Consensus One-step Multi-view Subspace Clustering

Journal:	*Transactions on Knowledge and Data Engineering*
Manuscript ID	TKDE-2020-05-0525
Manuscript Type:	Regular
Keywords:	Multi-view Clustering, Subspace Clustering

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 been attracted increasing attention in multimedia, machine learning and data mining communities. As one kind of the most important multi-view clustering algorithms, multi-view subspace clustering (MVSC) becomes more and more popular due to its strong ability in revealing the intrinsic low dimensional clustering structure hidden across views. Though demonstrating 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 the multi-view information being insufficiently utilized, leading to unsatisfying clustering performance. In this paper, we propose 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 be negotiated with each other to best serve clustering, leading to improved performance. An iterative algorithm is proposed to solve the resulting optimization problem. Extensive experimental 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, individual feature is insufficient for depicting data points while different features often contain complementary information which could be of benefit to exploring the underlying structure of data [1]. Therefore, multi-view clustering (MVC) has thus attracted attention in data mining and machine learning communities [2], [3], [4]. MVC aims at categorizing similar data points into the same cluster and dissimilar points into different clusters by combining the available multiple feature information, and searching for consistent clustering result across different views [5], [6], [7], [8], [9], [10], [11].

Numerous multi-view clustering methods have been proposed in recent years. According to [1], multi-view clustering can be summarized into four categories in terms of the mechanisms they base. The first category is co-training. This category of algorithms reach a consensus by iteratively exploiting prior knowledge or information learned form other views. There are two representative method: co-training multi-view spectral clustering [12] and co-regularized multi-view spectral clustering [13]. The second category refers to multi-kernel clustering [9], [14], [15], [16], [17], aiming at combining predefined kernels linearly or non-linearly to boost the clustering performance. The third prospective is multi-view graph clustering. This kind of method finds a unified graph and then spectral clustering or other graph-cut algorithms are performed on it [18], [19], [20], [21]. 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. The key idea of these methods is to learn a unified and discriminative indicator matrix, which clearly reveals the ideal cluster structure from multiple views. Most existing methods for multi-view subspace clustering integrate multi-view information in similarity or representation level by merging multiple graphs or representation matrices into a shared one. For example, [22] learns a shared sparse subspace representation by performing matrix factorization. Similarly, the centroid-based multi-view low-rank sparse subspace clustering method [23] induces low-rank and sparsity constraints on the shared affinity matrix across different views. [1] constructs 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, [24] and [25] induce Hilbert-Schmidt Independence Criterion(HSIC) and Markov chain to learn distinct subspace representations and then add them together directly or adaptively to get a unified representation.

Although these aforementioned subspace methods have achieved great 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, learning the

- *P. Zhang, X. Liu, W. Zhao, 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..*
- *X. Liu is the corresponding author.*

common affinity matrix or graph from original data may affect the clustering structure because the original data often consists of noise and redundancy [26]. 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 be not suitable for the clustering task. They are not able to 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 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 self-representation manner in each view. Based on the assumption that 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 resultant 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 resultant optimization problem. Extensive experiments on several multi-view benchmark datasets demonstrate the effectiveness of our method comparing to other state-of-the-art approaches.

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 $\mathbf{A}_{i,j}$ represent its i-th row and the ij-th element. The Frobenius norm of matrix $\mathbf{A}$ is denoted as $\|A\|_F$. 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 $\mathbf{A}^{\mathrm{T}}$, $Tr(\mathbf{A})$, 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 [27] 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 $\mathbf{AZ}$, the general formulation of subspace clustering can be expressed as:

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

where $\mathbf{L}(\cdot)$ and $\mathbf{\Omega}(\cdot)$ denote the reconstruction loss function and regularization term respectively. $\lambda > 0$ is a balance parameter. $\mathbf{Z}$ is also called the self-representaion 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 [28].

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.

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

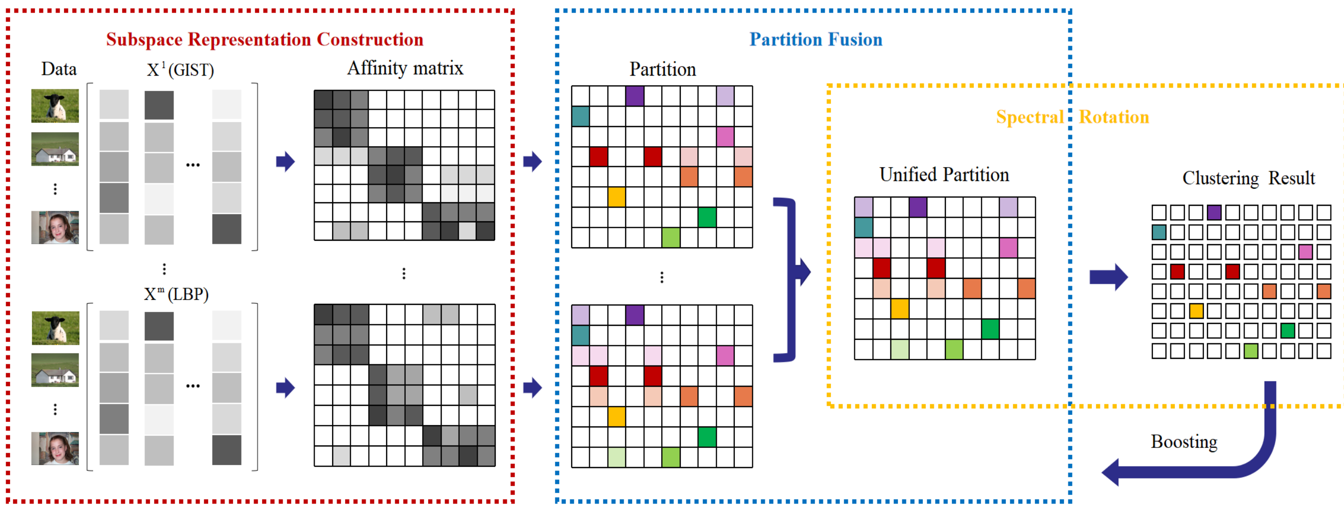


Fig. 1. Framework of proposed COMVSC. With the multi-view data input, COMVSC learns the partition information from corresponding affinity matrix. Then a partion-level fusion method is applied to integrate the complementary information accross the multiple views. The final clustering results can directly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

$\mathbf{Z}^v \in \mathbb{R}^{n\times n}$ 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 different view are similar theoretically, the magnitude of element values in $\mathbf{Z}^v$ are remarkably different [28]. 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) [25] explores the complementary information among multi-view features using Hilbert-Schmidt Independence Criterion(HSIC). Then diverse subspace representations are simply added together as the input of spectral clustering to generate the final result. Another strategy such as Multi-view Subspace Clustering (MVSC) [28] strengthens the consistency between different views by preforming 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 datas have noise and outliers, which resulting in a poor similarity matrix. Nevertheless, information in partition level reflect the intrinsic clustering structure. Therefore, considering to fuse multi-view information in partition level is a natural and novel idea. Furthermore, most of 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 [26]. 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 representaions we learned are designed for the clustering goal and more informative partition-level information are fused to obtain the final clustering result.

3 Proposed Approach

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

3.1 Representation Construction

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

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

$\mathbf{Z}^v$ 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 components in similarity matrix is equal to the multiplicity of eigenvalue 0 of corresponding Laplacian matrix [29].*

It 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 $n-k$. Naturely, we add a rank constraint in Eq. (4) to achieve this condition. The optimization problem becomes:

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

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 $D_{i,i}^v = \sum_{j=1}^n Z_{ij}^v$.

However, directly applying rank constraint $rank(\mathbf{L}^v) = n-k$ to Eq. (4) will make the optimization problem hard to tackle. According to Ky Fan's Theorem [30], we can obtain $\sum_{i=1}^k \sigma_i(\mathbf{L}^v) = arg\ \min_{\mathbf{F}^v \in \mathbb{R}^{n\times k}, \mathbf{F}^v\mathbf{F}^v = \mathbf{I}_k}\ Tr((\mathbf{F}^v)^{\mathrm{T}}\mathbf{L}^v\mathbf{F}^v)$, where $\sigma_i(\mathbf{L}^v)$ is the i-th smallest eigenvalues of $\mathbf{L}^v$ and $\mathbf{F}^v \in \mathrm{R}^{n\times k}$ is the clustering indicator matrix. It is obvious that the rank of $\mathbf{L}^v$ equals to $n-k$ when $\sum_{i=1}^k \sigma_i(\mathbf{L}^v) = 0$ Therefore, the problem can be reformulated as the following equivalent form:

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

The problem in Eq. (5) can be solved easily by transforming the rank constraint into trace.

3.2 Partition Fusion

As shown in Eq. (6), each view can get its individual clustering indicator matrix $\mathbf{F}^v$. Multi-view clustering holds the assumption that various clustering structures in different views should be analogous to each other [28]. 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 $\mathbf{F}^*$, which is so called partition fusion. Mathematically, this term can be formulated as:

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

3.3 Spectral Rotation

After obtaining the consensus representation $\mathbf{F}^*$, we induce a rotation matrix $\mathbf{R} \in \mathbb{R}^{k\times k}$ to incorporate representation learning, partition fusion and clustering process. This term can be written as:

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

where γ is considered as fuzzy coefficient and the vector $\mathbf{t}_c(c = 1\ldots k)$ is a $1\times k$ vector which c-th element equals to 1 and others are 0, identifying the k classes. $\mathbf{F}_{i,:}^*$ is the i-th row of $\mathbf{F}^*$, indicating the representation corresponding to the i-th sample. Inspired by [31], $Y_{i,c}$ 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 representaion $\mathbf{F}_{i,:}^*$ 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.

3.4 Unified Framework

Combining Eq. (6)(7)(8), we can fulfill our COMSC framework as follows:

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

By 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 best 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

$$\min_{\mathbf{Z}^v} \|\mathbf{X}^v - \mathbf{X}^v\mathbf{Z}^v\|_F^2 + \lambda\|\mathbf{Z}^v\|_F^2 + Tr((\mathbf{F}^v)^{\mathrm{T}}\mathbf{L}^v\mathbf{F}^v) \tag{10}$$

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

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

Note that $Tr((\mathbf{F}^{\mathrm{T}}\mathbf{L}\mathbf{F}) = \sum_{i,j=1}^n \frac{1}{2}Z_{i,j}\|\mathbf{f}_{i,:} - \mathbf{f}_{j,:}\|^2$. We denote $Q_{i,j} = \|\mathbf{f}_{i,:} - \mathbf{f}_{j,:}\|$, therefore $\mathbf{Q}_{i,:} \in \mathbb{R}^{1\times n}$. The vector form of Eq. (11) can be rewritten as:

$$\min_{\mathbf{Z}_{:,i}} \|\mathbf{X}_{:,i} - \mathbf{X}\mathbf{Z}_{:,i}\|^2 + \lambda\mathbf{Z}_{:,i}^{\mathrm{T}}\mathbf{Z}_{:,i} + \frac{1}{2}\mathbf{Q}_{i,:}\mathbf{Z}_{:,i} \tag{12}$$

By setting the derivative of Eq. (12) with respect to $\mathbf{Z}_{:,i}$ to zero, we can get the following closed-form solution:

$$\mathbf{Z}_{:,i} = (\mathbf{X}^{\mathrm{T}}\mathbf{X} + \lambda\mathbf{I})^{-1}(\mathbf{X}^{\mathrm{T}}\mathbf{X}_{:,i} - \frac{1}{4}\mathbf{Q}_{i,:}^{\mathrm{T}}) \tag{13}$$

Note that the inverse only needs to be caculated once as long as λ is fixed.

4.2 Update consensus cluster indicator matrix $\mathbf{F}^*$

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

$$\begin{aligned} \min_{\mathbf{F}^*} \sum_{v=1}^{m} \|\mathbf{F}^v - \mathbf{F}^*\|_F^2 + \sum_{i=1}^{n}\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\|\mathbf{t}_c - \mathbf{F}^*_{i,:}\mathbf{R}\|_2^2 \\ s.t.\ (\mathbf{F}^*)^{\mathrm{T}}\mathbf{F}^* = \mathbf{I}_k \end{aligned} \tag{14}$$

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

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

$$\max\sum_{i=1}^{n}\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\mathbf{R}^{\mathrm{T}}_{c,:}(\mathbf{F}^*)^{\mathrm{T}}_{:,i} \tag{15}$$

By expanding Eq. (15) element-wise, we can get: $\sum_{i=1}^{n}\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\mathbf{R}^{\mathrm{T}}_{c,:}(\mathbf{F}^*)^{\mathrm{T}}_{:,i} = Tr(\mathbf{R}^{\mathrm{T}}(\mathbf{F}^*)^{\mathrm{T}}\mathbf{G})$, where $\mathbf{G}_{i,:} = \sum_{c=1}^{k}(Y_{i,c}^{\gamma})\mathbf{t}_c$.

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 $\max\ Tr(\mathbf{R}^{\mathrm{T}}(\mathbf{F}^*)^{\mathrm{T}}\mathbf{G})$, 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* $\mathbf{B} = \mathbf{U\Sigma V}^{\mathrm{T}}$. *Accordingly, the constrained problem*

$$\max_{\mathbf{A}}\ Tr(\mathbf{A}^{\mathrm{T}}\mathbf{B})\ s.t.\ \ \mathbf{A}^{\mathrm{T}}\mathbf{A} = \mathbf{I} \tag{16}$$

has closed form solution:

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

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

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

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

The solution of maximize Eq. (16) can be obtained when $\mathbf{Q\Sigma} = \mathbf{I\Sigma}$, 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. (14) can be transformed into the follow matrix form:

$$\begin{aligned} \min_{\mathbf{F}^*}\sum_{v=1}^{m} -Tr((\mathbf{F}^v)^{\mathrm{T}}\mathbf{F}^*) - Tr(\mathbf{R}^{\mathrm{T}}(\mathbf{F}^*)^{\mathrm{T}}\mathbf{G}) \\ s.t.\ \ (\mathbf{F}^*)^{\mathrm{T}}\mathbf{F}^* = \mathbf{I}_k. \end{aligned} \tag{17}$$

Eq. (17) can be transformed into Eq. (18) by simple transformation.

$$\max_{\mathbf{F}^*}\ Tr((\mathbf{F}^*)^{\mathrm{T}}\mathbf{N}) \quad s.t.\ \ (\mathbf{F}^*)^{\mathrm{T}}\mathbf{F}^* = \mathbf{I}_k, \tag{18}$$

where $\mathbf{N} = \sum_{v=1}^{m}\mathbf{F}^v + \mathbf{G}\mathbf{R}^{\mathrm{T}}$. 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 $\mathbf{N}$ respectivelly.

4.3 Update cluster indicator matrix $\mathbf{F}^v$

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:

$$\begin{aligned} \min_{\mathbf{F}^v}\ Tr((\mathbf{F}^v)^{\mathrm{T}}\mathbf{L}^v\mathbf{F}^v) - 2Tr((\mathbf{F}^v)^{\mathrm{T}}\mathbf{F}^*) \\ s.t.\ \ (\mathbf{F}^v)^{\mathrm{T}}\mathbf{F}^v = \mathbf{I}_k \end{aligned} \tag{19}$$

Equation Eq. (19) can be relaxed into the following form, where λ_{max} is the largest eigenvalue of $\mathbf{L}^v$.

$$\begin{aligned} \max_{\mathbf{F}^v}\ Tr((\mathbf{F}^v)^{\mathrm{T}}(\lambda_{max}\mathbf{I} - \mathbf{L}^v)\mathbf{F}^v) + 2(Tr(\mathbf{F}^v)^{\mathrm{T}}\mathbf{F}^*) \\ s.t.\ \ (\mathbf{F}^v)^{\mathrm{T}}\mathbf{F}^v = \mathbf{I}_k \end{aligned} \tag{20}$$

Proposition 3. *If* $\mathbf{A}$,$\mathbf{B}$ *are positive semidefinite matrices,* $f(\mathbf{W}) = Tr(\mathbf{W}^{\mathrm{T}}\mathbf{AWB}) + Tr(\mathbf{W}^{\mathrm{T}}\mathbf{C})$ *is a convex function. Problem can be solved by optimizing* $\max_{\mathbf{W}^{\mathrm{T}}\mathbf{W}=\mathbf{I}} Tr(\mathbf{W}^{\mathrm{T}}\mathbf{M})$ *iteratively, where* $\mathbf{M} = f'(\mathbf{W}) = 2\mathbf{AWB} + \mathbf{C}$. *[32]*

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

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

```
while not converged do
    M = 2(λ_max I − L^v)F^v + 2F*.
    Solve max_{(F^v)^T F^v = I} Tr((F^v)^T M) to update F^v.
        Perform SVD on M, M = U_2 Σ V_2^T
        F^v = U_2 V_2^T
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:

$$\begin{aligned} \min_{\mathbf{R}}\sum_{i=1}^{n}\sum_{c=1}^{k}(Y_{i,c})^{\gamma}\|\mathbf{t}_c - \mathbf{F}^*_{i,:}\mathbf{R}\|_2^2 \\ s.t.\ \ \mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}_k. \end{aligned} \tag{21}$$

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

$$\max_{\mathbf{R}}\ Tr(\mathbf{R}^{\mathrm{T}}(\mathbf{F}^*)^{\mathrm{T}}\mathbf{G}) \quad s.t.\ \mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}_k, \tag{22}$$

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

$$\max_{\mathbf{R}}\ Tr(\mathbf{R}^{\mathrm{T}}\mathbf{H}) \quad s.t.\ \ \mathbf{R}^{\mathrm{T}}\mathbf{R} = \mathbf{I}_k \tag{23}$$

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 $\mathbf{H}$ respectivelly.

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 row of the matrix $\mathbf{Y}$, can be transformed into solving the following problem:

$$\begin{aligned} &\min \sum_{c=1}^{k} (Y_{i,c})^{\gamma} \|\mathbf{t}_c - \mathbf{F}^*_{i,:}\mathbf{R}\|_2^2 \\ &s.t.\ \ Y_{i,c} \geq 0, \mathbf{Y}_{i,:}\mathbf{1}_k = 1 \end{aligned} \tag{24}$$

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. (24) 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.

$$Y_{i,c} = \frac{(P_{i,c})^{\frac{1}{1-\gamma}}}{\sum_{c=1}^{k} (P_{i,c})^{\frac{1}{1-\gamma}}} \tag{25}$$

The entire optimization is summaried 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 lowwer-bounded. As a result, the proposed algorithm can be verified to be convergent.

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. The complexity of first step (similarity learning) is $\mathcal{O}(mn^3 + (m+1)nk^3)$. Then the second partition fusion process costs $\mathcal{O}(k^3)$ while the last one is $\mathcal{O}(nk^2)$. Overall, the complexity of our Algorithm 2 is $\mathcal{O}(T(mn^3 + mnk^3 + k^3))$, where T is the number of 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 this joint manner ensures that the similarity learning and the clustering task are integrated together.

6 Experiments

In this section, we extensively evaluate the clustering property of the proposed method on four 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.

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 $c\times c$ dimensional identity matrix. Initialize label matrix $\mathbf{Y}$ with only one 1 in each row.

```
1: while not converged do
2:     Update Z^v by Eq. (13).
3:     Update F* by solving Eq. (18).
4:     Update F^v by solving Eq. (19).
5:     Update R by solving Eq. (22).
6:     Update Y by Eq. (25).
7: end while
8: return clustering labels Y. In each row, the column
   number of the maximize element is exactly the cluster
   the data point belonging.
```

6.1 Datasets Description

Four public multi-view benchmark datasets including MSRC-v1 [33], Caltech7, Caltech20 and Wikipedia Articles [34] are used in our experiments. Specifically, the key information of them is summarized in Table 1.

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, 512-dimensional GIST, 256-dimensional LBP, 210-dimensional SIFT, 100 dimensional HOG, 48-dimensional CMT features from each image.

Caltech7 is a 101 categories image dataset. We selected the widely used 1474 images in 7 classes, i.e., Face, Motorbikes, Dolla-Bill, Garfield, Snoopy, Stop-Sign and Windsor-chair. Following previous work [35], we extract 48 dimension Gabor feature, 40 dimension wavelet moments (WM), 254 dimension CENTRIST feature, 1984 dimension HOG feature, 512 dimension GIST feature, and 928 dimension LBP feature from each image.

Caltech20 contains widely used 20 classes: Brain, Camera, Face, Ferry, Rhino, Pagoda, Snoopy, Wrench, Stapler, Leopards, Hedgehog, Garfield, Binocular, Motorbikes, Windsor Chair, Car-Side, Dolla-Bill, Stop-Sign, Yin-yang, and Water-Lilly. Six kinds of features are extracted as reported in Caltech7.

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 a Latent Dirchlet Allocation model are extracted from the data.

6.2 Compared Methods

We compare our proposed COMVSC with the following methods, including a baseline and 6 state-of-the-art multi-view subspace clustering algorithms.

- **FeatConcate** 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 (**Co-reg**) [36] utilizes a co-regularization term to make the partitions in different views agree with each other. Two schemes, i.e., centroid-based method and pairwise method, are proposed to accomplish this goal.
- Robust Multi-view K-Means clustering (**RMKMC**) [7] integrates multiple representations adaptively and induces structured sparsity-inducing norm to make it more robust to outliers.
- Latent Multi-view Subspace Clustering (**LMSC**) [37] conducts subspace clustering on latent representation learned from multi-view features to generate a common subspace representation.
- Multi-view Low-Rank Sparse Subspace Clustering (**MLRSSC**) [23] learns a joint subspace representation across all the views by conducting sparsity and low-rank constraint on each affinity matrix.
- multiple Partition Aligned Clustering (**mPAC**) [10] learns 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**) [38] 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.

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:

$$\text{ACC} = \frac{\sum_{i=1}^{n} \delta(p_i, map(q_i))}{n} \tag{26}$$

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.

TABLE 1
Details of the used multi-view datasets.

View	Wiki	MSRC-v1	Caltech7	Caltech20
1	SIFT(128)	CENT(1320)	HOG(1984)	HOG(1984)
2	LDA(10)	GIST(512)	LBP(928)	LBP(928)
3	-	LBP(256)	GIST(512)	GIST(512)
4	-	SIFT(210)	CENTRIST(254)	CENTRIST(254)
5	-	HOG(100)	Gabor(48)	Gabor(48)
6	-	CMT(48)	WM(40)	WM(40)
Views	2	6	6	6
Cluster	10	7	7	20
Points	693	210	1474	2386

Given two variables P and Q, NMI is defined as

$$\text{NMI}(P, Q) = \frac{I(P, Q)}{\sqrt{(H(P)H(Q))}}, \tag{27}$$

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:

$$\text{F}_1\text{-score} = 2 \times \frac{Precision \times Recall}{Precision + Recall}, \tag{28}$$

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

6.3 Experimental Results and Analysis

The three evaluation metrics (ACC, NMI, and F-score) of the compared algorithms on the four real-world datasets are displayed in Table 3-5. The best result is highlighted in red, while the second-best is reported in blue.

As shown in Table 2, our COMVSC method outperforms other reported methods at various degrees. To be specific, our method exceeds the second-best method by 2.16%, 2.26% and 2% in terms of ACC, NMI and F-score on Wikipedia Articles.

TABLE 2
The ACC, NMI and F-score of the compared algorithm on Wikipedia Articles

Methods	ACC	NMI	F-score
FeatConcate	0.5592	0.5467	0.4753
Co-reg_centroid [36]	0.5442	0.4708	0.4417
Co-reg_pairwise [36]	0.5623	0.4913	0.4593
MLRSSC [23]	0.1531	0.0083	0.1968
LMSC [37]	0.5931	0.5188	0.5014
RMKMC [7]	0.5657	0.5553	0.4794
mPAC [10]	0.5426	0.4725	0.4344
FMR [1]	0.5758	0.5176	0.4838
GMC [38]	0.4488	0.4170	0.2835
ours	0.6147	0.5779	0.5214

In most cases, FeatConcate method performs the 1st 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 accross mulitple views.

From Table 2, we can find that the subspace-based methods (LMSC, mPAC, FMR and our COMVSC) outperforms the graph-based method (GMC) and the multi-view K-means (RMKMC) methods. This is mainly because the partition-level information is more informative and less noisy than the similarity-level information. And the more informative multiple partition information is utilized to best serve the clustering task, which proves that fusing partition information is an effective approach in dealing with multi-view clustering problem.

As seen in Table 3, our method performs the best in terms of F-score and the second in terms of ACC and NMI. The reason why GMS achieves the best result on ACC is that the quality of the single view is relatively higher on MSRC-v1 than on other datasets. Moreover, our COMVSC method surpasses other reported multi-view clustering methods 1.9%, 5.68% and 6.47% respectively in terms of ACC, NMI and F-score. Compared with mPAC, our algorithm with less hyper-parameter could achieve superior performance. This is attributed to that the proposed merging multiple partitions into a consensus one could preserve the local structure of each indicator and thus facilitate the ultimate multi-view clustering.

TABLE 3
The ACC, NMI and F-score of the compared algorithm on MSRC-v1.

Methods	ACC	NMI	F-score
FeatConcate	0.4541	0.4217	0.3856
Co-reg_centroid [36]	0.7991	0.6991	0.6827
Co-reg_pairwise [36]	0.6933	0.6032	0.5866
MLRSSC [23]	0.4852	0.3642	0.4021
LMSC [37]	0.7476	0.6450	0.6384
RMKMC [7]	0.7095	0.6167	0.5894
mPAC [10]	0.8143	0.7508	0.7320
FMR [1]	0.8524	0.7490	0.7381
GMC [38]	0.8952	0.8200	0.7997
ours	0.8714	0.8076	0.8029

From Table 4 and Table 5, we can observe that the proposed method achieve the best ACC on both Caltech7 and Caltech20. When it comes to F-score, COMVSC also performs the best result on dataset Caltech7. While on Caltech20, it is second only to the best value 1.83% and goes above the rest methods more than 6.49%.

TABLE 4
The ACC, NMI and F-score of the compared algorithm on Caltech7

Methods	ACC	NMI	F-score
FeatConcate	0.5010	0.3735	0.4939
Co-reg_centroid [36]	0.4024	0.3510	0.4236
Co-reg_pairwise [36]	0.4323	0.4641	0.4634
MLRSSC [23]	0.7592	0.5505	0.6998
LMSC [37]	0.5217	0.5527	0.5382
RMKMC [7]	0.3725	0.4054	0.4447
mPAC [10]	0.7164	0.5652	0.6725
FMR [1]	0.4423	0.4752	0.4849
GMC [38]	0.6920	0.6920	0.7217
ours	0.7890	0.4769	0.7417

In summary, the above experimental results have well demonstrated the effectiveness of our proposed COMVSC comparing to other state-of-the-art methods. We attribute the superiority of the proposed method to two aspects: i)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 make full use of high-quality labels to guide the generation of similarity matrices in next iteration and further improve the performance. ii) Comparing with the existing similarity-fusion methods, the proposed COMVSC fuses multiple subspace information in partition level and learns a consensus partition, which verifies the advantages of combing high-level information, which is more informative and less noise and redundancy.

TABLE 5
The ACC, NMI and F-score of the compared algorithm on Caltech20

Methods	ACC	NMI	F-score
FeatConcate	0.3165	0.3629	0.2782
Co-reg_centroid [36]	0.3859	0.4964	0.3444
Co-reg_pairwise [36]	0.3801	0.5378	0.3386
MLRSSC [23]	0.4606	0.3857	0.4252
LMSC [37]	0.5088	0.6356	0.4347
RMKMC [7]	0.4116	0.5097	0.3737
mPAC [10]	0.5478	0.5911	0.5179
FMR [1]	0.3864	0.5334	0.3399
GMC [38]	0.4564	0.4809	0.3403
ours	0.5532	0.3780	0.4996

These two factors contribute to the significant improvements in clustering performance.

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 MSRC-v1, as can be seen in Fig. 2, the clustering structures become clearer with the algorithm iterating, which indicates the feasiblity and effectiveness of the proposed method.

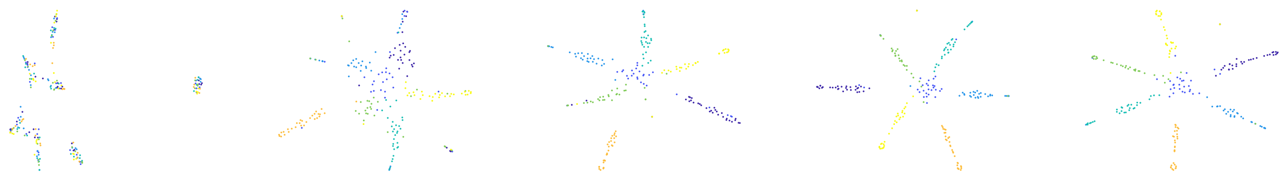
Fig. 3 shows the increasing performance of COMVSC within iterations on MSRC-v1 in terms of ACC NMI and F-score, which also verifies that the labels matrix derived from the consensus representation further guides the update of the consensus representation. The better consensus representation leads to better view-specific representations. They facilitate each other at each iteration to generate promising performance.

6.5 Parameter analysis

We also conduct the parameter sensitivity analysis. The proposed COMVSC method has two hyper-parameters $\{\lambda, \gamma\}$. λ is tuned in the range of$\{10, 20, 30\}$, while γ is tuned in the range from 1.2 to 1.8 with incremental step 0.2. Taking Caltech7 and Wikipedia Articles as examples, we indicate the influence of parameter λ and γ on accuracy in Fig. 4. We can observe that our method's clustering performance is quite stable under the range of $\lambda \in \{10, 20\}, \gamma \in \{1.6, 1.8\}$ across different datasets.

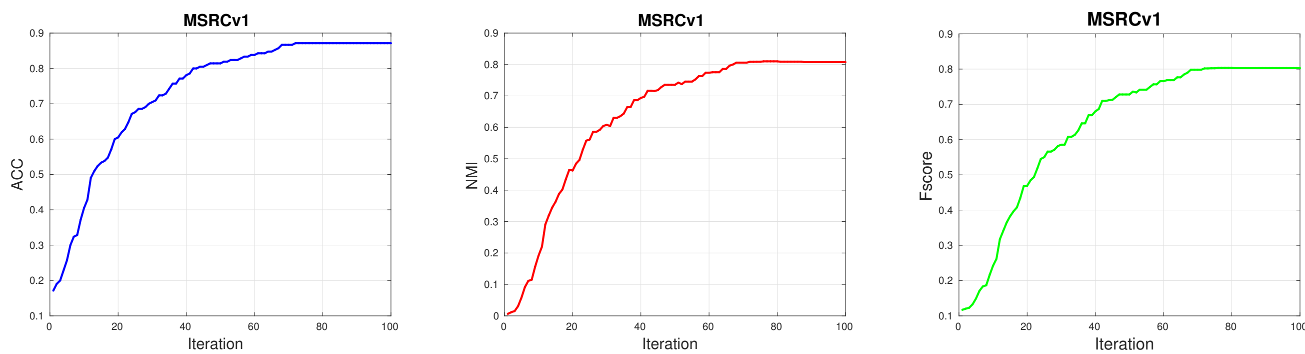
6.6 Ablation Study

To further illustrate the effectiveness of spectral rotation, we conduct an ablation study on the respective term. Specifically, we evaluate the clustering performance without spectral rotation term(w/o rotation) and compare it with



(a) 1 iter (b) 25 iter (c) 50 iter (d) 75 iter (e) 100 iter

Fig. 2. The cluster structure evolution illustration on MSRC-v1. 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.



(a) ACC (b) NMI (c) F-score

Fig. 3. The performance variation curve of the proposed algorithms on the MSRC-v1. In this figure, the ACC(a), NMI(b), and F-score(c) are illustrated.

COMVSC. Observed from Table 6, the spectral rotation term contributes to better local clustering structure and therefore leads to outstanding performance.

TABLE 6
The effectiveness of spectral rotation term on MSRC-v1

Method	MSRC-v1			Wikipedia Articles		
	ACC	NMI	F-score	ACC	NMI	F-score
w/o rotation	0.4619	0.5311	0.4417	0.5960	0.5618	0.4988
COMVSC	**0.8714**	**0.8076**	**0.8029**	**0.6147**	**0.5779**	**0.5214**

7 Conclusion

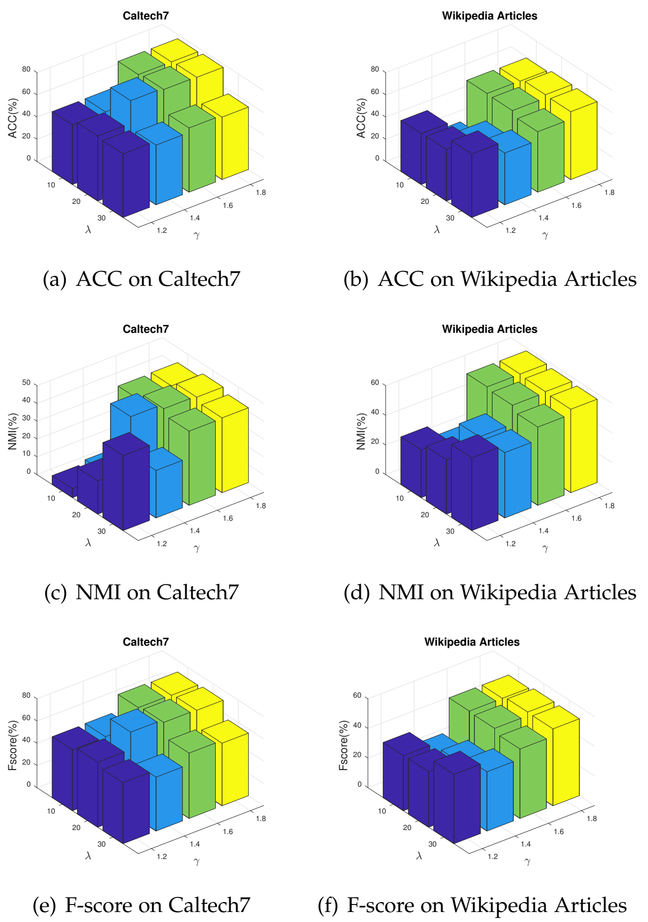
In this paper, we propose a novel one-step consensus multi-view subspace clustering method. Distinct to existing subspace methods, we fuse multiple subspace information in partition level. Furthermore, similarity learning, partition fusion and the clustering processes are combined into a unified framework,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

Acknowledgment

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

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. Wang, J. Lai, and P. S. Yu, "Multi-view clustering based on belief propagation," *IEEE Trans. Knowl. Data Eng.*, vol. 28, no. 4, pp. 1007–1021, 2016.
[3] Y. Ren, X. Yan, Z. Hu, and Z. Xu, "Self-paced multi-task multi-view capped-norm clustering," in *Neural Information Processing - 25th International Conference, ICONIP 2018, Siem Reap, Cambodia, December 13-16, 2018, Proceedings, Part IV*, ser. Lecture Notes in Computer Science, L. Cheng, A. C. Leung, and S. Ozawa, Eds., vol. 11304. Springer, 2018, pp. 205–217.
[4] 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, 2019.
[5] G. Chao, S. Sun, and J. Bi, "A survey on multi-view clustering," *arXiv preprint arXiv:1712.06246*, 2017.
[6] D. Huang, C. Wang, J. Wu, J. Lai, and C. Kwoh, "Ultra-scalable spectral clustering and ensemble clustering," *IEEE Trans. Knowl. Data Eng.*, vol. 32, no. 6, pp. 1212–1226, 2020.
[7] H. H. Xiao Cai, Feiping Nie, "Multi-View K-Means Clustering on Big Data," *IJCAI*, vol. 31, no. 5, pp. 724–730, 2013.
[8] 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.
[9] 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.
[10] Z. Kang, Z. Guo, S. Huang, S. Wang, W. Chen, Y. Su, and Z. Xu, "Multiple Partitions Aligned Clustering," *IJCAI*, pp. 2701–2707, 2019.
[11] 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.
[12] H. D. Iii, "A co-training approach for multi-view spectral clustering abhishek kumar."



(a) ACC on Caltech7 (b) ACC on Wikipedia Articles

(c) NMI on Caltech7 (d) NMI on Wikipedia Articles

(e) F-score on Caltech7 (f) F-score on Wikipedia Articles

Fig. 4. Illustration of parameter sensitivity of the proposed algorithm on Caltech7 and Wikipedia Articles in terms of ACC NMI and F-score.

[13] A. Kumar, P. Rai, and H. Daume, "Co-regularized multi-view spectral clustering," in *Advances in neural information processing systems*, 2011, pp. 1413–1421.
[14] 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.
[15] 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.
[16] 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.
[17] Y. Chen, X. Xiao, and Y. Zhou, "Jointly learning kernel representation tensor and affinity matrix for multi-view clustering," *IEEE Transactions on Multimedia*, 2019.
[18] 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.
[19] 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.
[20] 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.
[21] 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.
[22] 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.
[23] M. Brbić and I. Kopriva, "Multi-view low-rank sparse subspace clustering," *Pattern Recognition*, vol. 73, pp. 247–258, 2018.
[24] 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.
[25] X. Cao, C. Zhang, H. Fu, and S. Liu, "Diversity-induced Multi-view Subspace Clustering," *CVPR*, 2015.
[26] 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.
[27] 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.
[28] H. Gao, F. Nie, X. Li, and H. Huang, "Multi-view subspace clustering," *ICCV*, vol. ICCV, pp. 4238–4246, 2015.
[29] 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.
[30] Fan and K., "On a theorem of weyl concerning eigenvalues of linear transformations: Ii," *Proc Natl Acad Sci U S A*, vol. 36, no. 1, pp. 31–35, 1950.
[31] Y. Pang, J. Xie, F. Nie, and X. Li, "Spectral clustering by joint spectral embedding and spectral rotation," *IEEE transactions on cybernetics*, 2018.
[32] F. Nie, X. Wang, and H. Huang, "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.
[33] 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.
[34] 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 Trans Pattern Anal Mach Intell*, vol. 36, no. 3, pp. 521–535, 2014.
[35] Y. Li, F. Nie, H. Huang, and J. Huang, "Large-scale multi-view spectral clustering via bipartite graph," in *Twenty-Ninth AAAI Conference on Artificial Intelligence*, 2015.
[36] A. Kumar, P. Rai, and H. D. Iii, "Co-regularized multi-view spectral clustering."
[37] 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.
[38] H. Wang, Y. Yang, and B. Liu, "Gmc: Graph-based multi-view clustering," *IEEE Transactions on Knowledge and Data Engineering*, 2019.

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..

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.