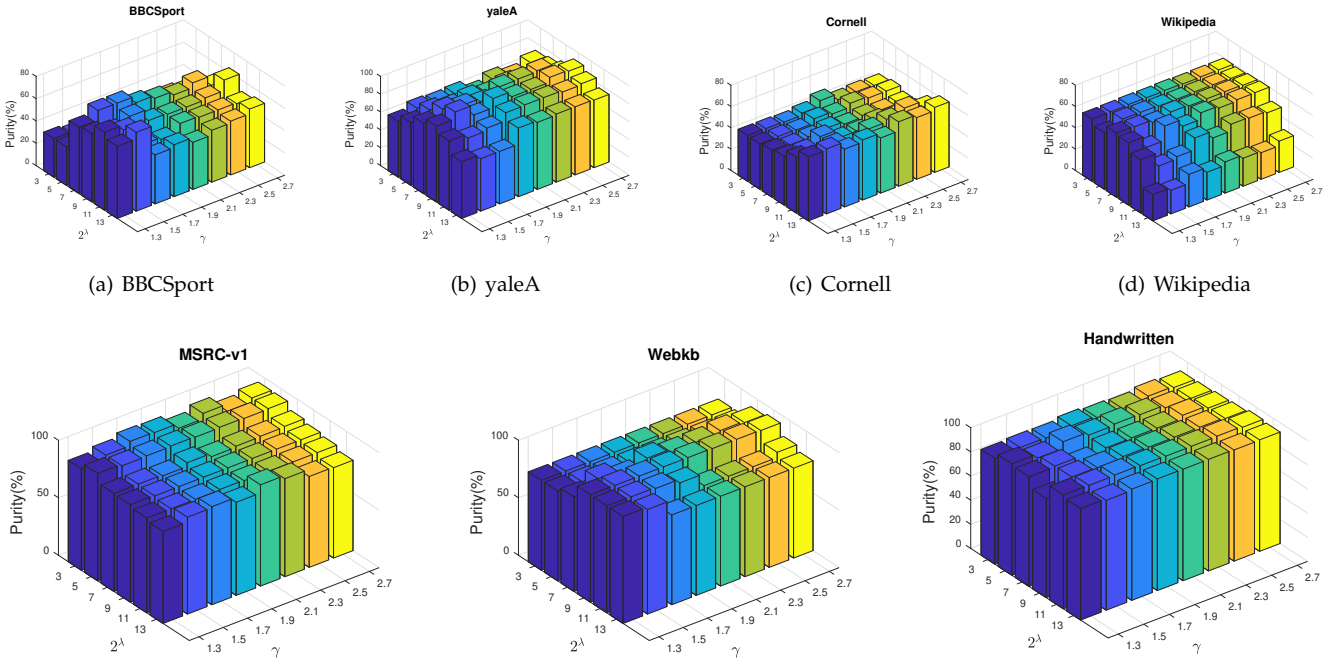
The notable results demonstrate the effectiveness and importance of the one-step strategy. It integrates all in a framework, enabling end-to-end multi-view clustering for better clustering through internal mutual negotiation and facilitation.

7 Conclusion

In this paper, we propose a novel consensus one-step multi-view subspace clustering method. Distinct to existing subspace methods, we fuse multiple subspace information in partition level. Furthermore, similarity learning, partition fusion, and the clustering processes are combined into a unified framework, which can be guaranteed to achieve an optimal solution. And the three steps can be negotiated with each other to best serve clustering, leading to improved performance. Experiment results on several multi-view benchmark datasets demonstrate the effectiveness and superiority of our proposed method. In the future, we will consider large-scale multi-view subspace clustering and adaptive fusion methods.

Acknowledgment

This work was supported by the National Natural Science Foundation of China (project no. 61773392 and 61672528).



(a) BBCSport (b) yaleA (c) Cornell (d) Wikipedia

(e) MSRC-v1 (f) Webkb (g) Handwritten

Fig. 6: Evaluating parameter sensitivity using purity.

REFERENCES

[1] R. Li, C. Zhang, Q. Hu, P. Zhu, and Z. Wang, "Flexible multi-view representation learning for subspace clustering," *IJCAI*, vol. 2019-Augus, pp. 2916–2922, 2019.

[2] C. Tang, X. Zhu, X. Liu, and L. Wang, "Cross-view local structure preserved diversity and consensus learning for multi-view unsupervised feature selection," in *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 33, 2019, pp. 5101–5108.

[3] G. Chao, S. Sun, and J. Bi, "A survey on multi-view clustering," *arXiv preprint arXiv:1712.06246*, 2017.

[4] C. Wang, J. Lai, and S. Y. Philip, "Multi-view clustering based on belief propagation," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 4, pp. 1007–1021, 2015.

[5] Y. Ren, X. Yan, Z. Hu, and Z. Xu, "Self-paced multi-task multi-view capped-norm clustering," in *International Conference on Neural Information Processing*. Springer, 2018, pp. 205–217.

[6] Z. Zhang, L. Liu, F. Shen, H. T. Shen, and L. Shao, "Binary multi-view clustering," *IEEE transactions on pattern analysis and machine intelligence*, vol. 41, no. 7, pp. 1774–1782, 2018.

[7] C. Tang, J. Chen, X. Liu, M. Li, P. Wang, M. Wang, and P. Lu, "Consensus learning guided multi-view unsupervised feature selection," *Knowledge-Based Systems*, vol. 160, pp. 49–60, 2018.

[8] D. Huang, C. Wang, J. Wu, J. Lai, and C. Kwoh, "Ultra-scalable spectral clustering and ensemble clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 32, no. 6, pp. 1212–1226, 2019.

[9] H. H. Xiao Cai, Feiping Nie, "Multi-view k-means clustering on big data," *IJCAI*, vol. 31, no. 5, pp. 724–730, 2013.

[10] C. Tang, X. Liu, X. Zhu, J. Xiong, M. Li, J. Xia, X. Wang, and L. Wang, "Feature selective projection with low-rank embedding and dual laplacian regularization," *IEEE Transactions on Knowledge and Data Engineering*, pp. 1–1, 2019.

[11] Y. Wang, X. Lin, L. Wu, W. Zhang, Q. Zhang, and X. Huang, "Robust subspace clustering for multi-view data by exploiting correlation consensus," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 3939–3949, 2015.

[12] S. Wang, X. Liu, E. Zhu, C. Tang, J. Liu, J. Hu, J. Xia, and J. Yin, "Multi-view clustering via late fusion alignment maximization," *IJCAI-19*, no. June, pp. 3778–3784, 2019.

[13] Z. Kang, Z. Guo, S. Huang, S. Wang, W. Chen, Y. Su, and Z. Xu, "Multiple partitions aligned clustering," *IJCAI*, pp. 2701–2707, 2019.

[14] C. Tang, X. Liu, X. Zhu, E. Zhu, Z. Luo, L. Wang, and W. Gao, "CGD: Multi-view clustering via cross-view graph diffusion." in *AAAI Conference on Artificial Intelligence*, 2020, pp. 5924–5931.

[15] B.-Y. Liu, L. Huang, C.-D. Wang, S. Fan, and S. Y. Philip, "Adaptively weighted multiview proximity learning for clustering," *IEEE Transactions on Cybernetics*, 2019.

[16] Y. Yang and H. Wang, "Multi-view clustering: A survey," *Big Data Mining and Analytics*, vol. 1, no. 2, pp. 83–107, 2018.

[17] H. D. Iii, "A co-training approach for multi-view spectral clustering abhishek kumar."

[18] A. Kumar, P. Rai, and H. Daume, "Co-regularized multi-view spectral clustering," in *Advances in Neural Information Processing Systems*, 2011, pp. 1413–1421.

[19] X. Liu, X. Zhu, M. Li, L. Wang, E. Zhu, T. Liu, M. Kloft, D. Shen, J. Yin, and W. Gao, "Multiple kernel k-means with incomplete kernels," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1–1, 2019.

[20] X. Liu, D. Yong, J. Yin, W. Lei, and E. Zhu, "Multiple kernel k-means clustering with matrix-induced regularization," in *Thirtieth Aaai Conference on Artificial Intelligence*, 2016.

[21] J.-T. Tsai, Y.-Y. Lin, and H.-Y. M. Liao, "Per-cluster ensemble kernel learning for multi-modal image clustering with group-dependent feature selection," *IEEE Transactions on Multimedia*, vol. 16, no. 8, pp. 2229–2241, 2014.

[22] Y. Chen, X. Xiao, and Y. Zhou, "Jointly learning kernel representation tensor and affinity matrix for multi-view clustering," *IEEE Transactions on Multimedia*, 2019.

[23] S. Zhou, X. Liu, M. Li, E. Zhu, L. Liu, C. Zhang, and J. Yin, "Multiple kernel clustering with neighbor-kernel subspace segmentation," *IEEE transactions on neural networks and learning systems*, vol. 31, no. 4, pp. 1351–1362, 2019.

[24] F. Nie, J. Li, and X. Li, "Parameter-free auto-weighted multiple graph learning: A framework for multiview clustering and semi-supervised classification." in *International Joint Conference on Artificial Intelligence*, 2016.

[25] Z. K, N. F, W. J, and Y. Y, "Multiview consensus graph clustering," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society*, vol. 28, no. 3, p. 1261, 2019.

[26] C. Wei, Z. Xiang, Z. Guo, Y. Wu, and W. Wei, "Flexible and robust co-regularized multi-domain graph clustering," in *Acm Sigkdd International Conference on Knowledge Discovery and Data Mining*, 2013.

[27] C. Tang, X. Zhu, X. Liu, M. Li, P. Wang, C. Zhang, and L. Wang, "Learning a joint affinity graph for multiview subspace clustering," *IEEE Transactions on Multimedia*, vol. 21, no. 7, pp. 1724–1736, 2018.

[28] Y. Guo, "Convex subspace representation learning from multi-view data," *Proceedings of the 27th AAAI Conference on Artificial Intelligence, AAAI 2013*, pp. 387–393, 2013.

[29] M. Brbić and I. Kopriva, "Multi-view low-rank sparse subspace clustering," *Pattern Recognition*, vol. 73, pp. 247–258, 2018.

[30] R. Xia, Y. Pan, L. Du, and J. Yin, "Robust multi-view spectral clustering via low-rank and sparse decomposition," *Proceedings of the National Conference on Artificial Intelligence*, vol. 3, pp. 2149–2155, 2014.

[31] X. Cao, C. Zhang, H. Fu, and S. Liu, "Diversity-induced multi-view subspace clustering," *CVPR*, 2015.

[32] X. Zhu, S. Zhang, R. Hu, W. He, C. Lei, and P. Zhu, "One-step multi-view spectral clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 31, no. 10, pp. 2022–2034, 2018.

[33] E. Elhamifar and R. Vidal, "Sparse subspace clustering: Algorithm, theory, and applications," *IEEE transactions on pattern analysis and machine intelligence*, vol. 35, no. 11, pp. 2765–2781, 2013.

[34] H. Gao, F. Nie, X. Li, and H. Huang, "Multi-view subspace clustering," *ICCV*, vol. ICCV, pp. 4238–4246, 2015.

[35] X. Zhang, X. Zhang, H. Liu, and X. Liu, "Multi-task multi-view clustering," *IEEE Transactions on Knowledge and Data Engineering*, vol. 28, no. 12, pp. 3324–3338, 2016.

[36] G. Tuysuzoglu, D. Birant, and A. Pala, "Majority voting based multi-task clustering of air quality monitoring network in turkey," *Applied Sciences*, vol. 9, no. 8, p. 1610, Apr 2019. [Online]. Available: http://dx.doi.org/10.3390/app9081610

[37] D. F. Wulsin, S. T. Jensen, B. Litt *et al.*, "Nonparametric multi-level clustering of human epilepsy seizures," *The Annals of Applied Statistics*, vol. 10, no. 2, pp. 667–689, 2016.

[38] B. Mohar, Y. Alavi, G. Chartrand, O. Oellermann, and A. Schwenk, "The Laplacian spectrum of graphs," *Graph Theory, Combinatorics and Applications*, vol. 2, p. 5364, 1991.

[39] Fan and K., "On a theorem of weyl concerning eigenvalues of linear transformations: II," *Proceedings of The National Academy Of Sciences Of The United States Of America*, vol. 36, no. 1, pp. 31–35, 1950.

[40] Y. Pang, J. Xie, F. Nie, and X. Li, "Spectral clustering by joint spectral embedding and spectral rotation," *IEEE transactions on cybernetics*, 2018.

[41] Z. Ren and Q. Sun, "Simultaneous Global and Local Graph Structure Preserving for Multiple Kernel Clustering," pp. 1–13.

[42] F. Nie, X. Wang, and H. Huang, "Clustering and projected clustering with adaptive neighbors," in *Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 977–986.

[43] ——, "Clustering and projected clustering with adaptive neighbors," *Proceedings of the ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, no. December, pp. 977–986, 2014.

[44] D. Greene and P. Cunningham, "Practical solutions to the problem of diagonal dominance in kernel document clustering," in *Proceedings of the 23rd International Conference on Machine Learning*, pp. 377–384.

[45] D. Cai, X. He, Y. Hu, J. Han, and T. Huang, "Learning a spatially smooth subspace for face recognition," in *2007 IEEE Conference on Computer Vision and Pattern Recognition*. IEEE, pp. 1–7.

[46] G. Bisson and C. Grimal, "An architecture to efficiently learn co-similarities from multi-view datasets," in *International Conference on Neural Information Processing*. Springer, pp. 184–193.

[47] N. Xu, Y. Guo, J. Wang, X. Luo, and X. Kong, "Multi-view clustering via simultaneously learning shared subspace and affinity matrix," *International Journal of Advanced Robotic Systems*, vol. 14, no. 6, pp. 1–8, 2017.

[48] J. C. Pereira, E. Coviello, G. Doyle, N. Rasiwasia, G. R. G. Lanckriet, R. Levy, and N. Vasconcelos, "On the role of correlation and abstraction in cross-modal multimedia retrieval," *IEEE transactions on pattern analysis and machine intelligence*, vol. 36, no. 3, pp. 521–535, 2014.

[49] V. Sindhwani, P. Niyogi, and M. Belkin, "Beyond the point cloud: From transductive to semi-supervised learning," in *Proceedings of the 22nd International Conference on Machine Learning*, pp. 824–831.

[50] A. Oliva and A. Torralba, "Modeling the shape of the scene: A holistic representation of the spatial envelope," *International journal of computer vision*, vol. 42, no. 3, pp. 145–175, 2001.

[51] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on pattern analysis and machine intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[52] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International journal of computer vision*, vol. 60, no. 2, pp. 91–110, 2004.

[53] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 886–893.

[54] H. Yu, M. Li, H.-J. Zhang, and J. Feng, "Color texture moments for content-based image retrieval," in *Proceedings. International Conference on Image Processing*, vol. 3. IEEE, 2002, pp. 929–932.

[55] A. Kumar, P. Rai, and H. D. Iii, "Co-regularized multi-view spectral clustering."

[56] C. Zhang, Q. Hu, H. Fu, P. Zhu, and X. Cao, "Latent multi-view subspace clustering," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017.

[57] H. Wang, Y. Yang, and B. Liu, "GMC: Graph-based multi-view clustering," *IEEE Transactions on Knowledge and Data Engineering*, 2019.

[58] Z. Kang, W. Zhou, Z. Zhao, J. Shao, M. Han, and Z. Xu, "Large-scale multi-view subspace clustering in linear time," in *The Thirty-Fourth AAAI Conference on Artificial Intelligence, AAAI 2020, The Thirty-Second Innovative Applications of Artificial Intelligence Conference, IAAI 2020, The Tenth AAAI Symposium on Educational Advances in Artificial Intelligence, EAAI 2020, New York, NY, USA, February 7-12, 2020*. AAAI Press, 2020, pp. 4412–4419. [Online]. Available: https://aaai.org/ojs/index.php/AAAI/article/view/5867

[59] Z. Kang, X. Zhao, C. Peng, H. Zhu, J. T. Zhou, X. Peng, W. Chen, and Z. Xu, "Partition level multiview subspace clustering," *Neural Networks*, vol. 122, pp. 279–288, 2020.

[60] Z. Wen and W. Yin, "A feasible method for optimization with orthogonality constraints," vol. 142, no. 1-2, pp. 397–434.

Pei Zhang is a graduate student in National University of Defense Technology (NUDT), China. Her current research interests include multi-view learning, incomplete multi-view clustering and deep clustering.

Xinwang Liu received his PhD degree from National University of Defense Technology (NUDT), China. He is now Professor at School of Computer, NUDT. His current research interests include kernel learning and unsupervised feature learning. Dr. Liu has published 60+ peer-reviewed papers, including those in highly regarded journals and conferences such as IEEE T-PAMI, IEEE T-KDE, IEEE T-IP, IEEE T-NNLS, IEEE T-MM, IEEE T-IFS, NeurIPS, CVPR, ICCV, AAAI, IJCAI, etc. More information can be found at https://xinwangliu.github.io/.

Sihang Zhou received his PhD degree from National University of Defense Technology (NUDT), China. He is now lecturer at College of Intelligence Science and Technology, NUDT. His current research interests include machine learning and medical image analysis. Dr. Zhou has published 20+ peer-reviewed papers, including IEEE T-IP, IEEE T-NNLS, IEEE T-MI, Information Fusion, AAAI, MICCAI, etc.

Wentao Zhao received his Ph.D. degree from National University of Defense Technology (NUDT) in 2009. He is now a Professor at NUDT. His research interests include network performance optimization, information processing and machine learning. Since 2011, Dr. Zhao has been serving as a member of Council Committee of Posgraduate Entrance Examination of Computer Science and Technology, NUDT. He has edited one book entitled ”Database Principle and Technology” and several technical papers such as Communications of the CCF, AAAI, IJCAI, FAW, etc.

En Zhu received his PhD degree from National University of Defense Technology (NUDT), China. He is now Professor at School of Computer Science, NUDT, China. His main research interests are pattern recognition, image processing, machine vision and machine learning. Dr. Zhu has published 60+ peer-reviewed papers, including IEEE T-CSVT, IEEE T-NNLS, PR, AAAI, IJCAI, etc. He was awarded China National Excellence Doctoral Dissertation.

Zhiping Cai received the B.Eng., M.A.Sc., and Ph.D. degrees in computer science and technology from the National University of Defense Technology (NUDT), China, in 1996, 2002, and 2005, respectively. He is a full professor at School of Computer Science, NUDT. His current research interests include artificial intelligence, network security and big data. He is a senior member of the CCF and a member of the IEEE.